Identifying *Plasmodium falciparum* merozoite surface protein 1 variants in children with multiple malaria infections in Kilifi, Kenya

By

Reuben Mangi Yaa

I56/81536/2012

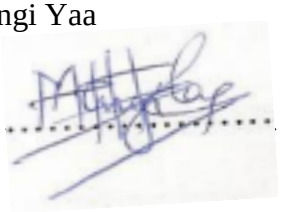
BSc. Biomedical Technology (UoN)

A thesis submitted in partial fulfilment of requirements for the award of degree of Master of Science in Biotechnology, University of Nairobi

DECLARATION

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

Reuben Mangi Yaa

Signature……  ……..……Date: 13th April 2016

SUPERVISORS' APPROVAL

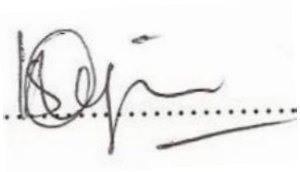
We confirm that this thesis has been submitted with our approval as university supervisors;

Dr. Isabella Oyier,

Kenya Medical Research Institute (KEMRI), Wellcome Trust,

Centre for Geographic Medicine Research Coast,

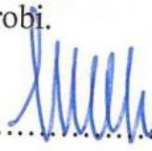
Kilifi.

Signature……  …….. Date: 13th April 2016

Dr. George Obiero,

Centre for Biotechnology and Bioinformatics (CEBIB),

University of Nairobi.

Signature……  ……. Date: 5th May 2016

ACKNOWLEDGEMENTS

I express my deep appreciation for the professional guidance and mentorship I was given by my supervisors. It was really a great privilege to work under their tutelage. Dr. Oyier, I thank you for your immediate intervention, and the alternative you offered after my previous project couldn't be realized. I am grateful for having been incorporated into the wider malaria and immunity integrated studies and for your subsequent supervision. Similarly, I was fortunate in receiving support from Dr. George Obiero, who played a pivotal role regarding administrative input and offered regular positive critics. Such input always "lightened up the path" for the successful completion of this study.

I also recognize the technical support that I got from CEBIB lab Technologists. I appreciate Ann Owiti, for the thorough orientation she gave me at the onset of the project. I am grateful to Mr. Edwin Rono, who kindly facilitated all the procurement procedures in acquiring the lab supplies and the sequencing.

My fellow students Kelvin Wachiuri, Jean Pierre Musabyimana and Akiso Mbendo, are my great friends and would regularly encourage me in my project progress.

My family supported and encouraged me throughout my MSc. advancement. I thank my father, Robert Yaa Mangi, my mother Rachel and my siblings, Ben, Chris, Matlida and Esther. I treasure and value the personal interest they showed in my studies. Their prayers, moral advice, encouragement and monetary support would at all times supplement my hopes and determination to finish my postgraduate studies.

I appreciate the input of Mrs. Val Nibbs, from Jersey UK for proof reading the grammar in my written submissions.

DEDICATION

This thesis is dedicated to my family

TABLE OF CONTENTS **PAGE**

ABBREVIATIONS

AMA-1 ……..……Apical membrane antigen 1

ABRA ……..**.....**Acid basic repeat antigen

bp…….Base pair

CEBIB……Centre for Biotechnology and Bioinformatics

CCD…… Coupled charged device

C-terminous…….....Carboxyl terminus

ddNTPs……Dideoxynucleotide triphosphates

dNTPs…….…..deoxynucleotide triphosphates

DNA……..……Deoxyribonucleic acid

DBP ……... Duffy binding protein

DnaSp……... DNA sequence polymorphism

EBA-175 ……Erythrocyte binding antigen 175

EIR……..Entomologic inoculation rate

EGF…….……...Epidermal like growth factor

ICB……..……...Interspecies conserved blocks

kDa……..Kilo Dalton

MEGA ……..Molecular Evolutionary Genetics Analysis

MUSCLE…….….. Multiple sequence alignment

MSP …….……Merozoite surface protein

GPI ……...……Glycosyl phosphatidyl inositol

EBLs ……Erythrocyte binding like proteins

$MgCl_2$ …….. Magnesium Chloride

µl……....Microlitre

ml…….Millilitre

^oC……...Celsius degrees

mA…….....Milliamperes

N-terminus……...…….Amino terminus

PCR…….……...Polymerase chain reaction

PHYLIP……..Phylogeny inference package

PVM …….Parasitophorous vacuole membrane

PNG……..Portable network graphics

***Pf*Rh** ……...……... *P. falciparum* reticulocyte binding protein homolog

RBCs ……....Red blood cells

RBLs .....……..... Reticulocyte binding homologue proteins

SCF……. ……Simple configuration files

SERA ……...……....Serine repeat antigen

SNP..... Single Nucleotide polymorphism

Taq…….. *Thermus aquaticus*

Tgo……Thermococcus gorgonarius

UC……USEARCH cluster format (UC)

USEARCH……..…….Unique search

LIST OF FIGURES **PAGE**

LIST OF TABLES **PAGE**

ABSTRACT

The *Plasmodium falciparum* Merozoite Surface Protein 1(*Pf* MSP1), a predominant antigen on the surface of the asexual blood stage of the parasite, plays a role in erythrocyte invasion. It elicits immune responses during exposure to natural *P. falciparum* infections, hence, it is a potential vaccine candidate. However, its extensive sequence diversity causes antigenic variability. Parasites that express variants other than that targeted by immune protection mounted as a result of a vaccine variant, evade the resultant host immune protection. This compromises the efficacy of allele-specific vaccines formulated to protect against a single variant. Due to this, *Pf* MSP1 has been extensively studied, including in Kenya. However, the extent of *Pf* MSP1 diversity in children with multiple infections are unknown in Kilifi which is a moderate to high malaria transmission zone. Parasite genomic DNA was extracted from 421 blood samples in 33 children aged below 5 years who had at least 9 multiple infections. The *Pf* $MSP1_{19}$kDa region was amplified by Polymerase Chain Reaction (PCR). Samples that generated good amplicons were sequenced and analysed for allelic polymorphisms using Sequencher 5.3 and MEGA 5.2 softwares. Sequence clustering and haplotype frequencies were done in 150 samples that generated unambiguous sequences using the Phyclust package in R programming and command line in executable USEARCH software. A total of 11 haplotypes distributed randomly across the infection episodes were identified. The Q-KSNG-L(36%), E-KSNG-L(26%) and Q-KSNG-F(18%) haplotypes were the most prevalent in this population. These haplotypes were mainly observed 2 to 3 times between consecutive infections in some children. Following such a pattern, the variants did not re-appear in the subsequent isolates from the respective child. This was due to the host immune responses that protected against those variants which were previously exposed to the child. The E-TSSR-L variant, corresponding to *P. falciparum* 3D7 isolate, a widely used allele for vaccine formulation, had a very low frequency (1.3%). These findings provide useful information in designing future multi-allelic malaria vaccines for deployment in endemics.

