Polymorphism analysis of *Plasmodium falciparum* merozoite *EBL*1, *RH4* and *MSP1* ligands in a severe malaria population in Kilifi, Kenya

MUSABYIMANA Jean Pierre

I56/83171/2012

Thesis submitted to the Centre for Biotechnology and Bioinformatics in partial fulfillment for the award of Master of Science degree in Biotechnology (Health option), University of Nairobi.

August 2015

DECLARATION

I declare that this research is my own work, and has not been submitted for examination in any other university.

Name: Jean Pierre, MUSABYIMANA

Registration: I56/83171/2012

Signature: _____Date_____

APPROVAL

This thesis has been submitted with our approval as University Supervisors:

Dr. Isabella L. Oyier, PhD

Centre for Biotechnology and Bioinformatics

University of Nairobi

Signature: _____Date_____

Prof. James O. Ochanda, PhD

Centre for Biotechnology and Bioinformatics

University of Nairobi

Signature: _____Date_____

ACKNOWLEDGEMENT

I would like to acknowledge Dr. Isabella Oyier, my primary supervisor, for being my inspiration and role model due to the unique and incomparable work she did in helping me to conceptualize this project from a crude idea and for her proactive guidance throughout the process of developing protocols, data analysis and writing this thesis. I would also like to acknowledge Prof. James Ochanda, who helped me to choose the course of health biotechnology and who has been guiding me since I started my studies at the Center for Biotechnology and Bioinformatics (CEBIB) as a foreign student. I also recognize help of all sorts from all CEBIB staff and students

I also acknowledge the RWANDA Ministry of Health (MoH) for funding my studies and the KEMRI-Wellcome Trust Collaborative Research Programme, Center for Geographic Medicine Research-Coast, Kilifi which supported this work through the Malaria Capacity Development Consortium Initiative grant awarded to Dr. Isabella Oyier. Many thanks also to the Programme for providing the samples and laboratory facilities. I give special thanks to Mrs Esther Kiragu, Mrs Omedo Irene Akinyi, Mr James Tuju, and Mr. Moses Mosobo for precious technical and moral support during my laboratory work.

This work is dedicated to all kids affected by malaria.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATONS

AMA : Apical membrane Antigen
CEBIB : Center for Biotechnology and Bioinformatics
CHO-K : Chinese Hamster Ovary Cell line K
CR1 : Complement Receptor 1
DBL-EBP : Duffy-binding like Erythrocyte-binding protein
DNA : Deoxyribo Nucleic Acid
EBA : Erythrocyte-binding Antigen
ebl : Eythrocyte-binding like
EGF : Epidermal Growth Factor
GPA : Glycophorin A
GPB : Glycophorin B
GPI : Glycosylphosphatidyl Innositol
IMC : Inner membrane complex
kDa : Kilo-Dalton
KDH : Kilifi District Hospital
KEMRI : Kenya Medical Research Center
msp : Merozoite Surface protein
PCR : Polymerase Chain Reaction
Pf : *Plasmodium falciparum*
RBC : Red Blood Cell
Rh : Reticulocyte Binding -like homologue
SNP : Single Nucleotide Polymorphism

ABSTRACT

Around 240 million people develop symptomatic malaria and around 1.24 million die annually from malaria globally. Over 80% of those deaths occur in sub-Saharan Africa. Symptomatic malaria in humans is initiated when merozoites, the free invasive blood-stage form, invade circulating erythrocytes. Erythrocyte invasion is facilitated by merozoite protein ligands such as erythrocyte binding ligand 1 (EBL1), merozoite surface protein 1 (MSP1) and Reticulocyte binding like homologue 4 (RH4) which bind the erythrocyte surface proteins Glycophorin B (GPB), Band 3 and Complement Receptor 1 (CR1), respectively. The merozoite proteins are also thought to be potential vaccine candidates since antibodies to these antigens have been detected in individuals exposed to malaria infection. This study aimed to analyze polymorphisms in genes coding the 3 ligands (mentioned above) receptor binding sites and compare genetic variation patterns between ligands and receptors in a population with severe malaria in Kilifi, Kenya.

The binding sites of MSP1, RH4 and EBL1 genes were amplified using conventional PCR, then sequenced by BIGDYE terminator 3.1 chemistry. Tests of neutrality and linkage disequilibrium were computed for all the 3 genes. This study identified 3 and 6 SNPs in RH4 and MSP1, respectively. EBL1 was found to be more polymorphic, with region I having 13 SNPs and region II, 8 SNPs. Of all the genes, only EBL1 like its receptor Glycophorin B (Tajima's D of -1.81 at exon 4, $p<0.05$) was found to be under purifying selection, with a Tajima's D of -1.87248 (p value<0.05) based on the receptor binding domain (region II), implying an excess of rare variants within the gene from the population studied. Receptors had similar neutrality values with their respective ligands. There was no significant selection in MSP1 and RH4 despite the mutations observed. These findings indicate that genetic variations have created different haplotypes of

which one or two are most predominantly circulating in the studied population. Proteins like RH4 and MSP1 show hope as potential vaccine candidates due to their previously demonstrated role of inducing immune response and the conserved nature of their receptor binding domains showcased in this study

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Malaria is a major disease for humans, of the approximately three billion people who are exposed annually, more than 240 million people develop symptomatic malaria (Murray *et al*., 2012). Annual malaria deaths are estimated at 1.24 million and over 80 percent of them occur in sub-Saharan Africa (Murray et al., 2012). Malaria is caused by a parasite of the Apicomplexa Phylum. Four *Plasmodium* species can infect humans and cause malaria, *P. vivax*, *P. malariae*, *P. ovale* and *P. falciparum*. *P. falciparum* causes the most severe form of malaria (Cai et al., 2012).

1.1. Parasite life cycle

Plasmodium has a complex life cycle (Figure 1.1) that occurs in both human and mosquito hosts. It undergoes a sexual stage in the mosquito and an asexual stage in humans. An infected *Anopheles* mosquito during its blood meal injects sporozoites into the blood which infects liver cells (hepatocytes) and thereafter develops into merozoites. Merozoites are released into the blood where they invade erythrocytes. In the erythrocyte, a merozoite becomes a trophozoite, which undergoes endomitosis and becomes a schizont. The schizont bursts (egress) and releases new merozoites into the blood and these merozoites can infect new erythrocytes. Instead of developing to trophozoites some merozoites can develop to gametes due to mechanisms that are yet to be well understood. Once these gametes (male and female) are taken up by a mosquito during a blood meal they can progress to sexual stage of the life cycle in the mosquito (Kuhn and McCarthy 2006).

Most of the disease symptoms are associated with the asexual parasite infection within the bloodstream. This is a stage initiated when merozoites invade circulating erythrocytes. Although this invasion process, which is our focus for this study is rapid, the parasite is still exposed to the host immune system because the merozoite spends some seconds in the bloodstream before entering the red blood cells (Cowman et al., 2012).

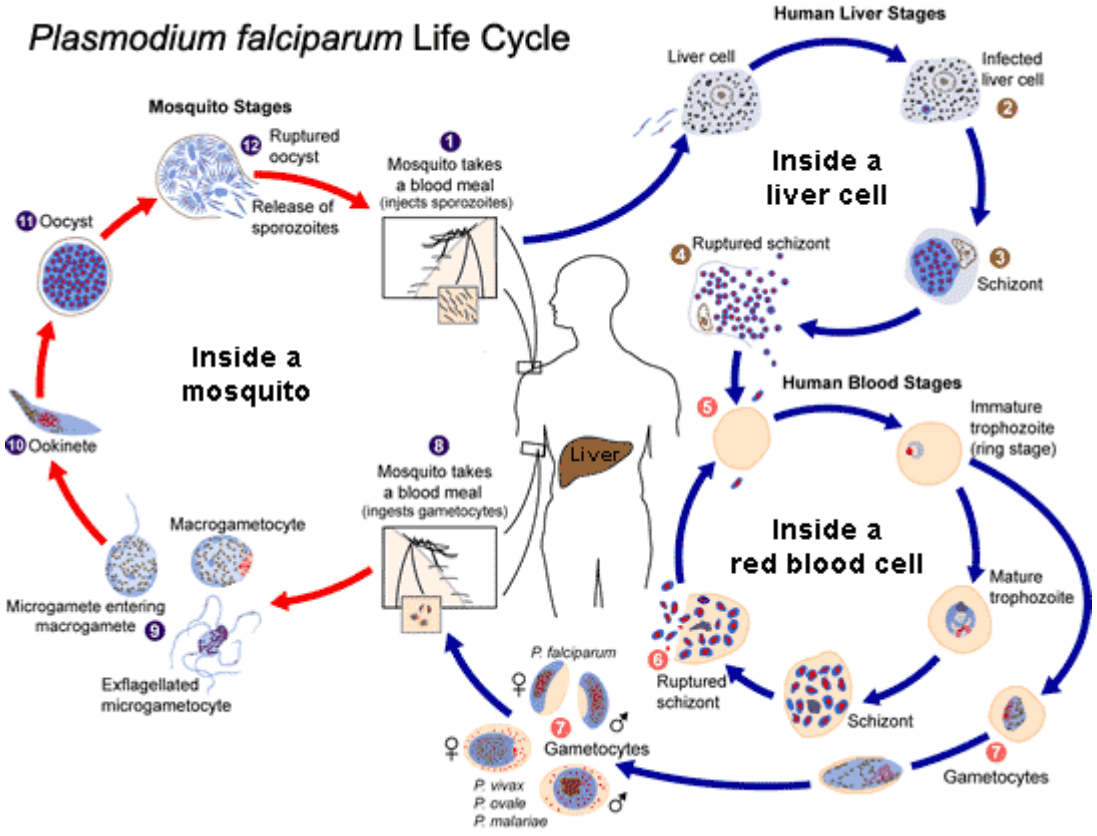


Figure 1.1. The life cycle of P. falciparum *(*1) The Anopheles mosquito bites a human and injects sporozoite forms. (2) These move to the liver and invade hepatocytes, in which (3-4) they develop to produce exoerythrocytic merozoite forms that are released into the blood stream. (5) Merozoites invade erythrocytes and grow into trophozoites and mature schizonts. (6) Merozoites are released that reinvade new erythrocytes. (7) Gametocytes, formed from the asexual blood stage, are taken up by a feeding mosquito into the gut where (8-9) they mature to form male and female gametes. (10) The fertilized zygote develops to an ookinete, then (11) oocyst and finally sporozoites that migrate to the salivary glands (12). http://www.cdc.gov/malaria/about/biology/

P. falciparum-infected individuals show a wide spectrum of clinical manifestations that range from a state of asymptomatic infection to severe life-threatening forms, such as hyperparasitaemia, hypoglycemia, cerebral malaria, respiratory distress, and vital organ dysfunction. Symptoms begin to manifest after merozoites appear in the blood and invade red blood cells, following a silent hepatocytic stage (Marsh et al., 1995).

1.2. Merozoite biology and erythrocyte invasion

The merozoite is the smallest cell within the *Plasmodium* life cycle (Bannister et al. 1986). Its organelle repertoire is conventional like other apicomplexan cells (Morrissette and Sibley 2002). It has an apical complex of secretory organelles (micronemes, rhoptries, and dense granules), mitochondrion, nucleus, and a plastid (apicoplast) (Figure 1.2; McFadden et al. 1996).

The mature merozoites are propelled from the bursting schizonts (Dvorak et al. 1975, Glushakova et al. 2005; Abkarian et al. 2011) and invade the erythrocyte through a complex process (Figure 1.3). They randomly associate with erythrocytes by a reversible attachment which results in the deformation of the erythrocyte surface. Attachment is followed by a very active process of reorientation that attaches the parasite apical complex to the host cell membrane forming an irreversible interface. This interface between merozoite and erythrocyte, also called the tight junction pushes the parasite inside the erythrocyte (Aikawa et al. 1978), a process called "internalization" facilitated by many proteins using different pathways (Gilson and Crabb 2009).

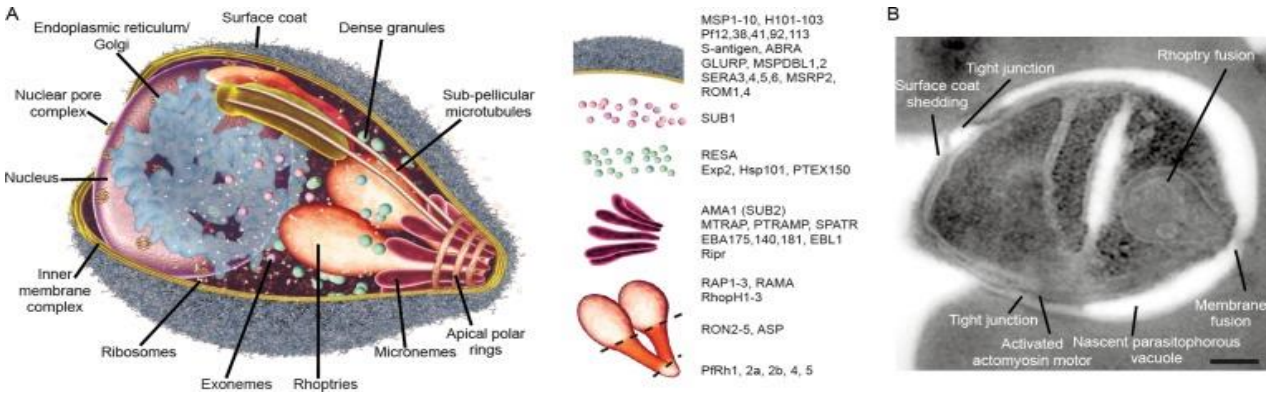


Figure 1.2: Three-dimensional diagram of a merozoite and its core secretory organelles. (A) The sectioned cell showing the major cellular architecture and organelle repertoire of the invasive merozoite, with dissected organelles listing core molecular constituents of these key invasion-related compartments. MSP proteins are localized on the surface. The secretory organelles are dense granules, micronemes, and rhoptries. There exist sub-compartmentalization within organelles (specifically the rhoptries). The rhoptries are divided into three segments, with PfRh family proteins in the most distal segment and RON2-5 in the next segment. RESA (Ring-infected erythrocyte surface antigen) is released from dense granules and exported to the infected red blood cell. The body of the rhoptry bulb contains lipids and other proteins involved in forming the parasitophorous vacuole, including RAP1-3 and RAMA. Micronemes secrete proteins including the PfDBL-EBP family. (B) A *P. falciparum* merozoite in the process of invading a human red blood cell (image courtesy of S. Ralph, University of Melbourne, Melbourne, Australia) (Cowman et al., 2012).

The invasion process revolves around the tight junction also called moving junction. This is a very key interface between parasite and host. This structure coordinates distinct stages after egress and attachment, facilitating invasion and post-invasion sealing of the parasite within the erythrocyte. This incompletely understood process of invasion, previously explained only at the cellular level, involves an arsenal of molecules, many of which are located on the surface or within micronemes and rhoptries, which form the molecular basis of invasion (Figure 1.2). The invasion is followed by a brief period of morphological spiking of the cell referred to as echinocytosis of the red cell stimulated by efflux of potassium and chloride ions. The erythrocyte returns to its normal shape within 10 minutes (Gilson and Crabb, 2009). The internalized parasite, now referred to as a ring, undergoes rapid and dramatic changes in shape after this process (Cowman et al., 2012).

1.3. Molecules involved in erythrocyte invasion

Since the genome sequence, proteomic and transcriptional information of *P. falciparum* and other *Plasmodium spp*. was made publicly available, many proteins associated with the merozoite have been identified (Gardner et al. 2002). Some of these proteins such as the ones located on the surface of merozoite play role in the initial invasion steps. Being exposed to immune responses, these surface proteins are of interest because they are considered as prime vaccine candidates (Cowman et al., 2012).

The main and most abundant surface protein is MSP1and plays an initial role in the attachment of the merozoite to the erythrocyte (Egan et al. 1996; Cowman and Crabb 2006). MSP1 is functionally conserved and is associated with the parasite membrane via a Glycophosphatidylinositol (GPI) anchor (Gerold et al. 1996). Since its discovery (Holder et al., 1988), a greatly expanded repertoire of surface proteins have been assembled (Cowman, Berry, and Baum 2012). These include other merozoite surface proteins such as MSP3, MSP7 and a serine repeat antigen (SERA) protease like family. Nine of the SERA proteins can be found in *P. falciparum a*nd contain a papain-like protease domain as well as regions that are likely involved in protein–protein interactions with other GPI-anchored proteins such as MSP1 (Cowman et al., 2012). Although these proteins are abundant on the surface of the merozoite, they are not evenly spread and their functions are not fully known (Cowman et al., 2012).

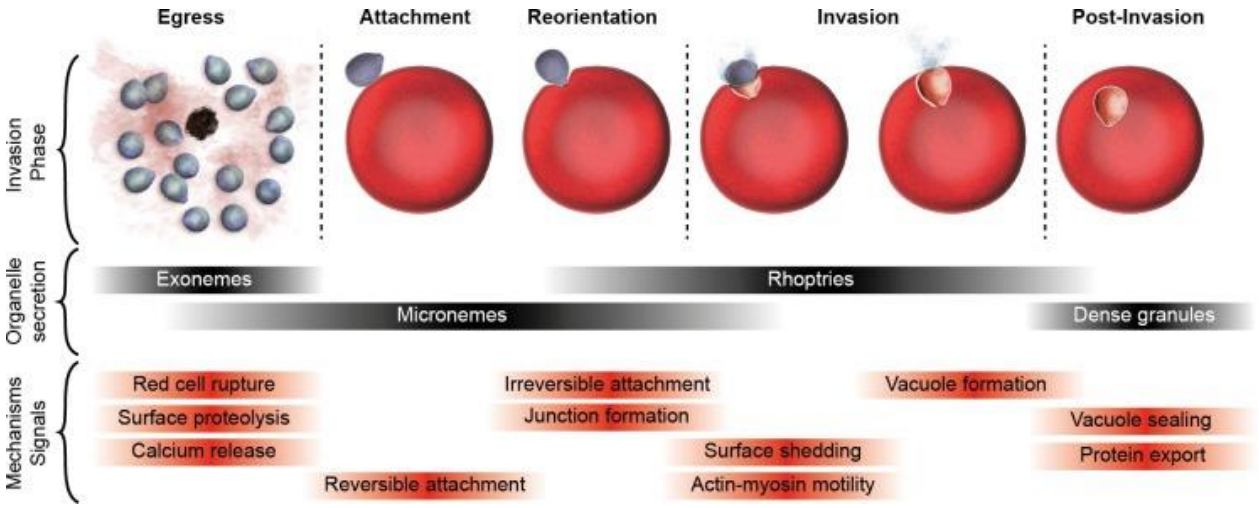


Figure 1.3: A time course of merozoite invasion of the erythrocyte from egress through post invasion. A cellular overview is linked with timing of organelle secretion and key mechanistic or signaling steps. After apical reorientation, the merozoite establishes a tight junction involving RON4 and AMA1 proteines. The tight junction is ultimately connected to the actomyosin motor. As the tight junction moves across the merozoite surface, proteins are shed off through the activity of proteases such as ROM4, ROM1, SUB1, and SUB2. The parasitophorous vacuole and membrane are mainly formed from the rhoptries engaging some red cell membrane components. These components expel their contents, forming the space into which the parasite can move under the action of the actomyosin motor. Post invasion sealing mechanisms of membranes after reaching the posterior end of the parasite remain unkown (Cowman et al., 2012).

The other proteins that have been shown to be important in invasion are located mainly in the micronemes and rhoptries (Cowman and Crabb 2006b). These proteins function directly in invasion in different pathways.

The invasion proteins can be divided into two classes: adhesins that function as ligands binding directly to specific receptors on the erythrocyte and invasins that function in the invasive process but do not necessarily bind directly to receptors on the host cell. Adhesins are the ones located in micronemes and rhoptries, and are in general are *Plasmodium*-specific or provide cell specificity which means they restrict parasites to the erythroid lineage in the case of merozoite invasion.

The main adhesins belong to two protein families which include the erythrocyte-binding proteins (EBP) and Reticulocyte Binding–like homologues (Rh), localizing to the micronemes and neck of the rhoptries, respectively (Figure 1.2). Different *P. falciparum* strains are able to invade using alternative host receptors due to the PfRh and PfEBP protein families (Rayner et al. 2000). Host receptors for some adhesins have been studied. Erythrocyte Binding Antigen 175 (EBA-175), Erythrocyte Binding Antigen 140 (EBA-140; also known as Baebl) and Erythrocyte Binding-like Ligand 1 (EBL1) which are members of the EBP family bind to glycophorin A, C and B, respectively (Mayer et al. 2009). PfRH4 and Rh5 belonging to PfRh family of proteins bind to complement receptor 1 (Tham et al., 2010) and erythrocyte surface CD147 or basigin (Crosnier et al. 2013), respectively.

Proteins involved in merozoite invasion with no necessity of receptors have also been identified in the merozoite. The best known of these invasins is the Apical Membrane Antigen-1 (AMA1) considered to be an important vaccine candidate due to its highly conserved nature and high concentration at the apical complex of the merozoite (Treeck et al. 2009). This protein shares similarity with EBLs. By interaction with the rhoptry neck proteins (RON), AMA1 participates in the tight junction formation (Cowman et al., 2012).

For this project, the focus was on at least one of the proteins expressed in each of the main parts of the merozoite which are so far known to play a role in the attachment and invasion of the erythrocyte using different pathways, during the blood stage of the life cycle. The proteins involved are among those of which erythrocyte receptors are known and thought to elicit the production of antibodies in humans (Crosnier et al. 2013; Pandey et al. 2013).

i) Erythrocyte binding ligand 1 and its gene family

EBL1 is a DBL-EBP family member. *P. falciparum* has four other DBL-EBP genes which are erythrocyte- binding antigen 175 (EBA-175), erythrocyte-binding antigen 140 (BAEBL/EBA-140), erythrocyte-binding antigen 181 (JESEBL/ EBA-181), and erythrocyte-binding antigen-165 (EBA-165) so far identified as a pseudogene. The erythrocyte-binding domains reside in the N-terminal cysteine-rich region (region II). This domain is duplicated into F1 and F2 domains which are both required for optimum binding to erythrocytes (Mayer et al. 2009).

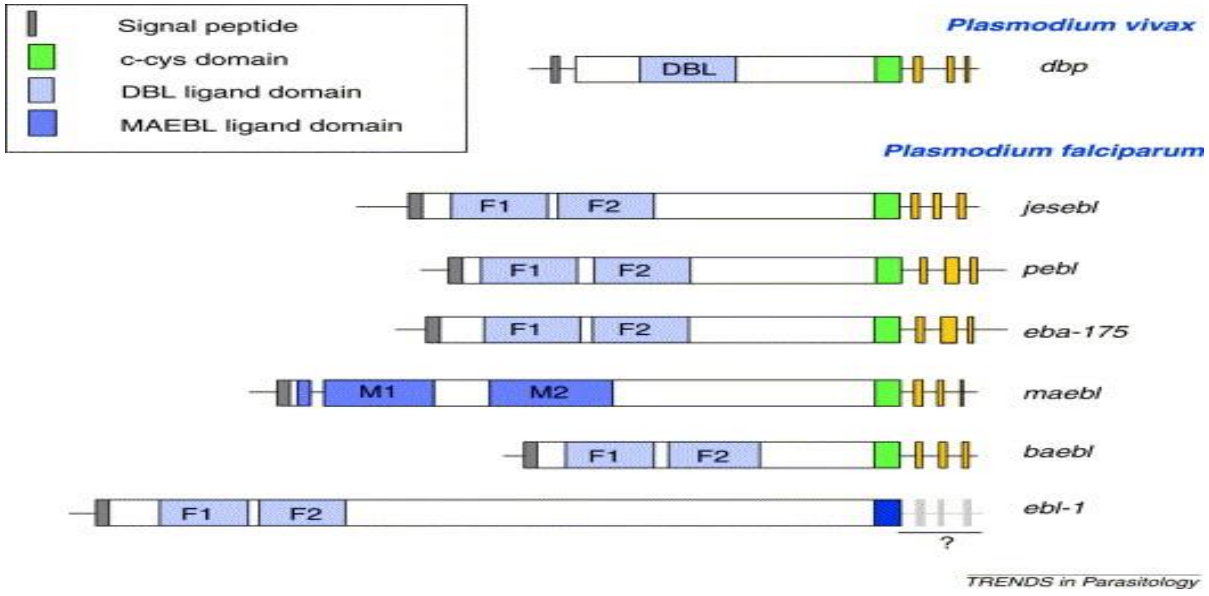


Figure 1.4: Summary of the Duffy-binding–like (DBL) domains of the ebl family gene structures and products of Plasmodium molecules sharing sequence similarity in one or several domains: Duffy-binding–like (DBL) domains of the ebl family; the carboxyl cysteine-rich region (C-cys) found in the ebl family; (DBP), Duffy-binding protein; F1 and F2, duplicated DBL domains; eba (EBA), erythrocyte-binding antigen; ebl (EBL), erythrocyte-binding–like. ebp (EBP) is a generic term for erythrocyte-binding protein; the product name is accompanied with the prefix corresponding to the Plasmodium species where it was identified (Adams et al. 2001).

Although EBL1 shares with the other members of the DBL-EBP family (Figure 1.4) a high degree of sequence homology in region II, it has been shown that it differs from them in the 3' cysteine rich region, the most conserved region of the DBL-EBP family, even between species, suggesting that EBL1 has a unique function (Mayer et al. 2009; Adams et al. 1992). Nevertheless, the EBL1 genes from approximately half of *P. falciparum* clones have 5 thymidines inserted into the Open Reading Frame (ORF; part of DNA that has potential to code for protein or peptide) that leads to a frameshift mutation (framing error) resulting in premature translational termination, suggesting that the gene is evolving to be a pseudogene (Mayer et al. 2009) and may not be essential for invasion. This is probably why its receptor, Glycophorin B has shown high level of polymorphism in individuals living in malaria-endemic regions (Gaur et al. 2003). Since EBL-1 is becoming pseudogene, its receptor can relax and accumulate mutations without net effect.

ii) Reticulocyte binding Homolog (*Pf*Rh) 4

The PfRH4 is a PfRH protein member that is made of 1716 amino acids protein (in the 3D7 strain) (Lopez-Perez et al. 2012). PfRH4 is essential in the sialic acid-independent pathway as it is demonstrated by the disruption of the gene in *Plasmodium falciparum* W2mef strain (Tham et al. 2010). By activating PfRH4 expression, the parasite is able to switch receptor usage from sialic acid–dependent to sialic acid-independent pathways, thus providing a mechanism for the parasite to invade via different pathways (Tham et al. 2011). It has been shown that this ligand, by its 160 kDa fragment, binds to the receptor CR1 (Tham et al. 2011). Reticulocyte binding Homolog (PfRh) 4 binding site has been identified as PfRH4.9 (Tham et al., 2012) with the size of 1817 base pairs. This region has been found to have low levels of polymorphisms (Reiling et al. 2012).

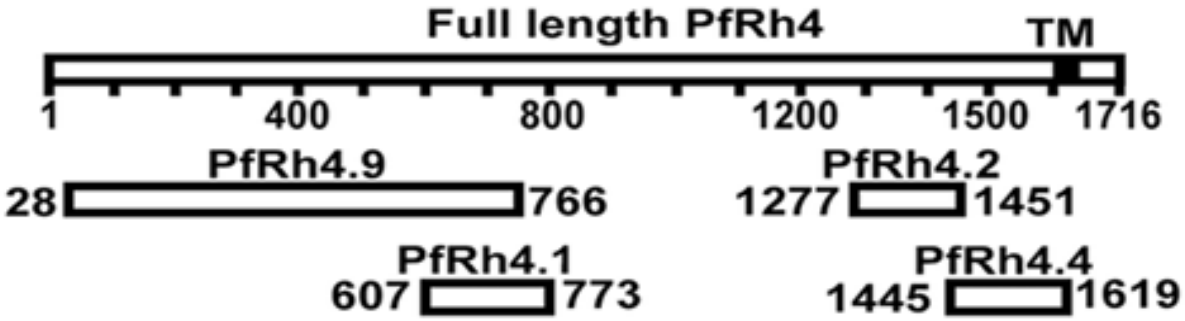


Figure 1.5: PfRH4 as a target of human antibodies. Schematic representation of the PfRH4 recombinant proteins used in Tham et al, 2012 study, relative to full length PfRH4. TM: transmembrane domain. Numbers represent amino acids (Reiling et al. 2012)

iii) Merozoite surface protein (MSP1)

The *P. falciparum* merozoite surface protein 1 (MSP1) is a major membrane protein attached to the merozoite surface at its C terminus through a glycosyl-phosphatidyl-inositol (GPI) anchor (Pan et al. 1999). It is among the first proteins to attach to the erythrocyte surface. It binds its receptor, Band 3, a protein at the surface of erythrocyte through its 42kDa protein fragment (Pan et al. 1999; (Goel et al. 2003). During schizogony or after merozoite release, full-length MSP1 is processed to give proteolytic 83, 30, 38 and 42 kDa products referred to as $MSP1_{83}$, $MSP1_{30}$, $MSP1_{38}$, and $MSP1_{42}$ fragments (Figure 1.6), respectively which together form a non-covalent complex on the merozoite surface. Subsequently, $MSP1_{42}$ undergoes secondary processing to give a 19 kDa fragment referred to as $MSP1_{19}$ (the C-terminal domain of $MSP1_{42}$), which is carried into the newly invaded RBCs while other MSP1 fragments are shed off (Pan et al. 1999; Gerold et al. 1996b).

Allele-specific genotyping of MSP1 showed a high genetic diversity in *P. falciparum* populations studied in different countries. Around twenty-five alleles of MSP1 were observed in different studies in Africa, in which, the K1 allelic family was consistently predominant, (Ayala et al., 2006; Soulama et al., 2009).

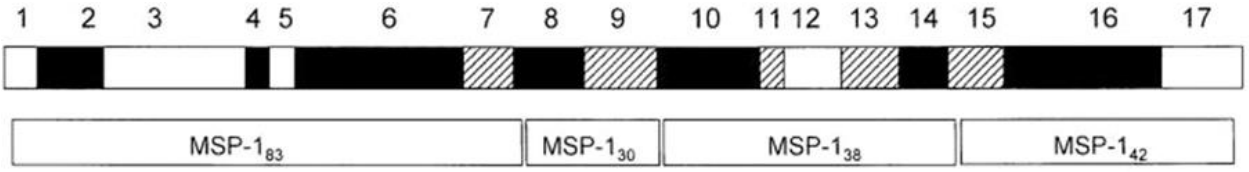


Figure 1.6: Schematic of recombinant MSP1 constructs relative to the MSP1 gene and its products. Shown is a diagrammatic representation of the complete MSP1 gene, divided into conserved (open) blocks 1, 3, 5, 12, and 17, semi-conserved (hatched) blocks 7, 9, 11, 13, and 15, and poorly conserved or polymorphic (filled) blocks 2, 4, 6, 8, 10, 14, and 16, as defined by Tanabe et al. (José et al. 1997). The positions of the MSP1 primary processing products ($MSP1_{83}$, $MSP1_{30}$, $MSP1_{38}$ and $MSP1_{42}$) are shown relative to the gene (José et al. 1997). $MSP1_{19}$ is a fragment that remains after cleavage of the $MSP1_{42}$ fragment.

Previous studies on MSP1 polymorphisms focused on Block 2 region. The blocks in MSP1 gene were set by convention of Tanabe (Peterson et al. 1988). This study focuses on Block 17 which is the region that encodes for the 19kDa C-terminal domain of the MSP1 protein, which is the domain that remains after undergoing the proteolytic maturation. It is also the region involved in interacting with band 3, the erythrocyte receptor (Goel et al. 2003).

1.4. Polymorphisms and natural selection

Although merozoite proteins have been thought to be potential antigen candidates for immunity, the polymorphisms in their genes have been considered as a challenge thus creating a dilemma when designing vaccines to these antigens (Conway, 1997). Polymorphism may create different haplotypes of a certain protein and one vaccine may not be able to successfully target all the existing haplotypes for one protein. This may on the other hand serve as the selective pressure favoring haplotypes that are not targeted by the vaccine (Polley, Chokejindachai, and Conway 2003). Some researchers proposed that either polymorphic regions of antigens be dismissed from further consideration as vaccines or that they deserve particular study to determine if they have functional significance which could implicate them as promising components of a multivalent vaccine (Conway, 1997).