CHAPTER 1

LITERATURE REVIEW

1.1 Background

Being one of the world's most lethal diseases, malaria remains a scourge to humanity since antiquity. The etiological agent of malaria is an obligate eukaryotic intra-erythrocytic protozoan of the genus *Plasmodium*. This genus has more than 100 different species. *P. falciparum, P vivax, P. knowlesi, P. malariae* and *P.ovale* are the main species in this genus that cause malaria in humans resulting in morbidity in the populations where they are endemic (Liu *et al*., 2010; Francia & Striepen, 2014). *Falciparum* malaria is the most severe form of this disease in humans. The erythrocytic stages of this parasite are responsible for all symptoms and pathologies of the disease (Cox, 2010). The infected erythrocytes in some cases do not circulate in blood but get sequestered and adhere to the blood vessels. Destruction of red cells and blockage of vessels leads to hypoxia and toxemia in the visceral organs like the brain, kidneys, spleen and lungs, and in some cases causing cerebral malaria when in brain or in the placenta, malaria in pregnancy. Generally, the symptoms of malaria are; fever, headache, malaise, fatigue, muscular pains, and occasionally nausea, vomiting and diarrhoea (Miller *et al*., 2002). All these symptoms can cause death if the patient is not treated in time.

Malaria thrives in the tropical areas of Asia, Africa, and Central and South America, where it is endemic. Even though, there have been major reductions in mortality and morbidity in these areas, malaria still remains a major paediatric killer in many parts of sub-Saharan Africa, which bears the greatest burden of disease (Nahlen *et al*., 2005). Currently, there are approximately 219

million malaria cases and mortality rates of approximately 1 million cases a year in the world with most of these being in children living in sub-Saharan Africa (Murray *et al*., 2012; World Health Organization, Tuschman, World Health Organization, & Global Malaria Programme, 2013). The 2010 Kenya Malaria Indicator Survey established that Kenya is among these countries, and nearly 28 million Kenyans live in areas of malaria risk, a majority of them are children under the age of 15 years who are regularly infected. The Kenyan lake and coastal zones have a high prevalence of malaria with high transmission rate due to the parasite thriving in the hot and humid climatic conditions of a tropical zone.

While anti-malarial drugs, insecticides on mosquitos and sleeping under long-lasting insecticide-treated bed nets have considerably reduced the burden of the disease in many areas (Gimnig *et al*., 2003), these measures are now becoming decreasingly efficacious because of anti-malarial drug resistance in the parasite, insecticide resistance in mosquitoes, and socio-economic deficits and warfare in human populations (Bloland, Organization, & others, 2001). Studies done by Delves *et al*., (2012) proved that free population movement within the globe is accountable for the introduction of resistant parasites to areas previously free of drug resistance.

1.2 Parasite life cycle

The malaria parasite undergoes a complex life cycle that requires both a human host and a female Anopheles mosquito as shown in Figure 1.1. When an infected female Anopheles mosquito bites a human during its blood meal, it injects saliva that contains anticoagulant into the skin. The saliva also has the infectious form of the parasite, known as the sporozoite (Bannister and London 2009). The sporozoites circulate freely in blood to the liver where they pass through Kuppfer cells and

invade the liver cells (Cowman and Crabb 2006). In the hepatocytes, they undergo a phase of asexual multiplication (exo-erythrocytic schizogony) and transform the infected hepatocyte into a schizont, which harbours the infectious form of the parasite. This occurs between one to two weeks. Mature schizonts rupture and release about 8-16 invasive merozoites during egress into the bloodstream (Cox, 2010).

The released merozoites rapidly invade the red blood cells (RBCs), take up haemoglobin and begin the asexual schizogony blood-stage cycle through ring, trophozoite and schizont stages to produce 16-32 merozoites that are released upon rupture of the infected cell to the extracellular compartment. The newly released merozoites invade other RBCs (Figure 1.1b). This infection cycle is repeated almost indefinitely and is responsible for the clinical symptoms of the disease. Unless the cycle is stopped, it may lead to the massive destruction of the host cells.

Some of the merozoites growing in the RBCs may develop into male and female sexual forms known as gametocytes that remain circulating in the peripheral bloodstream. These will be taken up along with blood by the mosquito during its feeding (Figure 1.1c). Once in the mosquito's mid gut, the gametocytes develop into sperm-like male gametes or large, egg-like female gametes. Fertilization of the gametes occurs there to form a motile zygote known as the ookinete. The ookinete penetrates the gut wall after sporogony and produces an oocyst filled with infectious sporozoites. Mature oocysts rupture and release thread-like sporozoites that migrate to the mosquito's saliva-producing glands for injection into a human host during the next blood meal, hence, the cycle starts over again (Cowman and Crabb 2006; Iyer *et al*., 2007; Miller *et al*., 2002).