Detection of Single Nucleotide Polymorphisms (SNPs) is one of the ways of determining genetic diversity. This is a situation whereby different variants differ by a single nucleotide at a particular locus in a genome. SNPs have been studied in different genome projects and evolutionary studies. In the *Plasmodium falciparum* genome, one of the recent studies by Mu et al (2007) found that SNPs are not distributed evenly across chromosomes; rather, some regions have consecutive genes without any SNPs, and other DNA segments have consecutive genes with multiple SNPs (Figure 1.7). The proportion of genes with SNPs varies from chromosome to chromosome, ranging from 47.1% (chromosome 13) to 67.5% (chromosome 7) in *P. falciparum* genome (Mu et al. 2007).

Interestingly, genes encoding surface antigens, cell adhesion molecules and proteins involved in drug interactions are mostly polymorphic because they are exposed to host immune system which works as a selective pressure and this has encouraged studies on specifically immunogenic and binding regions of these genes. The group of genes presenting immunogenic domains have a high ratio of nonsynonymous pairwise differences per nonsynonymous site (pN) relative to synonymous pairwise differences per synonymous site (pS), which suggests balancing, diversifying or partial directional selection. Additionally, estimates of Tajima's D, a measure of the frequency distribution of alleles, across chromosomes also identified some genes that show an excess of diversity indicative of balancing selection, such as eba-175, which has been shown to be under strong balancing selection (Escalante, Lal, and Ayala 1998; Kang et al. 2012).

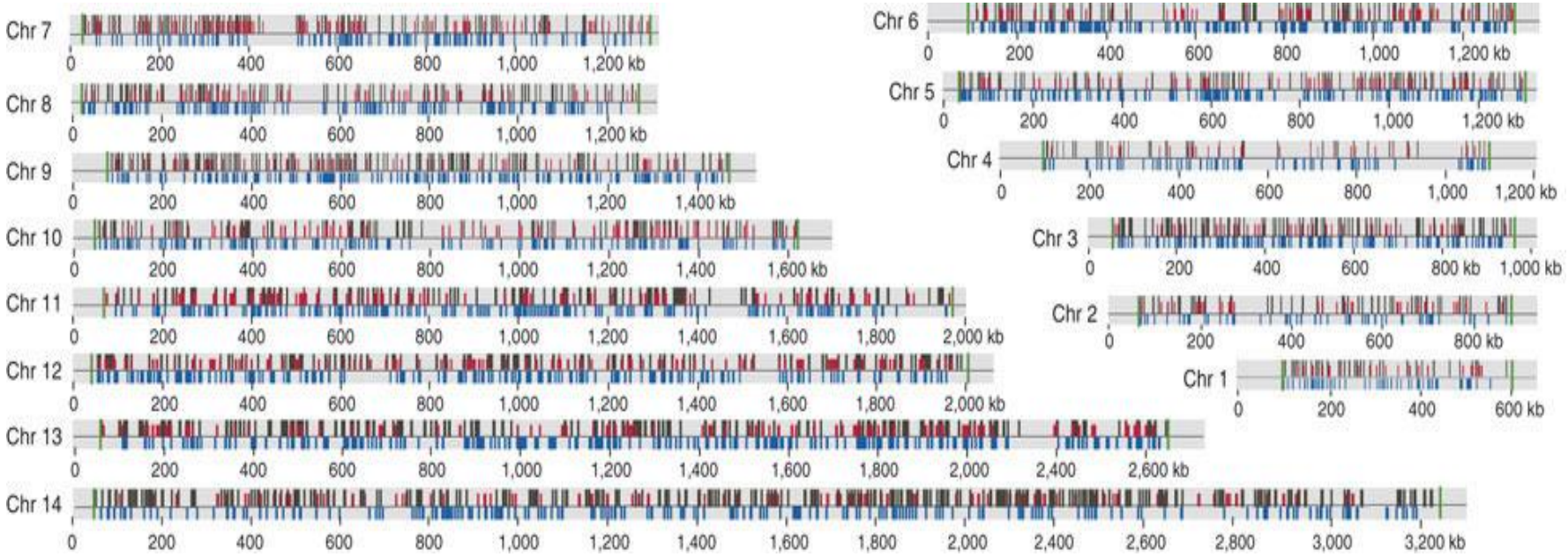


Figure 1.7: Physical maps showing distribution of polymorphic sites across 14 P. falciparum chromosomes. Vertical bars represent SNPs (black, nonsynonymous substitutions; red, synonymous substitutions) or microsatellites (blue, under the horizontal lines). Only one nonsynonymous SNP and one synonymous SNP are presented if there was more than one SNP in a gene (noncoding SNPs were grouped with synonymous SNPs). Most of the chromosomal ends (green vertical bars) were excluded because of gene families such as var, rifin and stevor. Diagram adopted from (Mu et al. 2007).

Immune responses have been listed as the powerful selective pressure for *Plasmodium falciparum* subpopulations (Plebanski et al. 2002; Amambua-ngwa et al. 2012; Wright and Rayner 2014). Little is known about the effect of receptor genetic diversity on the parasite ligands. Identification of antigen haplotypes more predominantly related to genetically diverse receptors may be important for the development of an effective vaccine.

1.5 Association of polymorphisms and invasion

Few studies have suggested a strong effect of polymorphisms on erythrocyte invasion. For instance, a study conducted in South America found that the erythrocyte invasion by parasites expressing the EBA-181 variant Dd2 (RVIQN) were highly sensitive to neuraminidase and chymotrypsin but resistant to trypsin-treatment. This was not observed in parasites containing other EBA-181 variants which showed resistance to neuraminidase and chymotrypsin treatment of erythrocytes (Lopez-Perez et al. 2012). HB3 (RVNQN) expressed on CHO-K1 cells is a variant that has been found unable to bind to untreated or enzyme-treated erythrocytes (Mayer et al. 2004). The same observation that polymorphism in a *P. falciparum* erythrocyte-binding ligand changes its receptor specificity had been underlined with the BAEBL/EBA-140 protein (Mayer et al. 2002) although this concept has been controversial to some degree in that some other researchers did not find evidence about the change of receptor specificity brought by polymorphisms (Maier et al. 2009). This provides a rationale for further research not only on the malarial ligands involved in erythrocyte invasion but also about the polymorphisms that create more alleles in the ligands and their receptors.

Glycophorin A (Sim et al. 1994) and Glycophorin B (Dolan et al. 1994; Mayer et al. 2009) have been identified as the erythrocyte receptors of EBA 175 and EBL1, respectively. These 2 pathways make part of a complex niche of pathways and the choice of one pathway against the other is not well understood. The 5 Ts insertion in the EBL1 gene has been abundantly found in geographically different isolates which suggests that the EBL1/GPB pathway is not used for invasion in different places. (Lopez-Perez et al. 2012; Drummond & Peterson 2005; Githui et al. 2010) .

Similar ligand polymorphism-invasion relationship was studied in the PfRH family of proteins. Some variants of PfRh5 were observed to be more associated with higher rate of invasion than others. For instance in a study conducted in South America it was hence observed that some of the most important PfRH4 polymorphisms such as modification of the DEVE repeat and 10 amino acids insertion were significantly associated with an invasion phenotype of neuraminidase, trypsin and chymotrypsin sensitivity (Lopez-Perez et al. 2012). Analysis based on only a single polymorphism in amino acid 203 in the C203Y variant confirmed association in that the parasites expressing all six cysteine residues in PfRh5 (Cys203 variant), like the 3D7 and Dd2 laboratory strains, showed a higher median rate of invasion into trypsin-treated RBCs (62%) than those containing only five cysteine residues (Tyr 203 variant) observed in the 7G8 laboratory strain (Lopez-Per ez et al. 2012).

This depicts that a complete understanding of the molecular interactions that are the basis of the invasion process is very crucial, not only in improving our knowledge about the basic biology of the malarial parasite, but also for the development of intervention strategies to counter the disease. This study was done to provide understanding on receptor-ligand interactions and the genetics governing invasion in endemic areas. The results presented in this thesis will thus hopefully contribute to the progress towards the development of malaria vaccines.

1.6. Research question and hypothesis

What are the polymorphisms in MSP1, RH4 and EBL1genes and how do they relate to their respective erythrocyte receptor diversity?

Our hypothesis is that *Plasmodium falciparum* merozoite MSP1, RH4 and EBL1 gene polymorphisms do not divert from neutrality proportionally with their respective receptors.

1.7. Objectives

1.7.1. General objective

To identify predominant *Plasmodium falciparum* merozoite EBL1, RH4, and MSP1 polymorphisms and their correlation with polymorphisms in their respective erythrocyte receptors in a severe malaria population in Kilifi, Kenya.

1.7.2. Specific objectives

Specifically, the study intended to:

1. Characterize *PfMSP1, PfEBL1* and *PfRH4* polymorphisms and their frequencies in a population with severe malaria in Kilifi, Kenya

2. Determine whether the polymorphisms in the 3 genes are under selection

3. Compare genetic variation patterns between ligands and receptors in the same population

1.8. Project justification

Merozoite proteins are potential vaccine candidates as they are recognized by the host's immune system before invading RBCs and there is evidence showing that antibodies to these proteins do inhibit merozoite invasion of erythrocytes (Lazarou et al. 2009; Triglia et al. 2011). There is also evidence to show that many of the proteins of the merozoite are highly polymorphic and thus present a challenge when developing vaccines to these genes (Holder et al. 1999).

Polymorphism analysis of micronemal (EBL1), rhoptry (RH4) and surface (MSP1) proteins with known receptors provides insights into the genetics of the merozoite invasion. Furthermore, this work contributes to the body of knowledge in understanding the interactions of these proteins with their receptors, the effect of genetic diversity and how this may impact merozoite antigen vaccine design.

CHAPTER 2

MATERIALS AND METHODS

2.1 Study site

The study was conducted in Kilifi County, in the coastal region of Kenya. Kilifi is a malaria endemic region, although transmission has been declining in recent years (O'Meara et al. 2008; Bejon et al. 2010). The study was conducted specifically in Kilifi County Hospital (KCH) which is part of the Kilifi Health and Demographic Surveillance System (KHDSS) covering an area of 891km^2 and having a population of 261,919, 18% of whom are under 5 years of age (Scott et al., 2012). The KHDSS is part of the KEMRI/ Wellcome Trust Research Programme (KWTRP). This study has been approved by the KEMRI ethical review board for integrated studies of natural immunity to malaria infections.

2.2 Study design and sampling

This cross sectional study involved DNA samples collected in 2001 from children admitted to the Kilifi County hospital's High Dependency Unit (HDU) suffering from severe malaria. The individuals had variable parasitemia, with a median of 190,000 parasites/µl. The age of the study participants ranged from 6 months to 8 years, with a median age of 3.5 years.

2.3 DNA extraction

Genomic DNA was previously extracted from heparinized venous blood samples of each individual using the DNA blood mini kit (QIAGEN). Briefly, the extraction process was performed as follows: The cellular components of blood were lysed using proteinase K. DNA was then tapped in the spin column matrix (QIAamp silica-gel membrane) by centrifugation. PCR inhibitors were removed in two wash steps using salted buffer followed by elution of DNA by water or TE buffer. The eluted DNA was diluted by 1:10 in TE buffer then stored at -20^{o} C.

2.4 Amplification of MSP1, EBL1 and RH4 genes

Using previously designed oligonucleotide primers (Table 2.1), Expand High Fidelity PCR system from Roche was used to amplify the regions of interest in the genes. The high fidelity Taq polymerase enzyme has polymerization and exonuclease activity which confer the ability to proofread removing errors that might have occurred during polymerization hence the name high fidelity. This means that there was high accuracy of adding the right nucleotide during elongation. Polymerase chain reaction (PCR) was performed in a final volume of 10µl.

Table 2.1. Amplification and sequencing primers

Gene	Region	Primer ID	Primer sequence
EBL1	Region I	OF	**ATGAATGTACCCCTGAATATAAAGTTCCTT**
		OR	**GTATTCGTCTTATTGGGGC**
	Region II	F	**GATGTGGATATATCCATAAAGG**
		R	**CCAGAACTATTAGGAGTATTC**
		F1	AGGAGCATGAAGAATCAA
		R1	CACATACACCTTCAGGT
MSP1	19 kDa	F1	**CAATGCGTAAAAAAACAATGTCC**
		R1	**TTAGAGGAACTGCAGAAAATACCA**
RH4	RH4.9	F3	**TCAAATAATGCTGTTAATGC**
		R1	**GTAAAATTTAGGTATGTATGTTTG**
		F4	CAAATTAAACGAAATCATAC

Sequences in bold were used for both PCR and sequencing while the other sequences indicate primers used for only sequencing.

Following a thawing process, aliquots of reagents used for PCR were vortexed and centrifuged prior to use to ensure homogenization. PCR reactions were prepared in clean and DNAse free conditions in the following steps: Two mixes were prepared in sterile 1.5 ml eppendorf tubes, one for the enzyme, its buffer and water, the other for primers, dNTPs, buffer and water. The use of two tubes instead of one, intended to avoid the degradation effect of Taq polymerase nuclease activity on primers before the start of the reaction. For every batch, mixes were prepared for multiple reactions and the volumes prepared included positive and negative controls and an extra volume of 1 microliter per batch of 10 samples as a consideration of the amount lost while pipetting.

For each PCR reaction, the first reaction mix (mix one) contained 1.0µl of 25mM MgCl 2 stock solution, 0.2µl of dNTP mix (10mM of dGTP, dATP, dTTP and dCTP), 0.3µl of both forward and reverse primers (10uM) and enough quantity of double distilled water (Sigma, UK) toping up to 4.5µl. The second reaction mix (mix 2) contained 0.13µl of Expand High Fidelity Enzyme solution, 1µl of Expand High Fidelity Buffer (10X) with 15mM $MgCl^2$ and 3.87µl of double distilled water (Sigma, UK) which made a total of 5 µl. The initial volume of template used was 0.5µl. For samples that failed to amplify, the volumes were increased to 0.7 or 1µl. The volume of double distilled water in the first reaction mix was adjusted accordingly to keep a final volume of 10µl.

Before running the samples, a PCR optimization process was performed by gradient polymerase chain reactions for all the genes. This aimed at optimizing conditions for each amplification primer set in order to achieve a single and clean band characterizing good yield of PCR products. Each primer set had its specific annealing temperature (Table 2.2). The amplification protocol for all the genes was as follows: DNA was initially denatured at 94^o C for 2 minutes, followed by 9 cycles of denaturation at 94o C for 15 seconds, annealing at temperatures varying between 43- 50^o C for 30 seconds depending on the region being amplified (Table 2.2), and then an extension step at 72^o C.

The round of 9 cycles was followed by other 25 cycles with a gradual increase of elongation time by 5 seconds at each successive cycle. This time increment was necessary due to the reduction in Taq polymerase activity speed due to prolonged exposure on varied temperature.

Table 2.2. Genes sequenced and amplification conditions

Gene	Gene ID (PlasmoDB)	chromosome	Region of interest	PCR Anneal temp	Length (bp)
EBL1	PF3D7_1371600	13	Region I	44	756
			Region II	45	1994
MSP1	PF3D7_0930300	9	$MSP1_{19}$	43	272
RH4	PF3D7_0424200	4	RH4.9	50	842

Gene ID numbers refer to PlasmoDB online database of the Plasmodium genome

2.5 Gel electrophoresis

Gel electrophoresis was performed to check the quality of amplification. The gels were prepared using 1X Tris-borate- EDTA (TBE) buffer at pH 8.0. This buffer was made of Tris base, boric acid and Ethylenediamine tetra-acetic acid (EDTA). A solution of 0.5 M EDTA with pH 8.0 required for the preparation of this buffer was prepared earlier on as follows. A stock solution of 500ml was made by dissolving 93.05 g of EDTA disodium salt in 400ml of deionized water on a magnetic stirrer. A pH of 8.0 was created by slowly adding NaOH to the solution at the same time checking the pH with a pH meter. The solution was then topped up to a volume of 500ml by adding more double distilled water.

A volume of one liter of TBE buffer was prepared in a sterile glass bottle for a concentration of 10X. A stock solution of 10X TBE was prepared as follows: 108g of Tris base and 55g of boric acid were weighed and dissolved in 600ml of deionized water using a magnetic stirrer. The amount of 40ml of the 0.5M EDTA prepared earlier on was added and the final volume was brought to 1 liter by adding deionized water.

The working solution was made from the stock by a dilution of 10 times using deionized water which gave a concentration of 1X. The agarose gel of 1% concentration was made by dissolving 1g of agarose powder (from Applied Genetic Technologies Corporation) in 100ml of 1X TBE buffer. The prepared gel was boiled in a microwave for 90 seconds then left standing to cool. Before the gel set, 5µl of ethidium bromide was added and the conical flask swirled to ensure even mixing. The gel was allowed to set on a gel tray having a number of interspaced combs, which helped in forming wells for loading samples.