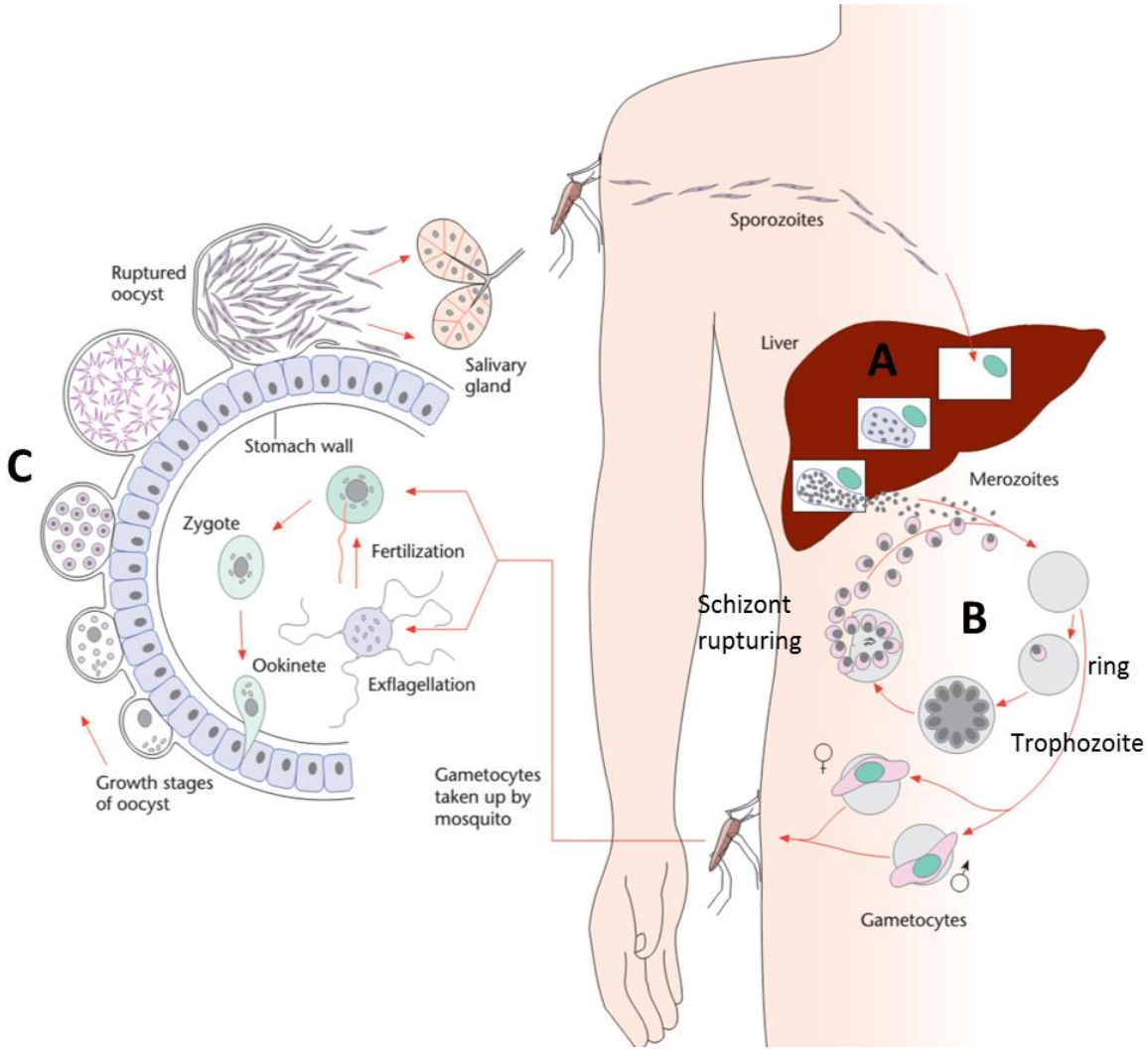


Figure 1.1: Life cycle of *P. falciparum* parasite

Main features of the life cycle of the malaria parasite Plasmodium falciparum, showing its different phases in vertebrate and mosquito hosts. (A) shows the liver stage of the parasite life cycle where the sporozoite invade the hepatocytes, mature and multiply in exo-erythrocytic asexual division and released in blood. (B) Shows the intra-erythrocytic cycle of the blood stage of the parasite, where the merozoites invade the red blood cells and others differentiate to gametocytes. (C) Shows the sporogonic cycle at the mosquito vector which multiply to sporozoite (Miller et al., 2002).

1.3 Invasion of erythrocytes by *P. falciparum* merozoite

Erythrocyte invasion by the parasite occurs at the merozoite stage of its life cycle. This process is complex and occurs rapidly in a cascade of events that involve a multi-step sequence in about 60 seconds as described by Dvorak *et al*., (1975). The specificity for invasion is manifested through receptor- ligand interactions between the host cell and the invading parasite (Brown 1977; Haldar & Mohandas 2007). The initial contact between the merozoite and erythrocyte which occurs by random collision contact, is a crucial step, as the parasite must distinguish between competent erythrocytes for invasion and other cell types. This recognition and primary adherence is relatively long distance, apparently of low affinity and is reversible (Bannister & Mitchell, 2003).

The initial contact requires interactions of merozoite surface proteins (MSP) and specific receptors on the host cell (Aikawa *et al*., 1978; Low *et al*., 2007). Reorientation ensues after initial attachment and stimulates a rapid 'wave' of erythrocyte membrane deformation to juxtapose the apical end of the merozoite to the erythrocyte membrane. This allows a much closer interaction.

The re-orientated 'apical' end expels apical secretory organelles called the micronemes, rhoptries and dense granules towards the junction of invasion. Contents of these organelles play an important role in invasion as well as host cell modification after invasion (Haldar & Mohandas, 2007) resulting in an irreversible junction formation between the parasite and the host cell. The formed tight junction moves from the apical to the posterior end of the merozoite in a complex series of events powered by the parasite actin-myosin motor (Keeley & Soldati, 2004). As invasion progresses, a depression of the erythrocyte membrane deepens and conforms to the shape of the invading merozoite (Iyer *et al*., 2007) allowing the parasite to push through and enter the

erythrocyte with the formation of an incipient parasitophorous vacuole membrane (PVM) as it sheds off its fuzzy coating covering the merozoite surface (Figure 1.2).

The PVM protects the parasite from the host-cell cytoplasm, thus it helps in forming an environment hospitable for the parasite development in the host cell. When the merozoite has completed its entry, the PVM seals and the junction fuses at its posterior end (Ford *et al*., 2007; Gilberger 2003).

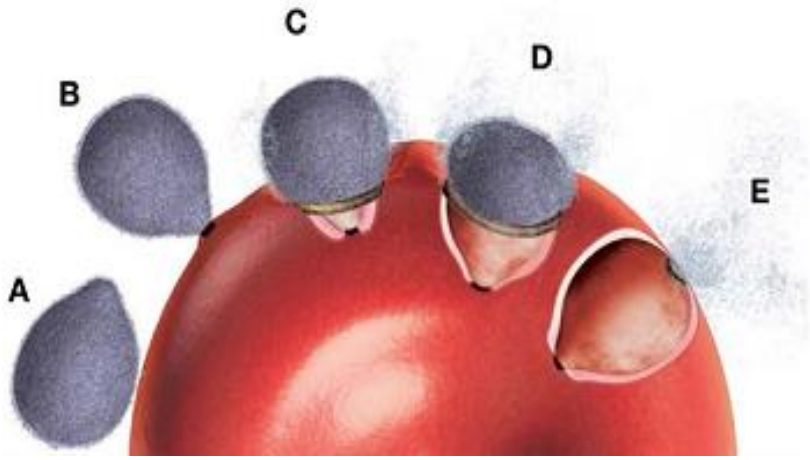


Figure 1.2: Process of RBC invasion
The process of erythrocyte invasion. (A) the merozoite makes an initial contact to the erythrocyte. (B) The merozoite reorients to juxtapose its apical end to the erythrocyte. (C) The merozoite indents the erythrocyte membrane and pushes in as it creates an incipient parasitophorus vacuole. (D) The invading merozoite sheds off its fuzzy coating through enzyme shadase. (E) The merozoite seals off the erythrocyte membrane once the tight junction reaches the posterior end of the parasite and initiate post invasion modification (Cowman and Crabb 2006; Cowman et al., 2012).

1.4 Molecular basis of invasion by merozoites

The merozoite is an ovoid cell and its apical end is a truncated cone shape (Figure 1.3). Its shape, surface proteins and its organelle contents are essential and adapted for erythrocyte invasion. The merozoite organelle repertoire, which include the apical complex of secretory organelles

(micronemes, rhoptries, and dense granules) play an essential role for the process of invasion. The contents initiate signal pathways both in the invading and the host cell to mediate the process of invasion. Dense granules are released soonest after invasion and include components of a putative protein translocon that is inserted into the PVM. Other ligands responsible for the process of invasion are glycosylphosphatidylinositol (GPI) anchored membrane proteins and their associated partners, and peripheral surface proteins, all of which are largely expressed on the merozoite surface coat. (Sanders, 2005).

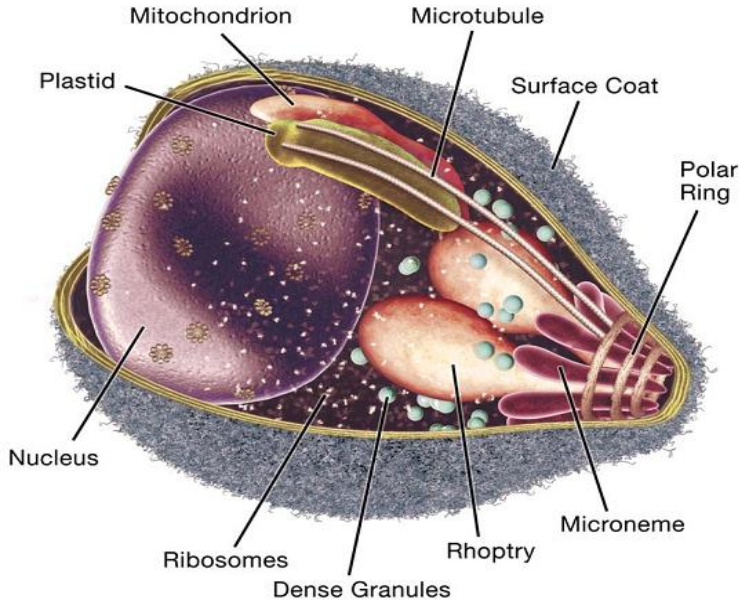


Figure 1.3: Fine structure of *Plasmodium falciparum* Merozoite
The molecular structure of the merozoite showing important organelles with their position in the merozoite. (Cowman and Crabb 2006).

Currently, there are eight GPI anchored proteins known and predicted to be located on the merozoite surface of *P. falciparum,* and a few have known potential erythrocyte ligands (Sanders, 2005) as seen in (Table 1.1). Besides the GPI anchor, many surface proteins share similarities including cysteine-rich domains that are of potential significance in adherence to the erythrocyte (Iyer *et al*., 2007). Merozoite Surface Proteins (MSP), one among the eight GPI anchored proteins,

is largely expressed on the merozoite surface, and so is the only class with higher chances of interacting with the RBCs. There exist up to 10 MSPs based on functional conservation, sequence divergence and domain arrangement. Of them, MSP1 is the most dominant and abundant protein, and hence, has been well studied and characterized (Cowman and Crabb 2006). It signals the invasion process by interacting on the erythrocyte surface with its receptor, protein band 3, which is coupled to G- protein. The exact roles in erythrocyte invasion for MSP 2 -10 are not clear, however, they are believed to play an indirect role such as forming complexes with MSP 1 for initial interaction with erythrocyte surface (Lin *et al*., 2016).

Other peripheral surface proteins involved in both primary and secondary contact include the acid basic repeat antigen (ABRA), Serine Repeat Antigen (SERA) and 6-Cys s48/45 protein family. The 6-Cys s48/45 include the Pf12, Pf41and Pf113. Ultimately, no well-defined functional role has yet been ascribed to these Pf blood-stage 6-Cys proteins, however, they have been recognized by immune sera from individuals naturally infected with malaria (Tonkin *et al*., 2013).

The microneme and rhoptry proteins which are secreted following reorientation play a role in secondary interaction to activate the process of invasion after the initial contact. They include the Apical Membrane Antigen 1 (AMA-1), the reticulocyte binding like protein homologues (RBLs) and Erythrocyte Binding like proteins (EBLs) as shown in Table 1.2 (Camus and Hadley 1985; Cowman and Crabb 2006). AMA-1, which is a trans-membrane protein located at the apical end of the merozoite and released prior to invasion. It plays a role in reorientation but also important in tight junction formation together with the RON protein family (Triglia *et al*., 2000). *P. falciparum* reticulocyte binding protein homolog (*Pf*Rh) which has five members (*Pf*Rh1, *Pf*Rh2a,

*Pf*Rh2b, *Pf*Rh4 and *Pf*Rh5) belongs to RBL family and are also important for secondary interactions with responsibilities for host-cell preferences (Triglia *et al*., 2000).

Table 1.1: Properties of some GPI anchored merozoite proteins and their ligands

GPI Anchored Surface Proteins	
Name	Features, Structure and receptors
MSP-1	Proteolytic processing and removal of bulky complex is essential for invasion. Has two Epidermal Growth Factor (EGF) like domains. The C-terminal domain Binds to protein band 3 on the host cell (Kauth et al., 2003)
MSP-2	It is highly polymorphic; two major alleles functionally identical; potential species-specific function
MSP-4	C-terminal single EGF domain
MSP-5	Not required for invasion; homolog of MSP-4 Surface
MSP-10	Surface and apical appearance; C-terminal double EGF module
Pf12	Member of 6-cys family found on the surface-only
Pf41	Cysteine-rich surface protein
Pf*113*	Putative surface protein

*Pf*Rh proteins were identified as orthologues of rhoptry proteins in *P.yoelii* and *P.vivax* and they function cooperatively with EBLs in invasion of human erythrocytes (Galinski *et al*., 1992; Preiser et al., 2002 ; Lopaticki *et al*., 2011).

The duffy binding-like (DBL) protein family, belong to the super family of EBLs (Miller *et al*., 2002; Triglia *et al*., 2000) and are characterized by a conserved cysteine-rich domain of approximately 35kDa that has been shown to directly mediate binding to host cell receptors (Gilberger, 2003). DBL proteins include Erythrocyte Binding Antigen (EBA) -175 (Camus and Hadley 1985), EBA-140, EBA-181 and EBA-165 (Gilberger, 2003). EBA175 is cardinal for junction formation signalling during invasion (Duraisingh *et al*., 2003). Sialic acid on glycophorins

of the erythrocytes are involved as receptor recognition for merozoite invasion after initial attachment. The receptors for EBA-175, EBA-140 and EBA-181 are glycophorin A, Glycophorin C and a trypsin resistant receptor respectively (Cowman and Crabb 2006). Furthermore, malaria merozoites can utilize independent pathways for invasion without sialic acid (Cowman and Crabb 2006; Kang *et al*. 2012).

Table 1.2: Properties of merozoite microneme and rhoptry proteins

Microneme Proteins	
Name	Features, Structure and receptors
AMA-1	PAN domains, polymorphisms surround conserved hydrophobic pocket (Triglia *et al*., 2000)
EBA-140/BAEBL	Binds glycophorin C (Gerbich antigen)(Maier et al., 2003)
EBA-175	Binds to glycophorin A; disruption in *P. falciparum* (W2mef) leads pathway switch to Rh4-dependent invasion/''handshake'' association between region II dimers creates grooves for GlyA glycan binding (Gilberger 2003)
EBA-181/JESEBL	Receptor for the RBCs is unknown.
EBL 1	Binds to glycophorin B (Mayer *et al*., 2009)
Rhoptry proteins	
Rh1,2a 2b,3, 4,5	Receptors for Rh1, Rh2a and Rh2b are unknown. Rh3 is a pseudogene. Rh4 and Rh5 mediate erythrocyte invasion through Complement receptor 1 and Basigin receptors, respectively (Cowman *et al*., 2012; Tham *et al*., 2010; Wright *et al*., 2014).

1.5 Immunity to invasion by merozoites

Clinical immunity to malaria is slow to develop and is short lived. One reason for this is the extensive diversity found in *Plasmodium* antigens, which facilitate parasite escape from host immune detection (Cowman *et al*., 2012; Marsh and Kinyanjui 2006). But still single infections

are however important in developing initial low levels of immunity (Langhorne *et al*., 2008). With repeated infection following recovery and increase in age, there is a shift from severe malaria to clinical malaria and eventually to asymptomatic parasitemia. Malaria exposure and also age have an effect on the natural acquisition of immunity (Langhorne *et al*., 2008).

Previous studies have shown that immunity towards the parasite is mounted to various stages of the parasite life cycle, sporozoite, schizont, merozoite and infected erythrocytes (Langhorne *et al*., 2008). But, because of its uniqueness towards its association with the disease clinical symptoms, the asexual stage merozoite invasion of erythrocytes represents an attractive target for both drug and vaccine endeavours towards malaria. The merozoite surface proteins, and in particular, MSP1 which is abundantly expressed on the merozoite surface, is one of the targets of immune attack by antibodies and has been shown to elicit both humoral and cell mediated immune responses during exposure to natural infections and in experimental immunization in mice and monkeys (Chang *et al*., 1996; Guevara Patiño *et al*., 1997). In some cases these antibodies have been shown to inhibit erythrocyte invasion. Thus, the selection pressure of the immune system on these proteins has been proposed to maintain the polymorphisms in these proteins (Ngoundou-Landji *et al*., 2010).