The gel was submerged in running TBE buffer (1X) in a gel electrophoresis tank, the PCR products (1µl for each gene fragment) were mixed with 2µl of the loading dye (6X Blue Orange from Promega) then loaded on the gel. 1kb DNA ladder (HyperLadder 1, Bioline UK) diluted at a half was used for estimating the size of PCR products. The ladder was used as the standard DNA marker. Electrophoresis was done for a period of 40 minutes at constant voltage (90 volts). The gels were viewed by digital photography under UV light using The Molecular Imager Gel Doc (Bio-Rad., UK). Only the samples containing a single band were considered for sequencing. DNA amplification process was repeated for samples of which there were no DNA bands in the gel.

2.6 Purification of the PCR products

Purification of PCR products was performed to remove remnants of primers, dNTPs and buffers that would interfere in the sequencing process. This was done using Ethanol/Sodium Acetate precipitation in 96 well plates. Purification buffer constituted of 3µl of Sodium Acetate, pH 5.2, 62.5µl of 95% ethanol and 24.5µl of distilled water were mixed to final volume of 90µl per sample. 90µl of the purification buffer was added to each well containing the PCR products.

The plates were then sealed with micro-seals (Bio-Rad) and incubated at -20^{o} C for 30 minutes. After incubation, the plates were centrifuged at 3000xg for 30 minutes at 4^{o} C on a centrifuge (5810R bench centrifuge, Eppendorf). Seals were removed and plates overlaid with absorbent paper towels then gently inverted. The inverted plates were centrifuged at 50xg for 1 minute at 4^{o} C. A volume of 150µL of ice cold (-20^{o} C) 70% ethanol was added into each well, then the plate was sealed and spun at 3000xg for 10 minutes at 4^{o} C. After this the plates were once again inverted over paper towels and excess fluid gently drained. The plates were again overlaid with clean paper towels, inverted and spun at 50xg for 1 minute at 4^{o} C. Following the washing steps, the plates were covered with fresh paper towels and left on the bench to air dry. Once the plates were completely dry, DNA was reconstituted using 20 µl of Elution buffer (TE Buffer) per well.

2.7 Big Dye Terminator Sequencing

Prior to Sanger sequencing by capillary electrophoresis, a sequencing PCR process was performed using fluorescent Big Dye terminators (BDT mix v3.1) to amplify the purified amplicons of the gene fragments of interest. This process is comparable to a PCR given that template DNA is reproduced to generate new strands starting at the site of the annealed primer. It differs from the conventional PCR in that only one primer is used instead of two and that apart from the typical dNTPs, there are other four dye-labelled dideoxynucleotides (ddNTPs). Once a ddNTP is integrated into the growing strand of DNA, synthesis stops due to the lack of a hydroxyl group that is required for two nucleotides to form a polymer. The ultimate products of this reaction consist of a set of different lengths of DNA fragments that are fluorescently labeled at the 3' end.

PCR prior to sequencing was carried out in 96 well plates. The master mix for one sequencing reaction was prepared by mixing the following reagents in a sterile eppendorf tube: 0.5µl of Big Dye Terminator (BDT) ready reaction mix v3.1, 1.75µl of 5X sequencing buffer, 1µl of primer (10µM) and PCR clean water (Sigma, UK). The volume of water and PCR product were adjusted accordingly to give a final volume of 10µl per reaction. For samples that had strong bands, 3µl of purified PCR product was used. The volume was increased to 4µl for samples that had faint bands.

The whole process was performed under minimal light due to sensitivity of the fluorescently labeled ddNTPs to light. Intense light may reduce coloring faculty of the fluorescent dyes.

Following preparation and tight covering by micro seals, plates were loaded onto a thermocycler. The cycling profile for the sequencing reaction was a follows: 25 cycles of denaturation at 96^{o} C for 30 seconds, annealing at 50^{o} C for 15 seconds and extension at 60^{o} C for 4 minutes, with a ramp rate of 1^{o} C per second between the different temperatures.

2.8 Big Dye PCR purification using ethanol/sodium acetate mixture

Purification of the products for the sequencing reaction was performed using ethanol/sodium Acetate precipitation in 96 well plates. This step was necessary to remove excess primers and ddNTPs and perfomed as described in section 2.6 apart from the reconstituion step which was done using 10µl of Hi-Di formamide reagent following plates air-drying.

2.9 Capillary electrophoresis

Following the reconstitution by Hi-Di formamide, capillary electrophoresis was performed in an automated 3130xl sequencer (Applied Biosystems, UK). The sequencer was accurate enough to separate DNA fragments that differ by just one base pair. Each of the four ddNTPs had a special fluorescent dye of a different colour attached to it. These dyes gave light at a different wavelength when excited using a laser beam. The resulting fluorescence was picked out by a charge-coupled device (CCD) camera and converted into a chromatogram. As the fluorescently labeled extension products from the sequencing reaction migrated through the polymer passed the laser detector, each base was detected as a colour signal.

2.10 Sequence editing, assembly and alignments

The sequencing files of the three genes referred to as the trace files generated by the 3130xl sequencer from Applied Biosystems were imported into Seqman Pro application (DNASTAR Lasergene Suite, Version 12) for analysis. For each sample, quality values (QV) for sequences were determined. The QV is a per-base estimate of the base caller accuracy. Per-base QVs are calibrated on a scale corresponding to: QV = -10 log 10 (Pe) where Pe is the probability of error (Cock et al. 2009). Sequences of good quality (QV> 25) were aligned into contigs and each primer trace file assessed for the quality of peaks and base calling. Reference sequences were used to scaffold the trace data generated from each primer for each gene. Corrections of bases called by the sequencer were done on the basis of the peaks of the electropherogram and independent of the reference sequence. Clean scaffolds were saved as consensus files.

A multiple alignment of the consensus files for all samples of each gene was carried out in the MegAlign application (DNASTAR, Lasergene Suite), Bioedit (Thomas A. Hall 1999) or MEGA (Tamura et al. 2013) using Clustal W method and the alignments were confirmed visually. Mis-aligned sequences were corrected manually. Multiple alignment was done to identify polymorphic positions in the samples for each gene under analysis. The alignments were saved in FASTA or PAUP format then imported into DnaSP V 5.10.01 software for statistical analysis.

2.11. Data analysis

2.11.1. Testing neutrality

Testing deviation from neutrality was performed basing on allele frequency distribution indices (Tajima's D and Fu and Li's D and F) computed using DnaSP5.10 software (Rozas et al. 2003). Tajima's D is based on the differences between the number of segregating sites and the average number of nucleotide pairwise differences. Fu and Li's D test statistic is based on the differences between the number of singletons (mutations appearing only once among the sequences, *hs*), and the total number of mutations, *h* (Fu and Li 1993). Fu and Li's F test statistic is based on the differences between the number of singletons and the average number of nucleotide pairwise differences between pairs of sequences, *k* (Fu and Li 1993). Sliding window analysis of genes that were significantly under selection was performed to show the regions of the genes which depart from neutrality and the regions that are neutral and randomly evolving.

2.11.2. Testing Linkage Disequilibrium

Linkage disequilibrium (LD) is defined as nonrandom association between nucleotide variants at different polymorphic sites in each gene region, and was studied using DnaSP5.10. The r2 (square of the correlation coefficient of allelic states at each pair of SNPs), D and D' indices were computed for significant association between different variable sites. Fisher's exact test was computed for significance. This analysis completely excluded sites containing alignment gaps, or polymorphic sites segregating for one nucleotide referred to as singletons.

2.11.3. Comparing genetic variation patterns between ligand and receptor data

Ligand and receptor neutrality tests were computed and compared. The erythrocyte receptor data was obtained from previous work done at CEBIB (C. Awuor thesis, 2014; I. Omedo thesis, 2013). The receptor data were previously collected from the same individuals.

CHAPTER 3

RESULTS

3.1. Gene amplification

This study was conducted using 93 DNA samples. Of all the 93, 61 samples for MSP1, 68 samples for EBL1 region I, 71 for EBL1 region II and 56 for RH4 were successfully amplified by PCR with the expected PCR product sizes. The amplicons of approximately 270 bp, 756 bp, 2 kb, and 842 bp were obtained for MSP1, EBL1 region I, EBL1 region II and RH4, respectively. The rest of the samples had low template concentrations and were unsuccessfully amplified despite the multiple PCR amplification attempts three to four attempts using varied template volumes.

3.2 Sequence generation

All the samples that were confirmed to have been amplified with a clear and single band as observed on gel pictures were considered for sequencing (Figure 3.1). Of the 61 samples of RH4 gene that were successively amplified, 49 yielded good sequences whereas, of the 56 RH4 samples, 46 generated good sequences. Of 68 EBL1 region I successively amplified, 46 yielded good sequence. Of the 71 EBL1 region II samples amplified, 58 were successfully sequenced. The full length of region II of EBLl gene was only obtained for 39 samples.

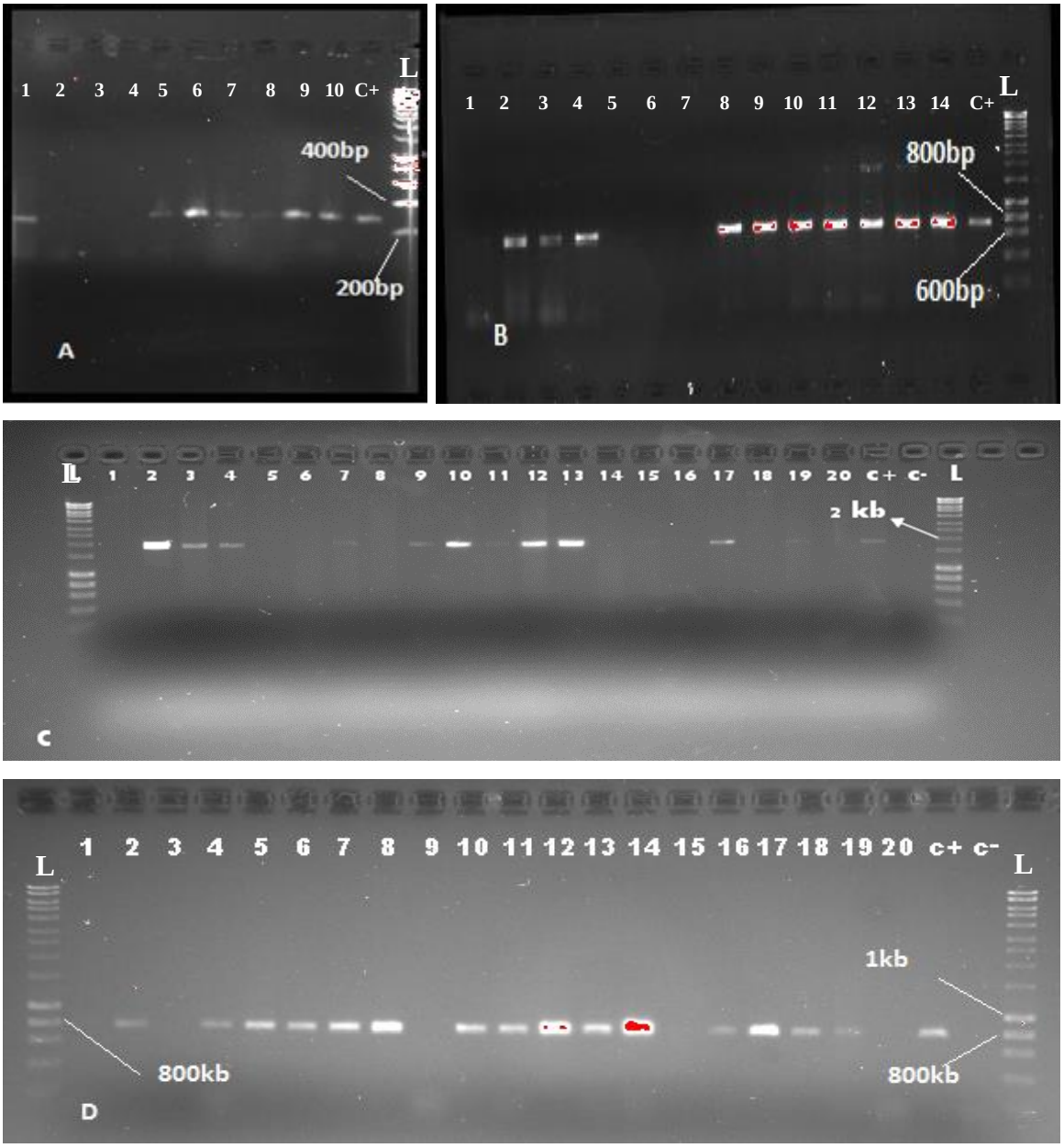


Figure 3.1. Gel pictures of amplification products of the MSP1 19kDa (A), EBL1 region I (B), EBL1 region II (C) and RH4.9 (D) regions. A) Clear DNA bands sized between 250 and 300 kb were obtained for MSP1 19kDa region. B) EBl1 region I produced DNA bands between 600 and 800kb. C) Around 2kb and D) approximately 800 bp for EBL1 region I and RH4.9 region, respectively. Lanes (L) represent Bioline's Hyperladder I used to determine amplicon size. Lanes 1-10 in (A), 1-14 in B, 1-20 in C and 1-20 in D represent samples. Samples with one identification number do not necessarily represent same sample.

3.3. Analysis of polymorphic sites in the genes

This study defined polymorphic or segregating sites as those differing from the reference sequence at specific positions. Positions in the sequences such as SNPs, insertions or deletions were analyzed using MegAlign (Lasergene 12, DNAstar, UK), Bioedit v7.2.5 and DnaSP 5.10. Using the multiple alignment of sequence isolates for each gene, different numbers of SNPs were observed in the genes sequenced. In a total of 46 RH4 isolates, 3 segregating sites were observed along a 798 kb sequence. The 19 kDa (C terminal) region of MSP1 had 6 segragating sitses (SNPs) as observed in the 276 kb amplicon.

It was found that EBL1 is highly polymorphic. Region I of EBL1 gene alone had 13 SNP positions within a length of 590 bp. Insertion of 5 Thymines at position 572 in the EBL1 gene (region I) was detected in 31 (75.6%) of the 41 sequences studied (Figure 3.2). This insertion introduces a stop codon in the downstream regions coding for a nonfunctional protein. Region II of the EBL1 gene had 9 polymorphisms along the size of 1931kb observed in this study.

Table 3.1. Synonymous and non-synonymous variations detected in MSP1-19, RH4.9 EBL1 region I and EBL1 region II

Gene region	Position	Ref. codon	Ref. amino acid	Variant codon	Variant amino acid
MSP1-19	1620	GAA	Glutamate	CAA	Glutamine
	1667	ACA	Threonine	AAA	Lysine
	1675	AGC	Serine	AAC	Asparagine
	1676	AGC	Serine	AAC	Asparagine
	1677	AGA	Arginine	GGA	Glycine
	1692	CTT	Leucine	TTT	Phenylalanine
	1702	AAC	Aspargine	AAA	Lysine
RH4.9	435	AAC	Asparagine	AAA	Lysine
	438	CAA	Glutamine	AAA	Lysine
	500	AAA	Lysine	ATA	Isoleucine
EBL1 region I	32	CAG	Glutamine	CAA	Glutamine
	39	CAA	Glutamine	TAA	STOP
	42	AAG	Lysine	AAA	Lysine
	62	AGA	Arginine	AAA	Lysine
	69	AAG	Lysine	AAA	Lysine
	72	GAA	Glutamate	AAA	Lysine
	73	GAT	Aspartate	AAT	Asparagine
	86	GAA	Glutamate	AAA	Lysine
	87	GAT	Aspartate	AAT	Asparagine
	92	GTG	Valine	GGG	Valine
	179	CTT	Leucine	TTT	Phenylalanine
	180	CGA	Arginine	CGG	Arginine
	195	AAT	Asparagine	AAC	Asparagine
EBL1 region II	341	ATA	Isoleucine	ATG	Methionine
	346	TAT	Tyrosine	TGT	Cysteine
	475	CAT	Histidine	CAC	Histidine
	509	TTA	Leucine	TTG	Leucine
	657	GGT	Glycine	TGT	Cysteine
	663	AAA	Lysine	TAA	STOP
	689	CAC	Histidine	TAC	Tyrosine
	745	AAA	Lysine	AAG	Lysine

Majority of the SNPs identified in these genes were non-synonymous. This means that the substitution of a nucleotide in the DNA led to the change of amino acid in the coded protein. A synonymous SNP means it does not result in a change in the amino acid sequence of the coded protein. Region I of the EBL1 gene has the most polymorphic of all the regions analyzed in this study and it contained 8 non-synonymous SNPs in the 13 SNPs observed. Of the 8 SNPs identified in Region II of this gene, 5 were non-synonymous. Some SNPs led to the introduction of the stop codons within the coded protein (Table 3.1). All the 7 SNPs observed in MSP1-19 and 3 SNPs observed RH4.9 regions were non-synonymous.