1.6 *Plasmodium falciparum* MSP1

Pf MSP1 is a high molecular mass protein. It is largely disordered in solution, but has a propensity to form amyloid-like fibrils under physiological conditions (Low *et al*., 2007). This protein is coded for by the *msp1* gene that is found in the middle of chromosome 9 of the parasite DNA. MSP1 is attached through its N-terminus to a GPI anchor. It interacts with the RBC receptor,

protein band 3, for invasion. It is synthesized as a large precursor during schizogony and is subsequently processed via proteolytic cleavage into four major polypeptides of approximately 83, 30, 38, and 42kDa from the N-terminus to C-terminus. During the invasion process, the C-terminal 42kDa fragment ($MSP1_{42}$) is further processed into 33kDa ($MSP1_{33}$) and 19kDa ($MSP1_{19}$) fragments (Figure 1.4), and the latter remains on the merozoite surface and is carried into the invaded erythrocytes, but all the other fragments are released from the merozoite surface (Cheong et al., 2013). *Pf* MSP1 is schematically organized into 17 blocks based on amino acid sequence divergence (Figure 1.4). Studies have also revealed that there is an existence of interspecies conserved blocks (ICBs) in MSP1 containing polymorphisms in the central variable region in different types of *Plasmodium* species (Escalante *et al.*, 1998; Kang *et al*., 2012 Mwingira *et al*., 2011).

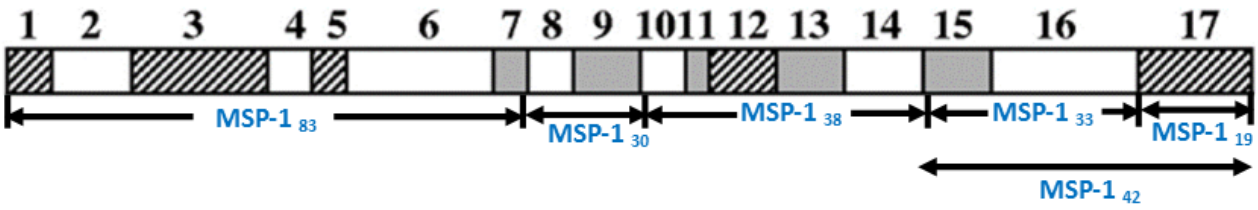


Figure 1.4: Mosaic arrangement of *P. falciparum* MSP1 protein

The sequence blocks are marked, the conserved blocks being hatched, semi-conserved blocks are filled, and oligomorphic regions are open. Sites where the precursor protein is cleaved to its major processing fragments MSP-1_{83}, MSP-1_{30}, MSP-1_{38}, and MSP-1_{42} as defined by Mazumdar et al., 2010, Roy et al., 2008) are indicated by arrows. A secondary proteolytic process cleaves MSP-1_{42} into MSP-1_{33} and MSP-1_{19}. The latter protein is anchored in the membrane of the parasite via a GPI moiety at its N-terminus.

Substantial data from studies with *Pf* MSP-1 and *in vivo* immunization studies of mice with other *Plasmodium* species has indicated that the protective immune responses are specifically directed against the C-terminal 19-kDa domain (Cheong *et al.*, 2013; Qari *et al*., 1998; Moncunill *et al*., 2013).

1.7 *Plasmodium falciparum* MSP1$_{19}$kDa

This is the smallest fragment that is attached to the merozoite at its N- terminus. The C-terminus of the MSP1$_{19}$ fragment found in block 17 (Figure 1.4) exhibits a remarkable structural and functional conservation across malaria parasite species. It has been shown to be the part of MSP1 that directly interacts with the G- coupled protein band 3 receptor on the erythrocyte. This fragment is folded into two epidermal growth factor (EGF)-like domains stabilized by 6 disulfide bonds.

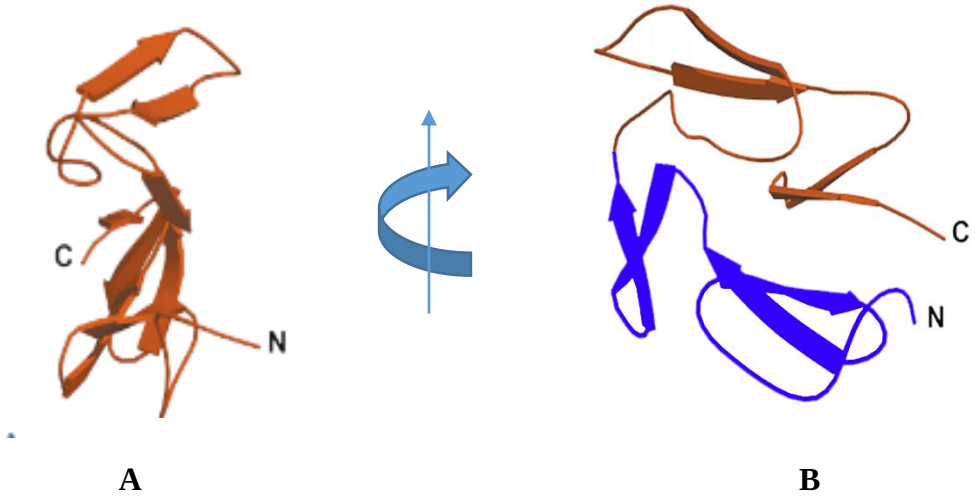


Figure 1.5: Crystal structure of MSP1$_{19}$kDa fragment.
An orthogonal view of the tertiary structure of block 17 of MSP-1 gene at 1.8 Å. The β-pleated sheats are shown as ribbons with an arrow for direction (B) shows the EGF-like domain 2 in red, and EGF-like domain 1 in blue. These domains are covalently bound (Pizarro et al., 2003; Chitarra et al., 1999).

The second EGF-like motif plays the binding role to the respective erythrocyte receptor. Antibodies to this region can block erythrocyte invasion *in vitro* and are associated with protection from clinical malaria in field studies (Egan *et al*., 1996; John *et al*., 2004).

The sequence of *Pf*MSP1$_{19}$kDa is highly conserved, however, six non-synonymous single nucleotide polymorphisms (SNPs) have so far been documented at amino-acid codon positions

1644, 1691, 1699, 1700, 1701, and 1716 (Takala *et al*., 2007) (Table 1.3), and it is unclear how these polymorphisms affects immunity towards malaria.

Since *Pf* MSP1$_{19}$kDa elicits immune responses (Guevara Patiño *et al*., 1997) in natural and experimental malaria infections (Cheong *et al*., 2013) and has been shown to have multiple polymorphisms it has been widely studied as a practical target for vaccine development. Such a vaccine would at least have an element focusing on preventing the interaction between merozoite surface ligands and the erythrocyte receptors, a short period in which the parasite is vulnerable to attack by the immune system (Cheong *et al*., 2013; Ford *et al*., 2007; Kim *et al*., 2014). However, the superlative approach is to have such a multi-allelic region incorporated with other antigens from the various stages of the parasite's complex life cycle (Osier *et al*., 2010; Smith *et al*., 2014) in order to have a more effective vaccine.

Studying *Pf* MSP1$_{19}$kDa in different geographical malaria endemic areas add on the existing information about this domain and will provide further knowledge on its population genetics to inform future malaria vaccines, this formed the core rationale for this study.