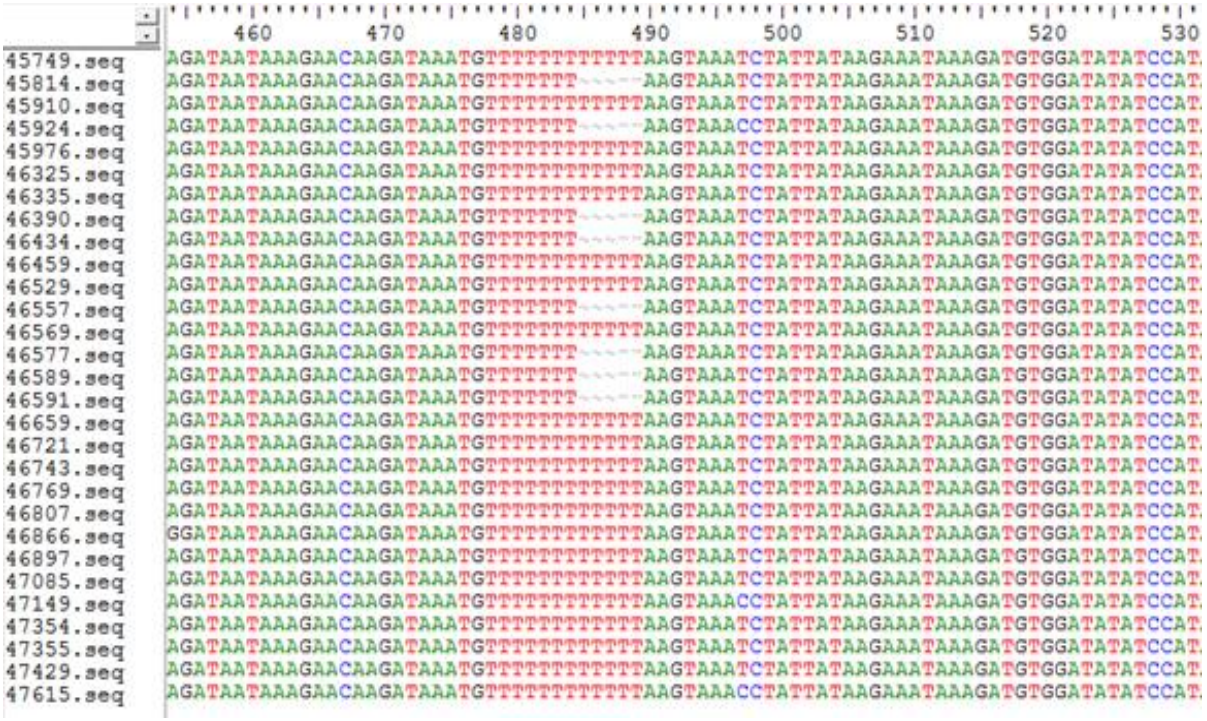


Figure 3.2: 5 Thymidines insertion in Region I of EBL1 gene. This insertion of 5Ts occurred at the position 572. Reference codon is based on 3D7 laboratory strain sequence.

3.3.1 Allele frequencies

Allele frequencies were calculated for every segregating site in the sequenced gene regions. The rates of transversions and transitions were equal in the MSP1-19 region while only transversions were observed for RH4.9 region. Transversion refers to the substitution of a purine (A or G) for a pyrimidine (C or T) or vice versa. Because this type of mutation changes the chemical structure dramatically, the consequences of this change tend to be more drastic than those of transitions. This is because the wobble position of the DNA, which to a large extent is responsible for the degeneracy of the code, is more tolerant of transition than a transversion, such that a transition mutation is more likely to lead to the same amino acid being encoded for or the amino acid of the same properties.

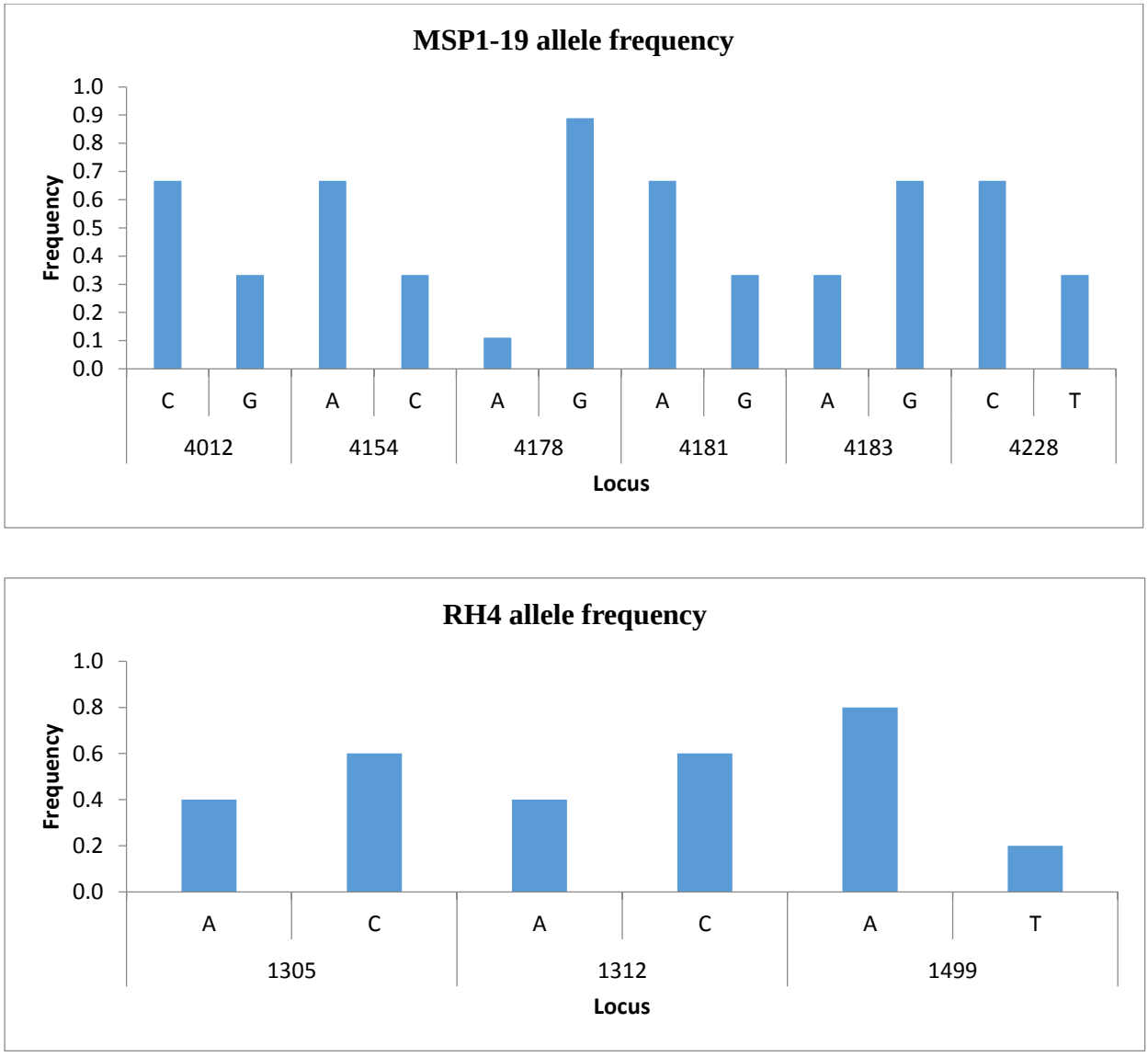


Figure 3.3: Allele frequencies in MSP1-19 and RH4.9 gene regions. The SNP loci were determined basing on Plasmodium falciparum 3D7 strain reference nucleotide sequence.

In the EBL1 gene, majority of the substitutions observed were transitions. In most of the cases A was substituted with G, A being less frequent in region I and more frequent in region II. The region II is known as the binding region and is Cysteine rich. Some SNPs introduced more cysteine amino acids compared to the reference sequence (3D7 strain). Unlike in MSP1 and RH4, in the EBL1 gene, genotypes having G instead of A in the transversion sites (mostly region I) were frequent at the rate higher than 75% (Figure 3.2 and figure 3.3). This means that the alleles involving the nucleotide A occurred in a low frequency (less than 25%).

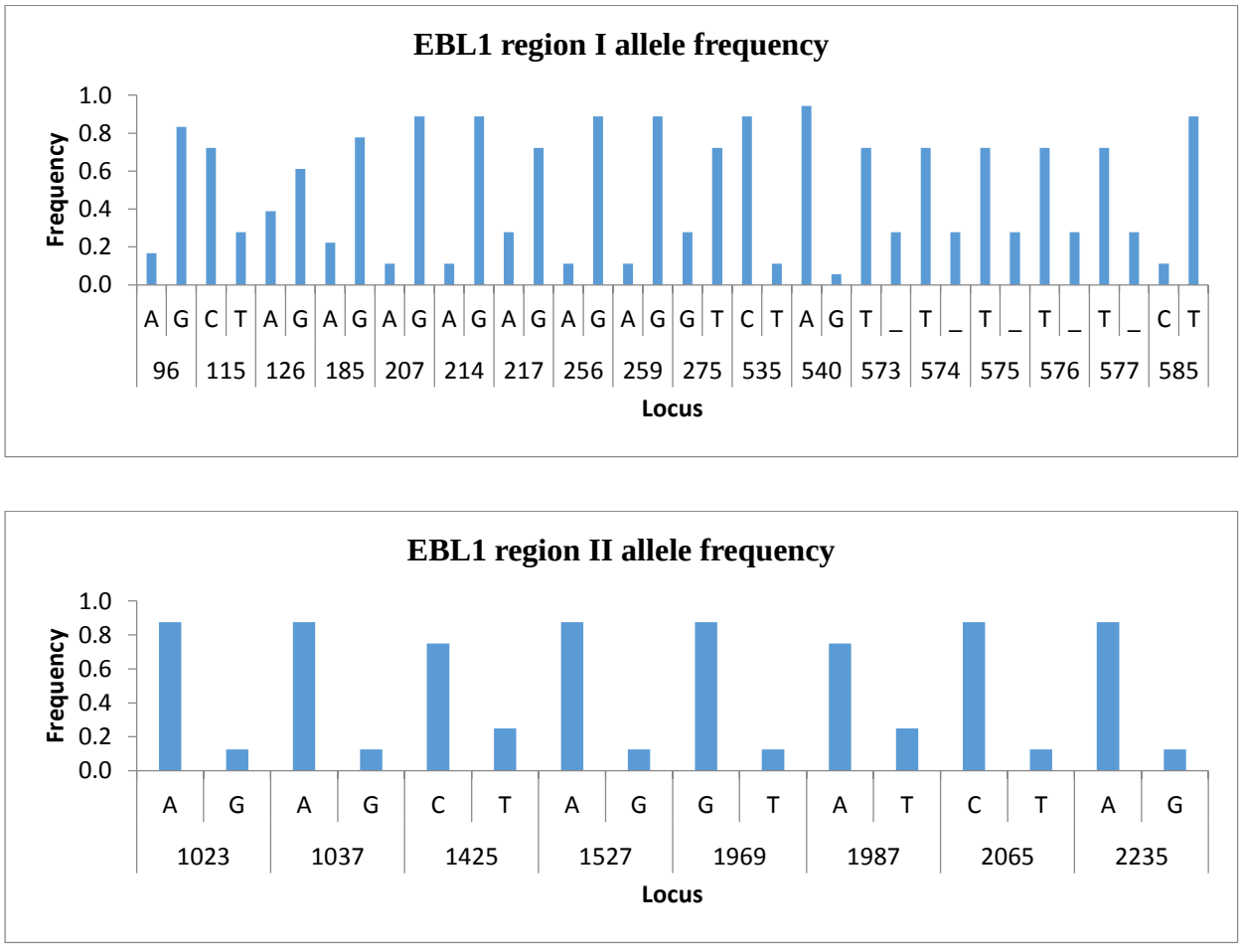


Figure 3.4: Allele frequencies for EBL1 region I and region II. The SNP loci were determined basing on Plasmodium falciparum 3D7 strain reference nucleotide sequences. The insertion sites in region I were considered as polymorphic sites and were presented as _ sign. For region II the positions were determined on the reference sequence that has no 5Ts insertion.

3.3.2. Haplotype frequencies

Different gene haplotypes were observed in our population at different frequencies. For all the 9 haplotypes observed in MSP1-19 region, only 2 haplotypes (Table 3.2) were predominantly circulating in the population at the rate of more than 20% each (Figure 3.4 A). MSP1 Haplotype 2 and 3 (Table 3.2) had similar amino acids (QKSNGLK). They only differed at codon position 1702 and this difference does not make a change in the coded amino acid. The haplotype QKSNGLK had a frequency of 34.7% and is much similar to haplotype EKSNGLK which is also predominantly (44.9%) circulating in the study population. The difference is only at one amino acid at position 1620, one having glutamine (Q) and the other, glutamate (E) (Table 3.2). Both haplotypes are predominant at the higher rate than 78% combined (Table3.2).

Table 3.2. MSP1 haplotypes based on 19 kDa region

H.c	Count (%)	Genotype	Haplotype						
		Codon position:	1620	1667	1675	1676	1677	1692	1702
1	2 (4.1)	CAAAGC	Q	K	N	N	G	L	K
2	10 (20.4)	CAGAGC	Q	K	S	N	G	L	K
3	7 (14.3)	CAGAGT	Q	K	S	N	G	L	K
4	1 (2.0)	CAGGGT	Q	K	S	S	G	F	K
5	1 (2.0)	CCGAAC	Q	T	S	N	R	L	K
6	1 (2.0)	CCGAGT	Q	T	S	N	G	F	K
7	22 (44.9)	GAGAGC	E	K	S	N	G	L	K
8	2 (4.1)	GAGGAC	E	K	S	S	R	L	K
9	3 (6.1)	GCGGAC	E	T	S	S	R	L	K

H.c: haplotype code; No.loci: number of loci; the total number of sequences considered is 49; Count: the absolute number per haplotype; Shaded residues represent the same haplotype but different nucleotide sequences.

In RH4 (based on region RH4.9), 5 haplotypes were identified (Table 3.3) of which 2 (NKK and NQK) were the most frequent at a rate higher than 89% combined.

Table 3.3. RH4 gene haplotypes based on RH4.9 region

H.c	Count (%)	Genotype	Haplotype		
		Codon position:	435	438	500
1	1 (2.2)	AAA	K	K	K
2	3 (6.5)	ACA	K	Q	K
3	25 (54.3)	CAA	N	K	K
4	16 (34.8)	CCA	N	Q	K
5	1 (2.2)	CCT	N	Q	I

H.c: haplotype code; No.loci: number of loci; the total number of sequences considered is 46; Count: the absolute number per haplotype.

To determine the haplotypes corresponding to region I of EBL1 gene we excluded %Ts insertion and only considered SNPs because the insertion causes open reading frame shift which introduces a stop codon meaning that the gene becomes a pseudogene. In region I, 14 haplotypes were observed in 16 genotypes (Table 3.4). 1 haplotype (QQKRKEDEDVLRN) was more predominant (44.5 %) but only present in 3 genotypes.