Table 1.3: Comprehensive MSP1$_{19}$kDa haplotypes in the literature to date

Comprehensive list of MSP-1 $_{19}$ Haplotypes								
	Amino Acid Position							
Haplotype	1644 (E/Q)	1691 (T/K)	1699 (S/N)	1700 (S/N)	1701 (R/G)	1716 (L/F)	Isolate/Country	Reference
1	Q	K	S	N	G	L	FVO, Wellcome	(Qari *et al*., 1998)
2	E	T	S	S	R	L	3D7, MAD20	(Takala *et al*., 2007)
3	E	K	S	N	G	L	FUP,Uganda-PA	
4	Q	K	S	N	G	F	T807, Thai	(Simpalipan *et al*., 2014)
5	Q	T	S	S	R	L	Indo	
6	E	K	S	S	R	L	Kenya-2	(Qari *et al.*, 1998)
7	E	K	N	N	G	L	Kenya	″
8	E	K	S	S	G	L	Kenya-1	″
9	E	K	S	N	G	F	Kenya-3	″
10	Q	K	N	N	G	L	Thai Variant-1	
11	E	T	S	S	G	L	India	
12*	E	T	S	S	R	F	Brazil-1	
13*	Q	T	S	S	R	F	Brazil-2	
14	E	T	S	N	G	L	Vietnam	
15*	E	K	S	Y†	G	F	India	
16*	Q	K	S	S	G	L	Mali-1	(Takala *et al*., 2007)
17*	Q	T	S	S	G	L	Mali-2	
18*	Q	K	S	S	R	L	Mali-3	
19*	Q	K	S	N	R	F	Mali-4	
20*	Q	K	S	S	G	F	Mali-5	
21*	E	K	S	N	R	L	Mali-6	
22*	E	K	N	N	G	F	Mali-7	
23	E	T	S	N	G	F	Kenya-4	(Kariuki *et al*., 2013)

Asterisk (*) shows haplotypes that have not yet been found in Kenya

E- glutamic acid *K- Lysine* *R-Arginine* *F-Phenylalanine*
Q-Glutamine *S- Serine* *G - Glycine*
T- Threonine *N-Asparagine* *L- Leucine*

1.8 Justification

Malaria has plagued humans for many thousands of years, and despite attempts to eradicate it through concerted programs of antimalarial drug treatment as well as vector control, the causative agent has adapted to these challenges. Developing a vaccine towards malaria would be the most practical way to arrest this quagmire. Since merozoite invasion is one of the stages in the parasite life cycle that is susceptible to immunity, it is an area that has been given attention by many researchers. However, efforts to develop vaccines based on blood stage antigens of the parasite have been slow and unsuccessful. Most of the vaccines for malaria are based on the *P. falciparum* 3D7 laboratory isolate variant, which is not always present or is at a low frequency in natural populations, leading to failure of the vaccine if there is allele-specific immunity.

Polymorphic variant analysis of MSP1 in multiple sequential infections would help in the determination of variant changes in the infections and if immunity to the variants offers protection to successive infections, thereby potentially influencing the variant type in the next infection. It is expected that due to immunity, multiple variant infections will be observed. This study defines variants in multiple infections and provides data on the potential impact of an msp1 vaccine on allelic changes in a population. This would inform the number of msp1 alleles for inclusion in a multi-allelic vaccine formulation.

1.9 Research question and hypothesis

1.9.1 Research questions

Are there *Plasmodium falciparum* $MSP1_{19}$ polymorphic differences in sequential clinical malaria infections?

1.9.2 Research hypothesis

There are no *Plasmodium falciparum* $MSP1_{19}$ polymorphic differences in sequential cases of clinical malaria infections.

1.10 Objectives

1.10.1 General objective

To determine *Pf* $MSP1_{19}$ genotypes in parasites obtained from individuals with multiple malaria infections during a single malaria season.

1.10.2 Specific objectives

- To identify *Pf* $MSP1_{19}$kDa gene polymorphisms in sequential clinical malaria infections.
- To determine allelic changes in individuals across the infections.
- To determine the population proportion of allele frequencies between infections.

CHAPTER 2

MATERIAL AND METHODS

2.1 Study design

The study was conducted in a longitudinal set up that measured allelic patterns of the *Pf* MSP1 gene in every infection from individuals in a cohort of children who experienced at least 9 malaria serial infections. This was the minimal number of multiple infections that gave a sufficient number of individuals for inclusion in this study which aimed to look at the impact of multiple infections in a malaria season. This sample set helped in determining if different variants are observed in multiple infections. This work was under a wider research of integrated studies in natural immunity to malaria.

2.2 Study site

This study was conducted in Junju location in Kilifi county,Kenya (Figure 2.1). The county is a moderate to high malaria transmission zone with high malaria transmission rate ranging from 20 to 100 infective bites per person per year (Mbogo *et al*., 2003). However, some recent studies have shown a reduction in transmission (Okiro *et al*., 2007). Junju is located in the southern part of the county (Figure 2.1b) and is a moderate to high malaria transmission area than the northern parts of the county (Mwangi *et al*., 2005). The county has a total area of 12,246 km² with a population of 1,109,735 inhabitants according to the 2009 census (county government of Kilifi, 2013). The 2008 Kilifi district report on short rainfall assessment shows that the county has a hot and humid

climate throughout the year, mean daily temperature ranges from 22.5^{0}C to 34^{0}C and a bimodal rainfall pattern. Thus, malaria in this region is seasonal coinciding with the rainfall.