Table 3.4. EBL1 haplotypes based on region I

H.c	Count (%)	Genotype	Haplotype											
		Codon position:	32	39	42	62	69	72	73	86	87	92	179	180	195
1	1 (3.4)	ACAAGAAAGTCAT	Q	Q	K	K	K	K	N	K	D	V	L	R	N
2	1 (3.4)	ACAAGGGGGGCAT	Q	Q	K	K	K	E	D	E	D	G	L	R	N
3	1 (3.4)	ACAAGGGGGTCAT	Q	Q	K	K	K	E	D	E	D	V	L	R	N
4	1 (3.4)	GCAGGGAGGGCAT	Q	Q	K	R	K	E	N	E	D	G	L	R	N
5	1 (3.4)	GCAGGGAGGTCAT	Q	Q	K	R	K	E	N	E	D	V	L	R	N
6	1 (3.4)	GCAGGGGGGTCAT	Q	Q	K	R	K	E	D	E	D	V	L	R	N
7	1 (3.4)	GCGAAGGGAGCAT	Q	Q	K	K	K	E	D	E	N	G	L	R	N
8	1 (3.4)	GCGGAGGGGTCAT	Q	Q	K	R	K	E	D	E	D	V	L	R	N
9	1 (3.4)	GCGGGAGAAGCAT	Q	Q	K	R	K	K	D	K	N	G	L	R	N
10	11 (37.9)	GCGGGGGGGTCAT	Q	Q	K	R	K	E	D	E	D	V	L	R	N
11	1 (3.4)	GCGGGGGGGTTAT	Q	Q	K	R	K	E	D	E	D	V	F	R	N
12	1 (3.4)	GCGGGGGGGTTGT	Q	Q	K	R	K	E	D	E	D	V	F	R	N
13	1 (3.4)	GTAGGGAGGGCAC	Q	*	K	R	K	E	N	E	D	G	L	R	N
14	1 (3.4)	GTGGGGAGGTCAT	Q	*	K	R	K	E	N	E	D	V	L	R	N
15	2 (6.9)	GTGGGGGGGTCAC	Q	*	K	R	K	E	D	E	D	V	L	R	N
16	3 (10.3)	GTGGGGGGGTCAT	Q	*	K	R	K	E	D	E	D	V	L	R	N

*H.c: haplotype code; No.loci: number of loci;The total number of sequences considered is 29; Count: the absolute number per haplotype; Shaded residues represent the same haplotype, but different nucleotide sequences; *STOP codon*

Interestingly, that was the case in the receptor binding domain (region II) of EBL1 gene. Only 1 (IYHLGKHK) of the 8 haplotypes observed in region II was predominantly frequent (83.8%) also in 3 different genotypes (Table 3.5). Most of the samples (3/4) that were confirmed not to have the 5Ts insertion at position 572 of EBL1 region I were proven to have this region II haplotype.

Table 3.5. EBL1 haplotypes based on the receptor binding domain (region II)

H.c	Count (%)	Genotype	Phenotype							
		Codon position:	341	346	475	509	657	663	689	745
1	27 (73.0)	AACAGACA	I	Y	H	L	G	K	H	K
2	1 (2.7)	AACAGTCG	I	Y	H	L	G	*	H	K
3	1 (2.7)	AACATATA	I	Y	H	L	C	K	Y	K
4	1 (2.7)	AACGGACA	I	Y	H	L	G	K	H	K
5	3 (8.1)	AATAGACA	I	Y	H	L	G	K	H	K
6	1 (2.7)	AATAGTCA	I	Y	H	L	G	*	H	K
7	2 (5.4)	AGCAGACA	I	C	H	L	G	K	H	K
8	1 (2.7)	GACAGACA	M	Y	H	L	G	K	H	K

*H.c: haplotype code; Count: the absolute number per haplotype (total number = 37); Shaded residues represent the same haplotype but different nucleotide sequences; *STOP codon*

These results show that in most of the cases, only one haplotype was predominantly circulating in the population of our study (Figure 3.4) regardless of the number of haplotypes in total. In the case where more than one haplotype are predominantly circulating, the difference between them is mostly at only one SNP (Table 3.5).

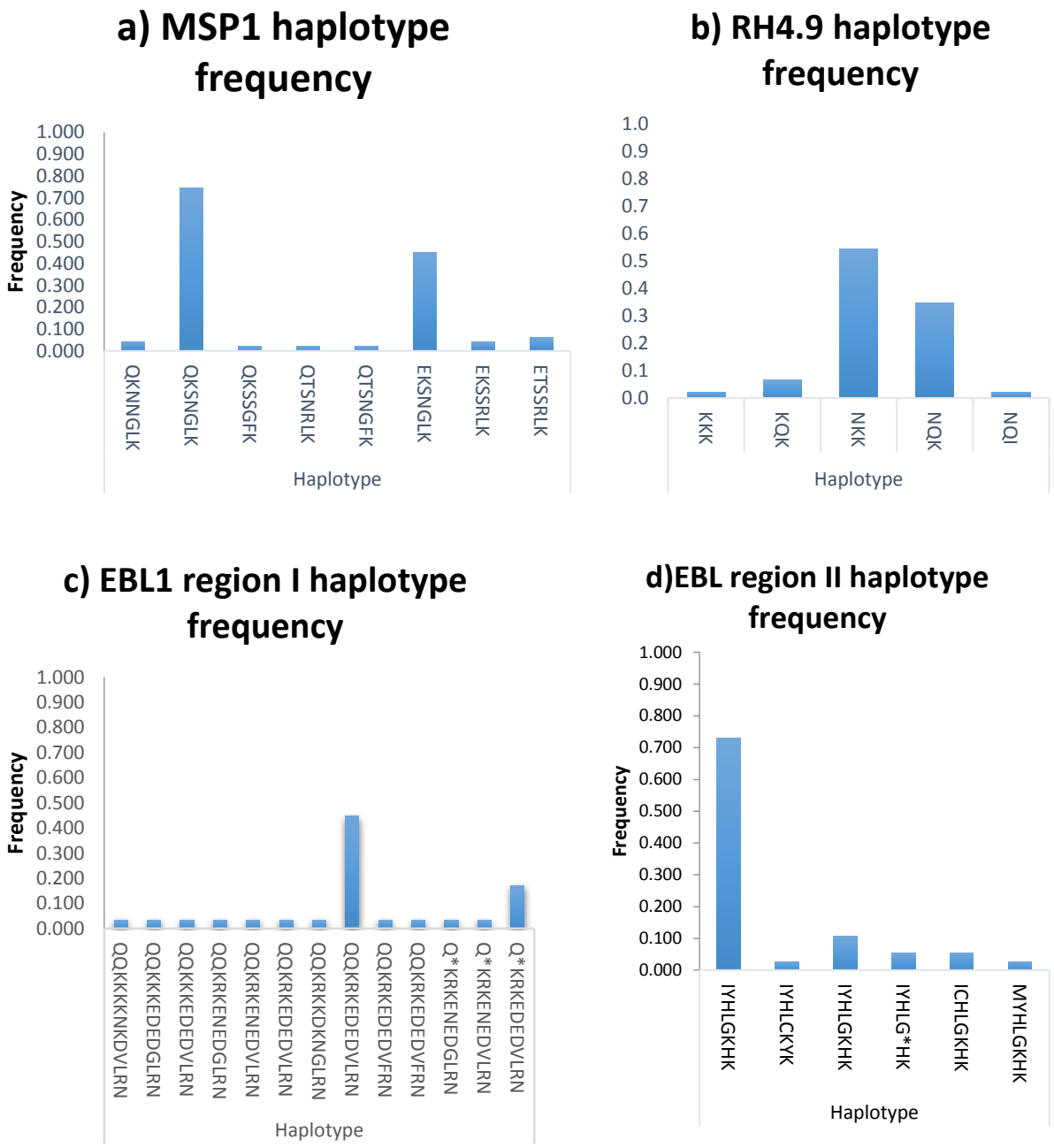


Figure 3.5: Haplotype frequencies for MSP1-19, RH4.9, EBL1 Region I and EBL1 region II. * Sign represent an introduction of a stop codon

3.4 Tests of neutrality

Departures from neutrality was tested based on the allele frequency indices such as Tajima's D, Fu and Li's D and F. MSP1 and RH4 did not show departure from neutrality which implied that these two genes are not under selection. The Tajima's D values were 0.32552 ($p>0.05$) and 0.07905 ($p>0.05$) for MSP1 and RH4, respectively. Tests showed that EBL1 is under purifying selection especially Region II which had a Tajima's D of -1.87248 ($P<0.05$). Despite the big number of segregating sites (S=13), region I of EBL1 gene was not proven to be under significant selection (Tajima's D = -0.59883; $p>0.05$). Detailed figures are displayed in table 3.2.

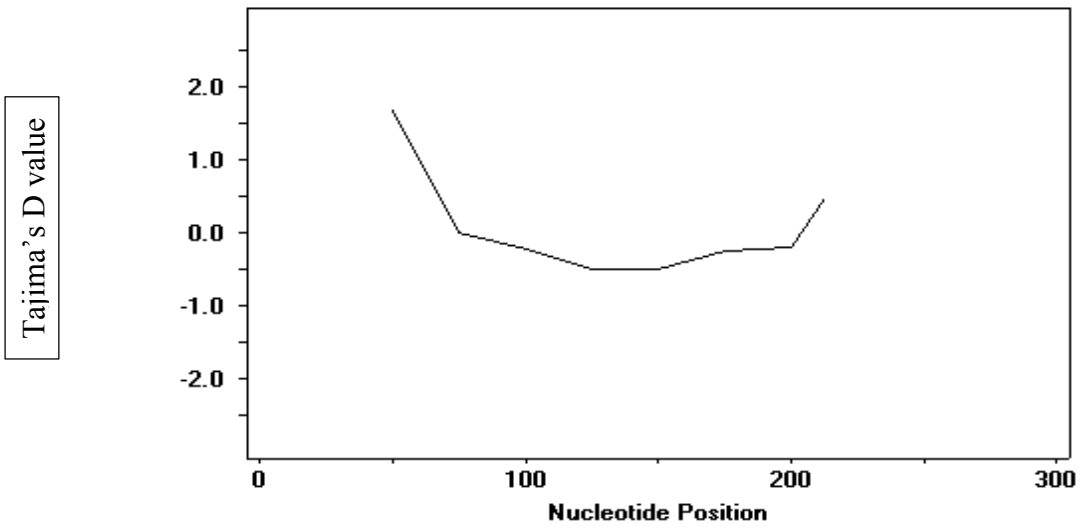
Table 3.6. Tests of Neutrality for MSP1, EBL1 and RH4 genes.

Gene	Region	n	size (bp)	S	Π	k	Tajima's D	Fu & Li D	Fu & Li F
MSP1	**19kDa**	49	276	6	0.00607	1.51701	0.32552	1.17625	1.06447
EBL1	**Region I**	29	590	13	0.00462	2.71429	-0.59883	1.05915	0.63827
	Region II	37	1712	8	0.0004	0.67868	-1.87248*	-2.02271*	-2.31445
RH4	**RH4.9**	46	798	3	0.00113	0.70821	0.07905	-0.39002	-0.29041

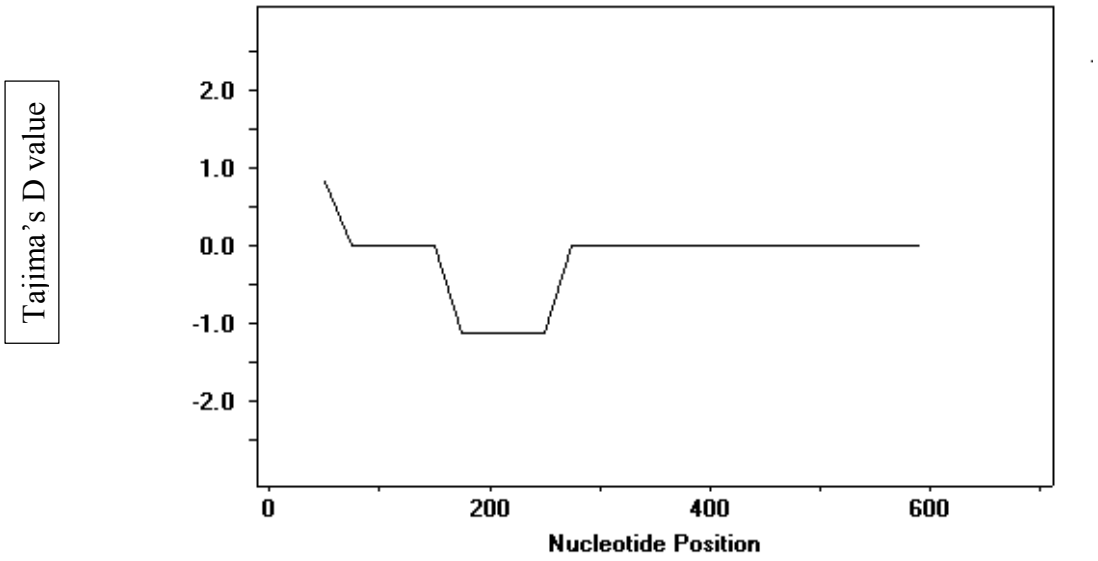
*The following abbreviations are used: π, Nucleotide diversity; k, Average number of nucleotide differences; S, number of polymorphic (segregating) sites. For Tajima's D, *, p<0.05; for Fu and Li's analysis, *, p < 0.05. These values were calculated using DnaSP 5.10.*

The same neutrality tests were computed for all regions using a sliding window of 100 sites long, with a step size of 25 bases. This analysis aimed at finding the region within the sequence that are responsible for any significant departure from neutrality. Subsequent results were presented in form of graphs using DnaSP 5.10 software (Figure 3.5).

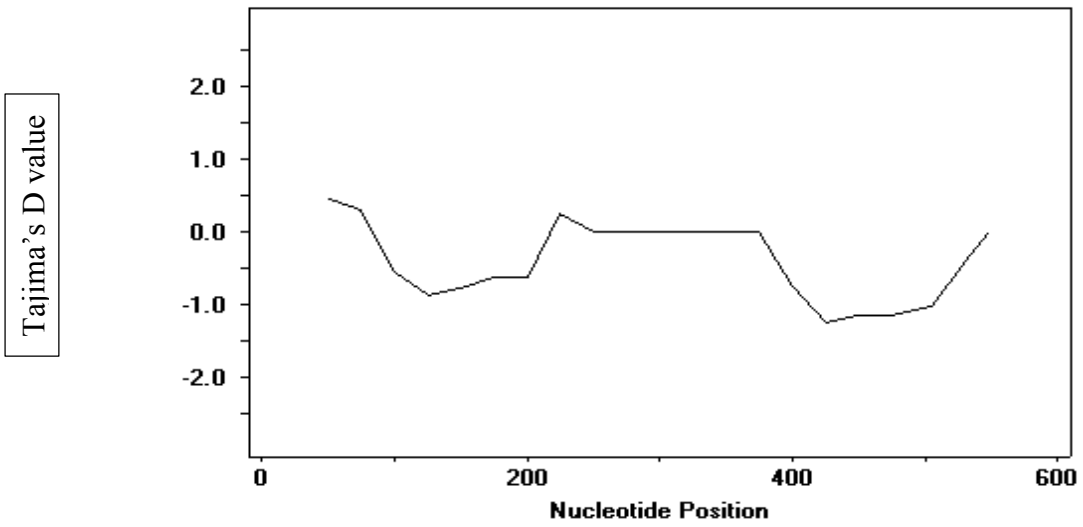
Different regions of EBL1 and RH4 showed to be under purifying selection, although there was no statistical significance for RH4 ($p>0.05$). No region of MSP1 showed to be under selection.



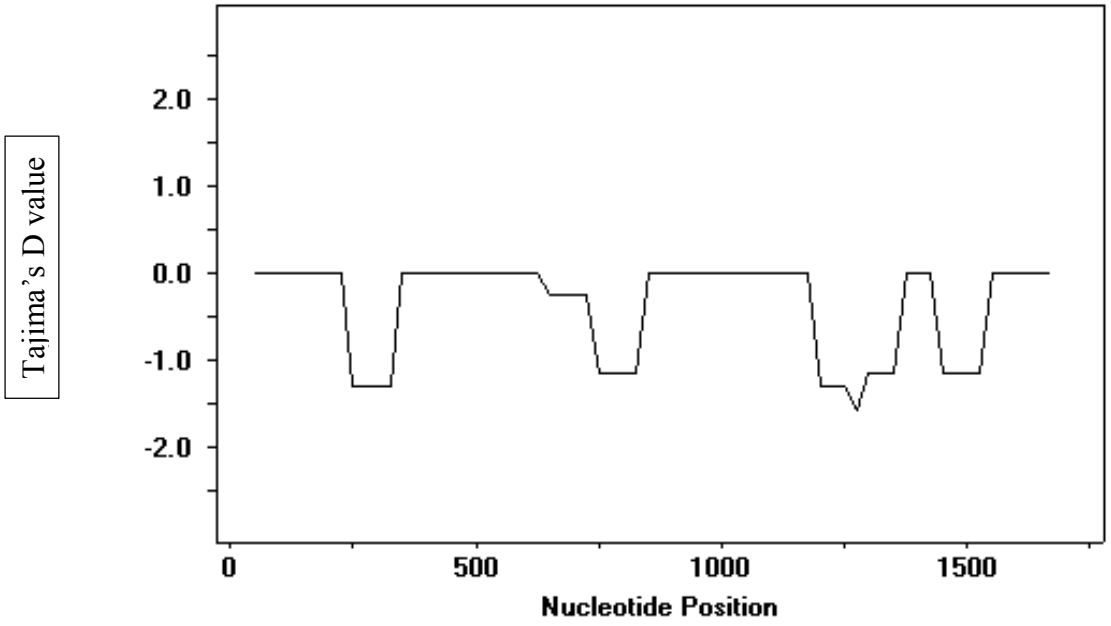
A) MSP1



B) RH4



C) EBL1 region I

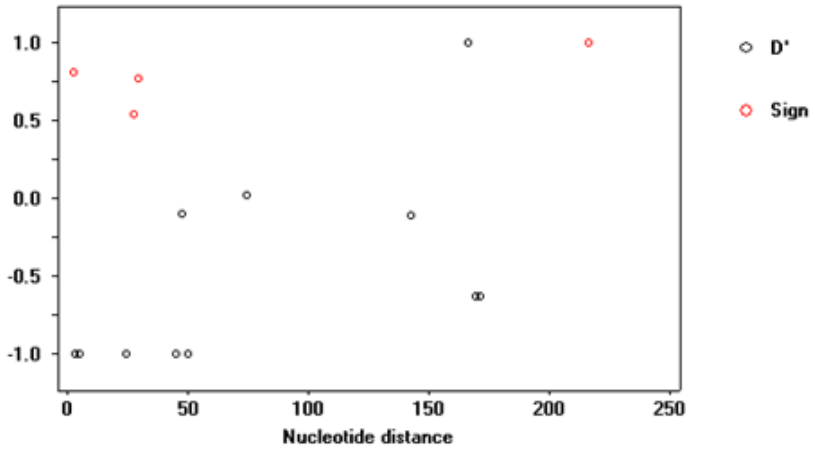


D) EBL1 region I I

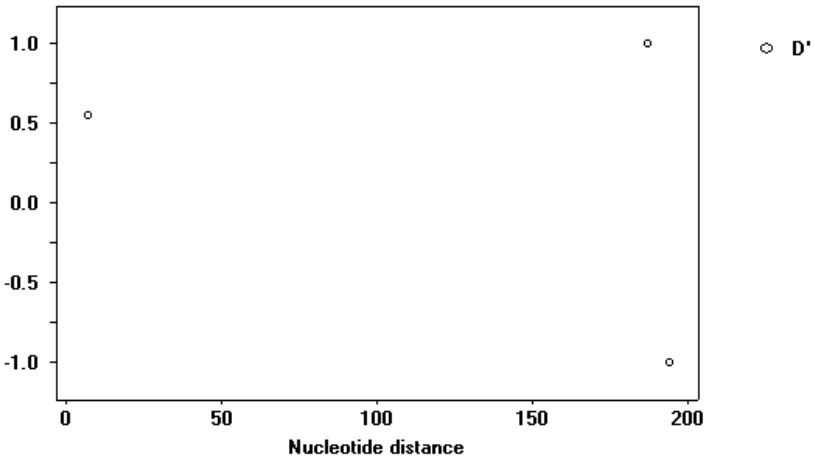
Figure 3.6: Tajima's D graph for MSP1 (A), RH4 (B), EBL1 region I (C) and region II (D) by sliding window analysis. The graphs were constructed in DnaSP software using a sliding window 100 sites long and a step size of 25 sites. The graph shows the value of D for each region of the sequenced gene. Only EBL1 region II had shown a significant departure from neutrality.

3.5 Linkage disequilibrium

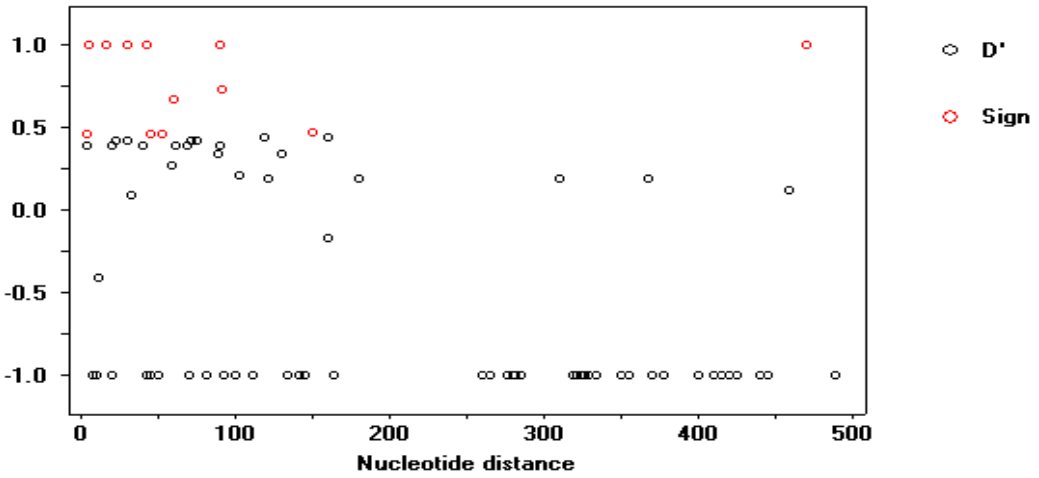
Linkage Disequilibrium (LD) was computed based on D, D' and R2 indices in the DnaSP software to study if there is nonrandom association between SNPs detected (Figure 3.6). Pairwise comparisons were computed between SNPs with significance of pairwise comparisons being determined by Fisher's exact and Chi square tests. This means that at every segregating site, the frequency of 2 nucleotide being associated in the same haplotype was studied.



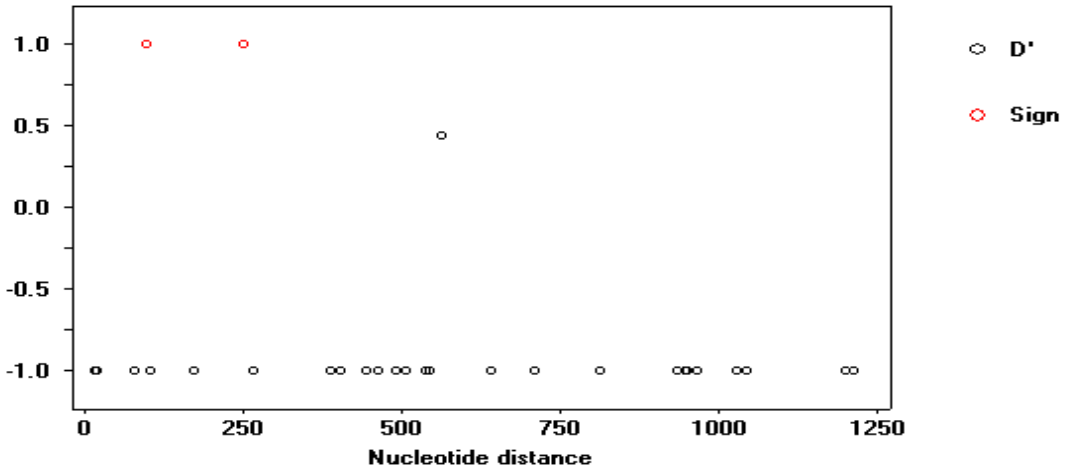
A) MSP1-19



B) RH4



C) EBL1 region I



D) EBL1 region II

Figure 3.7: Linkage disequilibrium index D' plots for MSP1-19, RH4.9, EBL1 region I, EBL1 region II. Each black circle represents the difference between observed and expected haplotypes frequencies. Sign (red circles), represents statistically significant differences as determined by the two-tailed Fisher's test.

For the MSP1-19 region, 15 pairwise comparisons were computed, of which 4 were significant based on both Fisher's exact and Chi square tests, suggesting that nucleotides in each of the 4 pairs studied at the segregating sites are not randomly associated. The MSP1 positions that were in LD are 4858 and 5074, 5000 and 5027, 5000 and 5029 and 5027 and 5029 based on the 3D7 reference sequence. For EBL1 region I, 7 of the 78 pairwise comparisons computed by Fisher's exact test were significant. The 7 comparisons were 96-126, 96-185, 115-585, 126-185, 126-217, 214-256, and 259-275. The significance was raised to 12 by computation with Chi square test. Nonrandom association was significant for only 2 (1969-2065 and 1987-2235) among the 28 pairwise comparisons computed for SNPs found in EBL1 region II.

3.6 Comparison of genetic patterns between ligands and receptors

Ligand and receptor neutrality tests results were compared. Exon 4 of the Glycophorin B gene showed evidence of being under purifying selection (Table 3.7). This conforms to the results of its ligand, EBL1 based on its receptor binding domain (Table 3.6).

Table 3.7. Tests of Neutrality for Band 3, Glycophorin B and Complement Receptor 1 genes.

Gene	Region	size (bp)	S	π	Tajima's D	Fu & Li D	Fu & Li F
Band3	Exon 17	197	1	0.00027	-0.75057	0.46966	0.1125
	Exon 18	253	3	0.00159	-0.43021	-0.6657	-0.69513
Glycophorin B	Exon 2		9	0.00807	-1.54	-0.004	-0.64
	Exon 3		2	0.00833	-0.43	-0.93	-0.91
	Exon 4		7	0.00451	-1.81*	-2.82*	-2.93*
CR1	Exon 4 +	184	1	0.00132	0.37574	0.46895	0.51576
	Exon 5	387	4	0.00039	-1.43282	-0.37596	-0.85753

The following abbreviations are used: π, Nucleotide diversity; k, Average number of nucleotide differences; S, number of polymorphic (segregating) sites. For Tajima's D, *, $p<0.05$; for Fu and Li's analysis, *, $p < 0.05$. These values were calculated using DnaSP 5.10.

CHAPTER 4

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

4.1 Discussion

The main aim of this study was to identify *P. falciparum* merozoite EBL1, RH4, and MSP1 polymorphisms and compare their neutrality test scores with those of their respective erythrocyte receptors in order to obtain an insight into the genetic impact they may have on each other in a severe malaria population in Kilifi, Kenya. Collection of *P. falciparum* sequence data enabled us to characterize the parasite ligands, based mainly on their binding regions. We were able to identify polymorphisms in the ligands and determine the haplotypes circulating in the severe malaria population, and their selection signatures. Data from ligands and receptors were both collected from the same individuals which provided a real situation to study the genetic basis of ligand-receptor interaction in the invasion of erythrocytes by *P. falciparum* merozoites.

To characterize the MSP1 gene, this study focused on the $MSP1_{19}$ fragment, the C-terminal domain of $MSP1_{42}$, because it is the binding site and is carried into the newly invaded RBCs while other MSP1 fragments are shed off after proteolytic maturation. The 6 SNPs identified in this region (aa1620 to 1702) were shown to be consistent and have been previously reported in the field isolates (Takala et al 2009). This conforms to the study done by applying high- throughput genotyping methods to a large sample of infections in Mali (Takala, et al 2009). A Tajima's D of 0.32552 can be interpreted as no evidence of departure from neutrality in this region. No-significance of positive Tajima's D has been previously reported in studies done in other places (Simpalipan et al. 2014).

Of the 9 MSP1 haplotypes obtained in this study based on the $MSP1_{19}$ region, 6 were previously reported in *P. falciparum* (Takala et al, 2009), and the other 2, QTSNRLK and QTSNGFK are unique to our population however they are rare variants at a frequency of 2%. Despite this number of haplotypes, only 2 very similar haplotypes (QKSNGLK and EKSNGLK) were predominantly circulating to the level close to 80% in the population studied. The predominant haplotypes are similar to the strains identified in different studies. QKSNGLK is similar to FVO or Wellcome laboratory strains (Holder et al. 1985) while EKSNGLK is similar to FUP or Uganda-PA laboratory strains (Chang et al, 1988).

The PfRH4 ligand binds to the erythrocyte surface CR1 receptor through its N- terminal region. This protein is 205-kDa (Stubbs et al. 2005). It has been previously observed that the binding domain of PfRH4 is an 88-kDa region of PfRH4 (aa 28 to 766) that was tagged with an amino-terminal six-His tag (RH4.9) (Tham et al. 2009). In this study we were able to sequence only the C terminal part of RH4.9 (aa 430 to 639). In their study, to delineate the binding region of PfRH4, Tham et al (2009) expressed three overlapping recombinant proteins spanning RH4.9: RH4.10 (aa 28 to 340), RH4.11 (aa 233 to 540), and RH4.12 (aa 493 to 700). Only RH4.10 and RH4.11 bound to the surfaces of the erythrocytes, suggesting that the most C-terminal region of RH4.9 is not required for binding (Tham et al. 2009; Reiling et al. 2012).

This study was able to show that the PfRH4 gene region that we sequenced had limited polymorphism. Only 3 SNPs were identified in that region, all being non-synonymous and at least 20% frequent. The Tajima's D and Fu and Li D indices showed no significant departure from neutrality. Thus, there is no evidence of significant selective pressure in the C-terminal region of PfRH4 based on what we sequenced. This means that this binding region is conserved and this is consistent with previous studies on PfRH4. In the study conducted by Reiling et al (2012) there were four SNPs in the PfRH4.9 region resulting in one synonymous and 3 non-synonymous changes determined from 12 isolates, the positons were not precise in their paper. Of the 3 SNPs observed in this study, 2 were reflected in the public database (PlasmoDB.org). Those 2 SNPs are located at position 435 and 438 referring to the 3D7 protein sequence. The SNP newly found in this study was at position 500 and was a singleton in the population studied.

The conserved nature of this region was also confirmed by the minimal number of haplotypes and low haplotype diversity index. This study identified only 5 haplotypes of PfRH4 gene with only 2 predominantly frequent (greater than 85%). In the study previously done on PfRH4 disruption showed that this protein has a key role in erythrocyte invasion compared to other proteins from Erythrocyte binding (EBL) and Reticulocyte binding-like homologue (RH) proteins families that were proven to be involved in the erythrocyte invasion. PfRH4 plays major role in the switching of invasion pathways by *P. falciparum* (Stubbs et al. 2005). Disruption of this gene in the W2mef strain variants inhibited the switching to sialic acid–independent invasion, and no parasites were observed after extended culture generation on neuraminidase-treated red cells (up to 40 days) or after the culture had been returned to untreated erythrocytes. The same result was observed for all W2mefD RH4 (RH4 disrupted) clones (Stubbs et al. 2005).

This information together with the fact that there is limited polymorphism in the region examined suggests that there is no selective pressure in the PfRH4 gene. This is consistent with previous studies and it is very crucial in the consideration of PfRH4 as potential candidate for malaria vaccine development. Unlike MSP1 and EBL1, there is limited variability, hence targeting RH4 would be potentially more effective due to limited antigenic variability as a vaccine target.

In the EBL1 ligand, three variants have been reported so far: i) normal EBL1 gene sequence similar to Dd2 strain; ii) EBL1 gene containing a 5 thymidine insertion (coded as 5 T's) similar to 3D7 strain; and iii) complete EBL1 gene deletion (coded deletion) similar to HB3 strain. The majority of field isolates previously identified from Colombia, Peru and Kenya had a 5 T's insertion, suggesting that no functional EBL1 protein is produced. That was confirmed in this study. In contrast, in some places like Brazil and Peru, isolates had a normal EBL1 gene sequence. We could also identify few isolates with normal EBL1 protein (having no 5 Ts insertion) in our study population. As observed in different studies, no PCR products were obtained for few isolates, and this has been interpreted as complete gene deletion similar to the HB3 strain (Lopez-Perez et al. 2012). In this study, we successfully amplified 71 samples among the 93 that constituted our sample size. We cannot argue confidently enough that this was not due to the deletion of this gene in the isolates for which we were unable to obtain PCR products after repeated PCR attempts. These observations together combined, suggest that the EBL1 ligand does not play an indispensable role in invasion by some isolates, but is potentially more utilized by other isolates (Lopez-Perez et al. 2012).

The EBL1 gene is the only gene characterized in this study that was confirmed to be under selection, precisely the receptor binding domain. The Tajima's D value of -1.87248 (Table 3.2) obtained based on region II, suggests that the gene is under purifying selection ($p<0.05$). This type of selection acts to remove genetic variation from the population. This is consistent with research previously done on this gene (Verra et al. 2006). Despite the truncated form of EBL1 and unconfirmed deletions observed in some isolates, the results show that other isolates in the same population may need it among other ligands for invasion. We can also argue that this may reflect underlying host genetic factors that drive selective pressure on EBL1 variants (Githui et al. 2010).