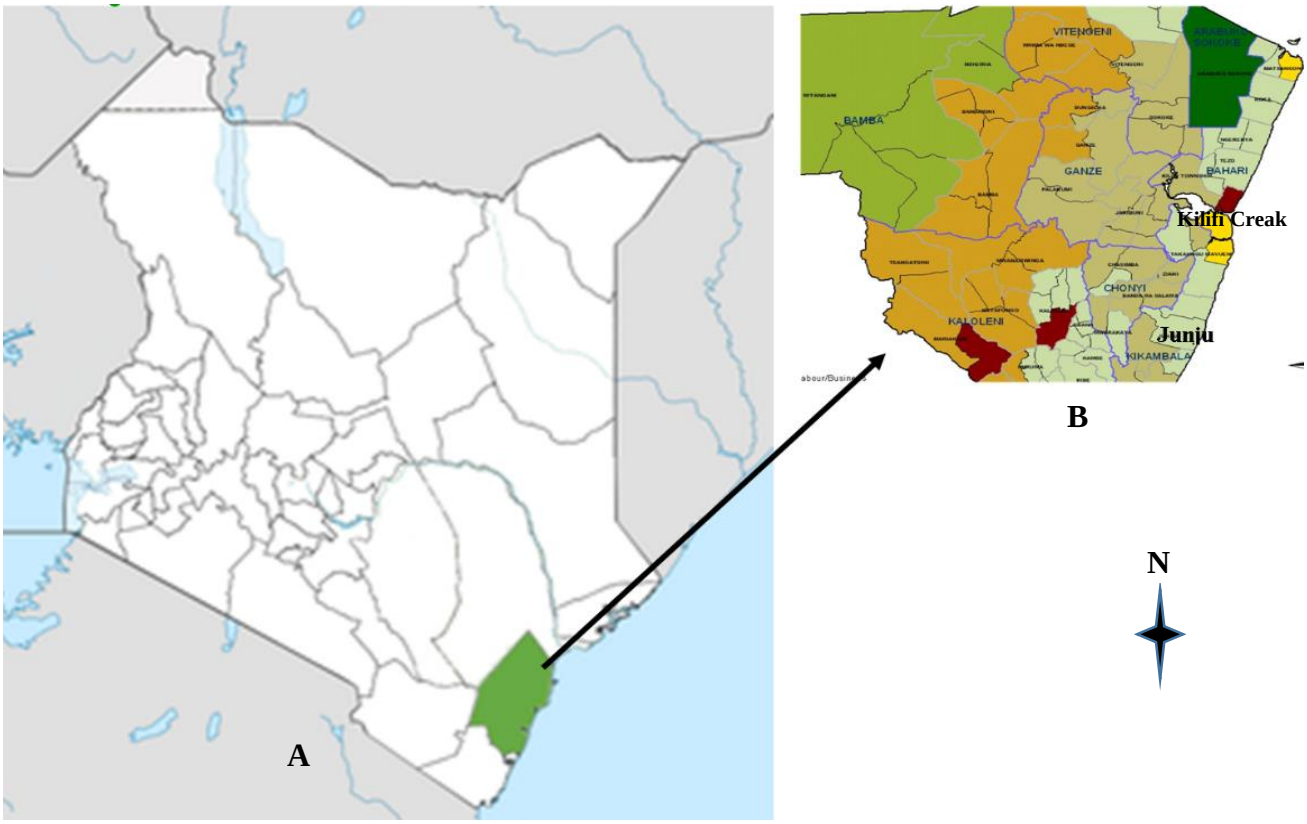


Figure 2.1: Map of Kenya showing Kilifi county and Junju
(A), is the Map of Kenya that shows county demarcations. The area marked green is Kilifi County which falls in the Coast region. (B), is the southern part of the county separated by the creek, and it shows the study site, Junju, which is found in Kilifi South sub-County.

2.3 Study population

Samples were collected in 2008 from children aged below 5 years during the malaria season. These participants regularly visited Kilifi district hospital through referral from Junju dispensary malaria study site every time they showed up with malaria symptoms. Venous blood was drawn from the subjects into indexed heparinized tubes and blood smears done to confirm presence of malaria

parasites. If positive for malaria, the samples would be stored at -80^{o}C for the study. A total of 33 children who met the inclusion criteria of having at least 9 malaria infection episodes were included in the study, and their respective samples subjected to successive analysis. A total of 421 samples were collected all together from the eligible study population.

2.4 DNA extraction

Total DNA from the blood samples was extracted using QIAamp DNA and Blood Mini kits as per manufacturer's instructions (Qiagen). The protocol was based on using proteinase K (Qiagen) to lyse and elute the cellular components of the whole blood while nucleic acid was adsorbed onto the QIAamp silica-gel membrane in a brief centrifugation step. Salt and pH conditions of the elution buffer of the lysate ensured that protein and other contaminants, which could inhibit polymerase chain reaction (PCR) were not trapped on the QIAamp membrane. DNA bound to the membrane was then washed in centrifugation steps using buffers which came with the kit. This was to significantly improve the purity of the extracted DNA by removing any residual contaminant. The purified DNA was eluted into labelled 1.5ml eppendorf tubes and stored at -20°C following 1:10 dilution using the QIAamp kit elution buffer.

2.5 Amplification of 19kDa region in *Pf* MSP1 gene

Pf MSP-1_{19}kDa region from the total DNA, which is approximately 250bp, was amplified by PCR using gene specific primers for MSP-1_{19} kDa region (Table 2.1) and Expand High Fidelity PCR system (Roche). This system is composed of a unique enzyme mix containing thermo-stable Taq and Tgo DNA polymerases with proofreading ability, polymerization and exonuclease activity.

Table 2.1: Amplification and sequencing primers for $MSP1_{19}$kDa region on MSP-1 gene

$MSP1_{19}$kDa region Primers			
Primer ID	**Oligonucleotide sequence**	**%G C content**	**melting temperature**
F1	5′-CAATGCGTAAAAAAACAATGTCC-3′	34.78%	50°C
R1	3′-TTAGAGGAACTGCAGAAAATACCA-5′	37.5%	52°C

Prior to running the PCR, thawed aliquots of homogenized reagents for PCR were used to prepare the master and the Taq mixes separately in amounts that would sum up to a final PCR reaction volume of 10µl per sample as seen in Table 2.2. Reagent preparation for the two mixes was done in two clean sterile 1.5ml eppendorf tubes on ice. The ice was used to keep the temperatures low to prevent the enzyme from starting the reaction until the mixture is taken to the thermo-cycler. The separate preparation of the mixes was intended to avoid the degradation effect of Taq polymerase nuclease activity on primers before the start of the reaction. The mastermix contained 25mM $MgCl_2$ labelled as buffer 4 in the Taq PCR kit (Applied Biosystems, USA), deoxynucleotide triphosphates (dNTP) mix for the bases (10mM of dGTP, dATP, dTTP and dCTP)s, 10µM primers and double distilled water (Sigma, UK) in volumes shown in Table 2.2. The Taq reaction mix contained the Expand High Fidelity Enzyme solution, Expand High Fidelity Buffer (10X) with 15mM $MgCl_2$ and double distilled water (Sigma, UK) in reaction volumes shown Table 2.2.

Table 2.2: Volume of PCR requirements in microlitres

PCR Mix	Volume in µl for PCR reaction requirements in PCR tube per sample								
	DD H_2O	Buffer4	dNTPs	Forward Primer	Reverse Primer	Taq	Buffer2	DNA template	Total
Master Mix	2.7	1.0	0.2	0.3	0.3	-	-	-	4.5
Taq Mix	3.87					0.13	1.0	-	5.0
								0.5	0.5
Total volume in PCR tube									**10.0**

DDH2O - Double Distilled water.

The amount of DNA template added was increased to 0.7 µl and 1.0 µl for repeats on samples that had not amplified on their previous runs.

All the PCR reactions were separately done in 0.2ml PCR tubes for every test sample in the thermocycler. However, prior to starting PCR, the mixes were vortexed and centrifuged to ensure homogeneity. A PCR optimization process was first performed by gradient polymerase chain reactions for the *Pf* MSP1 genes using *P. falciparum* 3D7 culture type as the positive control and double distilled water as negative control. The aim was to identify the desirable amplification conditions for the primers that could achieve a single and clear band characterizing good yield of PCR products. The amplification protocol chosen after trying a range of amplification conditions was as follows, DNA denaturation for 2 minutes at 94^{o}C, followed by 9 cycles of 24 denaturation at 94^{o}C for 15 seconds, primer annealing at 44^{o}C for 30 seconds and then an extension step at 72^{o}C for 2 seconds. The round of 9 cycles is followed by other 24 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 physical and chemical conditions.

Since the optimal amplification conditions for the target region of the *Pf* MSP1 gene using its specific primers was determined, it was set and saved as the PCR protocol for this study in the thermocycler. PCR for the 421 samples was then run in batches of between 22-30 samples including negative and positive controls. Distilled water and *Plasmodium falciparum* 3D7 were used as negative and positive controls, respectively. PCR mixes were therefore prepared using the protocol described in Table 2.2 for the batch size with one or two extra virtual samples to have excess reagent volumes. These extra volumes catered for any amount lost during pipetting. Volumes of 4.5µl of the master mix, 5.0µl of the Taq mix and 0.5µl of the test samples were aliquoted to each indexed PCR tube, respectively. These tubes were then inserted into the wells of the thermo-cycler (MJ Research PTC-100), and the PCR protocol was selected which lasted for 2 hours and 15 minutes.