Neutrality tests were previously computed for the erythrocyte receptors and the results conforms to the ones found for ligands. No evidence of departure from neutrality was found for both band 3 and CR1 as it was found for their ligands, MSP1 and RH4, respectively. Interestingly, EBL1 and its receptor Glycophorin B were found to be highly polymorphic and under purifying selection ($p<0.05$). Purifying selection removes diversity trying to conserve certain features that confer survival during evolution in case there is a selective pressure. It might be speculated that the EBL1 adapts to genetic variations that have been occurring in the Glycophorin B gene in humans or vice-versa.

4.2 Conclusion and recommendation

This study expands the understanding and current knowledge of the variations of $MSP1_{19}$, RH4.9 and EBL1 region I and region II polymorphisms in field isolates of *P. falciparum* in Kilifi population in Kenya. The variations have created different haplotypes of each gene, of which one or two are most predominantly circulating in the studied population. Proteins like RH4 and MSP1 so far show hope as potential vaccine candidates due to their previously demonstrated role of inducing immune response and the conserved nature of their receptor binding domains showcased in this study. It is still hard to understand the role of EBL1 pathway in invasion due to its varied characteristics, as a functional, then non-functional protein in the same population. Comparison of receptor and ligand neutrality features was very useful in providing a clue on the interaction. With a limited sample size, this study made us hypothesize that genetic variation in receptors can affect changes in the ligand although this needs to be further studied.

We recommend further research on the choice of pathways during merozoite invasion. Computational tools for studying host-pathogen interactions need to be established in order to have a standardized way of doing interaction analysis. We strongly recommend use of larger sample sizes in further endeavors in the field of host-pathogen interaction research.

REFERENCES

Adams, J. H., B. K. Sim, S. A. Dolan, X. Fang, D. C. Kaslow, and L. H. Miller. 1992. "A Family of Erythrocyte Binding Proteins of Malaria Parasites." *Proceedings of the National Academy of Sciences* 89 (15): 7085–89.

Adams, John H., Peter L. Blair, Osamu Kaneko, and David S. Peterson. 2001. "An Expanding Ebl Family of Plasmodium Falciparum." *Trends in Parasitology* 17 (6): 297–99.

Amambua-ngwa, Alfred, Kevin K A Tetteh, Magnus Manske, Natalia Gomez-escobar, Lindsay B Stewart, M Elizabeth Deerhake, Ian H Cheeseman, et al. 2012. "Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites" 8 (11).

Ayala, Everick Pinedo, Viviana V Terry, Hilja Cabrera, Lilia Vinetz, Joseph M. Vinetiz. 2006. "Polymerase Chain Reaction and Molecular Genotyping to Monitor Parasitological Response to Anti-Malarial Chemotherapy in the Peruvian Amazon" 74 (4): 546–53.

Bannister, L H, G H Mitchell, G A Butcher, E D Dennis, and S Cohen. 1986. "Structure and Development of the Surface Coat of Erythrocytic Merozoites of Plasmodium Knowlesi." *Cell and Tissue Research* 245 (2): 281–90.

Bejon, Philip, Thomas N Williams, Anne Liljander, Abdisalan M Noor, Juliana Wambua, Edna Ogada, Ally Olotu, et al. 2010. "Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya." *PLoS Medicine* 7 (7): e1000304.

Cai, Hong, Zhan Zhou, Jianying Gu, and Yufeng Wang. 2012. "Comparative Genomics and Systems Biology of Malaria Parasites Plasmodium." *Current Bioinformatics* 7 (4).

Cock, Peter J a, Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. 2009. "The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants." *Nucleic Acids Research* 38 (6): 1767–71.

Conway, D. J. 1997. "Natural Selection on Polymorphic Malaria Antigens and the Search for a Vaccine." *Parasitology Today (Personal Ed.)* 13 (1): 26–29. doi:10.1016/S0169-4758(96)10077-6.

Cowman, Alan F., Drew Berry, and Jake Baum. 2012. "The Cellular and Molecular Basis for Malaria Parasite Invasion of the Human Red Blood Cell." *The Journal of Cell Biology* 198 (6): 961–71.

Crosnier, Cécile, Madushi Wanaguru, Brian McDade, Faith H. Osier, Kevin Marsh, Julian C. Rayner, and Gavin J. Wright. 2013. "A Library of Functional Recombinant Cell-Surface and Secreted P. Falciparum Merozoite Proteins." *Molecular & Cellular Proteomics* 12 (12): 3976–86.

Dolan, S A, J L Proctor, D W Alling, Y Okubo, T E Wellems, and L H Miller. 1994. "Glycophorin B as an EBA-175 Independent Plasmodium Falciparum Receptor of Human Erythrocytes." *Molecular and Biochemical Parasitology* 64 (1): 55–63.

Drummond, Paul B, and David S Peterson. 2005. "An Analysis of Genetic Diversity within the Ligand Domains of the Plasmodium Falciparum Ebl-1 Gene." *Molecular and Biochemical Parasitology* 140 (2): 241–45.

Gardner, Malcolm J, Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W Hyman, Jane M Carlton, et al. 2002. "Genome Sequence of the Human Malaria Parasite Plasmodium Falciparum." *Nature* 419 (6906): 498–511.

Gaur, Deepak, Jill R Storry, Marion E Reid, John W Barnwell, and Louis H Miller. 2003. "Plasmodium Falciparum Is Able To Invade Erythrocytes through a Trypsin-Resistant Pathway Independent of Glycophorin B" 71 (12): 6742–46.

Gerold, P, L Schofield, M J Blackman, A A Holder, and R T Schwarz. 1996. "Structural Analysis of the Glycosyl-Phosphatidylinositol Membrane Anchor of the Merozoite Surface Proteins-1 and -2 of Plasmodium Falciparum." *Molecular and Biochemical Parasitology* 75 (2): 131–43.

Gilson, Paul R, and Brendan S Crabb. 2009. "Morphology and Kinetics of the Three Distinct Phases of Red Blood Cell Invasion by Plasmodium Falciparum Merozoites." *International Journal for Parasitology* 39 (1): 91–96.

Githui, Elijah K, David S Peterson, Rashid a Aman, and Abdirahman I Abdi. 2010. "Prevalence of 5' Insertion Mutants and Analysis of Single Nucleotide Polymorphism in the Erythrocyte Binding-like 1 (ebl-1) Gene in Kenyan Plasmodium Falciparum Field Isolates." *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 10 (6). Elsevier B.V.: 834–39.

Goel, Vikas K., Xuerong Li, Huiqing Chen, Shih-Chun Liu, Athar H. Chishti, and Steven S. Oh. 2003. "Band 3 Is a Host Receptor Binding Merozoite Surface Protein 1 during the Plasmodium Falciparum Invasion of Erythrocytes." *Proceedings of the National Academy of Sciences* 100 (9): 5164–69.

Holder, A A. 1988. "The Precursor to Major Merozoite Surface Antigens: Structure and Role in Immunity." *Progress in Allergy* 41: 72–97.

Holder, Anthony A., Michael J. Lockyer, Karel G. Odink, Jasbir S. Sandhu, Valentina Riveros-Moreno, Stephen C. Nicholls, Yvonne Hillman, et al. 1985. “Primary Structure of the Precursor to the Three Major Surface Antigens of Plasmodium Falciparum Merozoites.” *Nature* 317 (6034): 270–73.

José, By, A Guevara Patiño, Anthony A Holder, Jana S Mcbride, and Michael J Blackman. 1997. “Antibodies That Inhibit Malaria Merozoite Surface Protein–1 Processing and Erythrocyte Invasion Are Blocked by Naturally Acquired Human Antibodies” 186 (10).

Kevin Marsh, Dayo Forster, Catherine Waruiru Isiah Mwangi, Maria Winstanley, Victoria Marsh, Charles Newton, Peter Winstanley, Peter Warn, Norbert Peshu, Geoffrey Pasvol, Robert Snow. 1995. “Indicators of Life-Threatening Malaria in African Children” 332 (21): 1399–1404.

Kuhn, Susan M, and Anne E McCarthy. 2006. “Paediatric Malaria: What Do Paediatricians Need to Know?” *Paediatrics & Child Health* 11 (6): 349–54.

Lopez-Perez, Mary, Elizabeth Villasis, Ricardo L D Machado, Marinete M Póvoa, Joseph M Vinetz, Silvia Blair, Dionicia Gamboa, and Sara Lustigman. 2012. “Plasmodium Falciparum Field Isolates from South America Use an Atypical Red Blood Cell Invasion Pathway Associated with Invasion Ligand Polymorphisms.” *PloS One* 7 (10): e47913.

Maier, Alexander G., Jake Baum, Brian Smith, David J. Conway, and Alan F. Cowman. 2009. “Polymorphisms in Erythrocyte Binding Antigens 140 and 181 Affect Function and Binding but Not Receptor Specificity in Plasmodium Falciparum.” *Infection and Immunity* 77 (4): 1689–99.

Mayer, D C Ghislaine, Jian-Bing Mu, Xiaorong Feng, Xin-zhuan Su, and Louis H Miller. 2002. "Polymorphism in a Plasmodium Falciparum Erythrocyte-Binding Ligand Changes Its Receptor Specificity." *The Journal of Experimental Medicine* 196 (11): 1523–28.

Mayer, D C Ghislaine, Jian-Bing Mu, Osamu Kaneko, Junhui Duan, Xin-zhuan Su, and Louis H Miller. 2004. "Polymorphism in the Plasmodium Falciparum Erythrocyte-Binding Ligand JESEBL/EBA-181 Alters Its Receptor Specificity." *Proceedings of the National Academy of Sciences of the United States of America* 101 (8): 2518–23.

Mayer, D. C. Ghislaine, Joann Cofie, Lubin Jiang, Daniel L. Hartl, Erin Tracy, Juraj Kabat, Laurence H. Mendoza, and Louis H. Miller. 2009. "Glycophorin B Is the Erythrocyte Receptor of Plasmodium Falciparum Erythrocyte-Binding Ligand, EBL-1." *Proceedings of the National Academy of Sciences of the United States of America* 106 (13): 5348–52.

McFadden, G I, M E Reith, J Munholland, and N Lang-Unnasch. 1996. "Plastid in Human Parasites." *Nature* 381 (6582): 482. doi:10.1038/381482a0.

Morrissette, Naomi S., and L. David Sibley. 2002. "Cytoskeleton of Apicomplexan Parasites." *Microbiology and Molecular Biology Reviews* 66 (1): 21–38. doi:10.1128/MMBR.66.1.21-38.2002.

Mu, Jianbing, Philip Awadalla, Junhui Duan, Kate M McGee, Jon Keebler, Karl Seydel, Gilean a T McVean, and Xin-zhuan Su. 2007. "Genome-Wide Variation and Identification of Vaccine Targets in the Plasmodium Falciparum Genome." *Nature Genetics* 39 (1): 126–30.

Murray, Christopher JL Rosenfeld, Lisa C Lim, Stephen S Andrews, Kathryn G Foreman, Kyle J Haring, Diana Fullman, Nancy Naghavi, Mohsen Lozano, Rafael Lopez, Alan D. 2012. "Global Malaria Mortality between 1980 and 2010: A Systematic Analysis." *The Lancet* 379 (9814): 413–31.

O'Meara, Wendy P, Phillip Bejon, Tabitha W Mwangi, Emelda A Okiro, Norbert Peshu, Robert W Snow, Charles RJC Newton, and Kevin Marsh. 2008. "Effect of a Fall in Malaria Transmission on Morbidity and Mortality in Kilifi, Kenya." *The Lancet* 372 (9649): 1555–62.

Peterson, M Gregory, Ross L Coppel, M Bridget Moloney, and David J Kemp. 1988. "Third Form of the Precursor to the Major Merozoite Surface Antigens of Plasmodium Falciparum" 8 (6): 2664–67.

Plebanski, Magdalena, Owen Proudfoot, Dodie Pouniotis, Ross L. Coppel, Vasso Apostolopoulos, and Graham Flannery. 2002. "Immunogenetics and the Design of Plasmodium Falciparum Vaccines for Use in Malaria-Endemic Populations." *Journal of Clinical Investigation* 110 (3): 295–301.

Polley, Spencer D., Watcharee Chokejindachai, and David J. Conway. 2003. "Allele Frequency-Based Analyses Robustly Map Sequence Sites Under Balancing Selection in a Malaria Vaccine Candidate Antigen." *Genetics* 165 (2): 555–61.

Rayner, J C, M R Galinski, P Ingravallo, and J W Barnwell. 2000. "Two Plasmodium Falciparum Genes Express Merozoite Proteins That Are Related to Plasmodium Vivax and Plasmodium Yoelii Adhesive Proteins Involved in Host Cell Selection and Invasion." *Proceedings of the National Academy of Sciences of the United States of America* 97 (17): 9648–53.

Reiling, Linda, Jack S. Richards, Freya J. I. Fowkes, Danny W. Wilson, Watcharee Chokejindachai, Alyssa E. Barry, Wai-Hong Tham, et al. 2012. "The Plasmodium Falciparum Erythrocyte Invasion Ligand Pfrh4 as a Target of Functional and Protective Human Antibodies against Malaria." *PLoS ONE* 7 (9): e45253.

Shannon l. Takala, david l. Smith, Mahamadou A. Thera, Drissa Coulibaly, Ogobara K. Doumbo, Christopher V. Plowe. 2009. “Short Report: Rare Plasmodium Falciparum Merozoite Surface Protein 1 19-Kda (Msp-119) Haplotypes Identified In Mali Using High-Throughput Genotyping Methods” 76 (5): 855–59.

Sim, B., C. Chitnis, K Wasniowska, T. Hadley, and L. Miller. 1994. “Receptor and Ligand Domains for Invasion of Erythrocytes by Plasmodium Falciparum.” *Science* 264 (5167): 1941–44.

Simpalipan, Phumin, Sittiporn Pattaradilokrat, Napaporn Siripoon, Aree Seugorn, Morakot Kaewthamasorn, Robert D J Butcher, and Pongchai Harnyuttanakorn. 2014. “Diversity and Population Structure of Plasmodium Falciparum in Thailand Based on the Spatial and Temporal Haplotype Patterns of the C-Terminal 19-kDa Domain of Merozoite Surface Protein-1.” *Malaria Journal* 13 (1). Malaria Journal: 1–11.

Soulama, Issiaka, Issa Nébié, Alphonse Ouédraogo, Adama Gansane, Amidou Diarra, Alfred B Tiono, Edith C Bougouma, et al. 2009. “Plasmodium Falciparum Genotypes Diversity in Symptomatic Malaria of Children Living in an Urban and a Rural Setting in Burkina Faso.” *Malaria Journal* 8 (January): 135.

Stubbs, Janine, Ken M Simpson, Tony Triglia, David Plouffe, Christopher J Tonkin, Manoj T Duraisingh, Alexander G Maier, Elizabeth a Winzeler, and Alan F Cowman. 2005. “Molecular Mechanism for Switching of P. Falciparum Invasion Pathways into Human Erythrocytes.” *Science (New York, N.Y.)* 309 (5739): 1384–87.

Tham, Wai-Hong, Danny W. Wilson, Linda Reiling, Lin Chen, James G. Beeson, and Alan F. Cowman. 2009. “Antibodies to Reticulocyte Binding Protein-Like Homologue 4 Inhibit Invasion of Plasmodium Falciparum into Human Erythrocytes.” *Infection and Immunity* 77 (6): 2427–35.

Treeck, Moritz, Sonja Zacherl, Susann Herrmann, Ana Cabrera, Maya Kono, Nicole S. Struck, Klemens Engelberg, et al. 2009. “Functional Analysis of the Leading Malaria Vaccine Candidate AMA-1 Reveals an Essential Role for the Cytoplasmic Domain in the Invasion Process.” *PLoS Pathogens* 5 (3).

Verra, Federica, Watcharee Chokejindachai, Gareth D Weedall, Spencer D Polley, Tabitha W Mwangi, Kevin Marsh, and David J Conway. 2006. “Contrasting Signatures of Selection on the Plasmodium Falciparum Erythrocyte Binding Antigen Gene Family.” *Molecular and Biochemical Parasitology* 149 (2): 182–90.

Wright, Gavin J., and Julian C. Rayner. 2014. “Plasmodium Falciparum Erythrocyte Invasion: Combining Function with Immune Evasion.” *PLoS Pathogens* 10 (3): 1–7.