2.6 PCR products analysis by agarose gel electrophoresis

After PCR, the amplicons were assessed for their quality by searching for bands in their respective lanes upon separating them using electrophoresis in 1% agarose gel. Any sample that had a single clear band in its lane upon viewing gel was considered positive in comparison to the controls. Samples that had very faint bands, smears or no bands at all were considered negative. PCR process was repeated for these negative samples by improving DNA amplification specificity through DNA template amount adjustment from 0.5µl to 0.7µl and 1.0µl, and increasing the annealing temperature slightly to 45°C while monitoring their bands. If still they had no bands, then they were confidently declared negative.

The gel was first prepared using previously made 1X TBE (Tris-Borate- EDTA) buffer at pH 8.0. The buffer was made briefly as follows, first, was the preparation of a half a litre of 0.5M EDTA solution with pH 8.0 by re-suspending 93.05g of EDTA (Ethylenediaminetetraacetic acid) disodium salt in 400ml of deionized water on a magnetic stirrer. An amount of 9g of NaOH (Sodium Hydroxide) was then added while still stirring. The pH was adjusted and monitored to 8.0 using a pH meter by slowly adding 10N NaOH to the solution. Achieving the pH of 8.0 would ensure complete dissolving of the EDTA. The solution was then topped up to a volume of 500ml using double distilled water.

In a clean sterile bottle, a mixture 40ml of earlier on made 0.5M EDTA solution and 800ml of deionized water was used to dissolve 108g of Tris base and 55g of boric acid using a magnetic stirrer. The final volume was adjusted to litre by adding deionized water to make 10X of 1000ml TBE buffer.

The TBE working solution was regularly prepared in a volume of 500ml by topping up 50ml of 10X TBE stock solution to 500ml using deionized water to achieve a concentration of 1X in a clean sterile bottle.

A 1% (w/v) agarose was made by dissolving 0.3g of agarose powder (from Applied Genetic Technologies Corporation) in 30ml of 1X TBE buffer in a conical flask. This was boiled on a hot plate to ensure complete dissolving of the powder, then cooled using running water on the flask from the tap. Before the gel set, 2µl of SYBR® safe for DNA staining (from life Technologies) was added and the conical flask swirled to ensure even mixing. The lukewarm agar was poured in a level casting gel tray that had gel combs placed on one end to form wells for loading samples. This was left for 20 minutes for the agar to gel, after which the comb was removed and gel

submerged in the gel electrophoresis tank with TBE buffer (1X) as the electrophoresis running buffer. Mixes of 2µl for each PCR product and 2µl of the loading dye (6X Blue Orange from Promega) were loaded in the formed wells alongside 4 µl (microliter) of 1kb DNA ladder (HyperLadder 1, Bioline UK). The ladder was used as the standard DNA marker. Electrophoresis was done for a period of 30 minutes at 94 volts and at current of 400mA. To visualize the bands, a digital photography under UV (ultraviolet) light using The Molecular Imager Gel Doc (Bio-Rad., UK) was used. The images were captured, edited and saved in SCF (Simple configuration files) and PNG (Portable network graphics) picture formats. Positive samples were purified for sequencing.

2.7 Amplicon purification

Ethanol and Sodium Acetate precipitation method was performed on the successfully amplified PCR products to remove remnants of primers, dNTPs and buffers that hinder sequencing process. For purification, the PCR products were transferred from the PCR tubes to 96 well plates. The purification method involved constituting a purifying premix for the total number of amplified products to be purified plus an extra sample volume to cater for volume losses especially for ethanol which evaporates during pipetting. The premix constituted 3µl of 3M Sodium Acetate (Ambicon), pH 5.2, 62.5µl of 95% ethanol (Sigma-Aldrich, UK) and 24.5µl of distilled water making a final volume of 90µl per well. Into each well containing the PCR product, 90µl of the Ethanol and Sodium premix was added. After addition of the premix into each well, the plates were then sealed with micro-seals (Bio-Rad, USA) and incubated at -20°C for 30 minutes. After which they were spun at 3000 x g for 30 minutes at 4°C on a centrifuge (5810R bench centrifuge, Eppendorf). Seals were removed and plates were overlaid with absorbent paper towels then gently

inverted to drain them. Fresh paper towels were then placed on the inverted plates and spun at 50 x g for 1 minute at 4◦ C. A volume of 150µL of ice cold (-20◦ C) 70% ethanol was added into each well, then the plates were sealed and spun at 3000 x g for 10 minutes at 4◦C, after which they were inverted again over paper towels and excess fluid gently drained. The plates were overlaid with clean paper towels, inverted and spun at 50 x g for 1 minute at 4◦ C. Finally, the plates were covered with fresh paper towels and left on the bench to air dry. After this clean-up process, 16µl of deionized water was added to each well to dilute the amplicons.

2.8 Sequencing

Sequencing services were outsourced from Inqaba Biotechnical Industries (Pty), South Africa Since the approximate size for MSP 1_{-19}kDa is 250bp, therefore, the sequencing primers described in (Table 2.1) covered the whole desired region considering the fact that sanger read length are of approximately 800 bp (Schatz, 2010).

The purified samples were first amplified using fluorescent BigDye Terminator v3.1 chemistry and sequencing primers. The subsequent products were later purified, reconstituted in HiDi formamide reagent and then denatured by heating the plates in a thermocycler in readiness for analysis by capillary electrophoresis using Applied Biosystems 3130xl genetic analyser. Incorporation of any of the four ddNTPs which were coupled to a special fluorescent dye of a different colour during BigDye PCR terminated the elongation process forming the basis of analysis. The fluorescently labelled DNA fragments, moves across the path of a laser beam and causes the dyes attached to the fragments to fluoresce and emit light at a different wavelength. The dye signals are separated by a diffraction system, and a charge-coupled device (CCD) camera

detects the fluorescence and converts it to digital data in form of a chromatogram. The fluorescently labelled terminated products from the sequencing reaction migrate through the polymer and as they pass the CCD camera which contains a laser detector, the fluorescently labelled products are detected as a colour signal. The DNA fragments are assembled by the analyser based on their sizes and a chromatogram sequence of the product is generated.

2.9 Sequence editing, assembling and alignments

The trace files generated by the DNA analyser sequencer from Applied Biosystems were imported into Sequencher software version 5.3 for quality check, editing and assembling. Samples that generated chromatograms with low peaks were re-sequenced. Following this procedure, all the trace files were trimmed off to remove low quality bases at the extreme ends of the chromatograms using the default settings of the software.

For each sample, trimmed sequences for the reverse and forward orientations were assembled into contigs against *P. falciparum* 3D7 MSP-1 sequence accession number PF3D7_ 0930300 downloaded from PlasmoDB (www.plasmodb.org) as the reference for scaffolding using the default assembling algorithm. Corrections of bases on the resultant contigs were manually done to ensure the correct base calling for the subsequent consensus sequences. These ambiguities and disagreements were resolved by referring to the chromatogram files of the respective reads in that consensus sequence and independent of the reference sequence. Clean scaffolds were saved as consensus files in FASTA (fast all) formats.

A multiple alignment of the consensus files for all samples and the reference was carried using the MUSCLE (Multiple sequence alignment) algorithm in the MEGA 5.2 program (Tamura *et al*.,

2011). Misaligned sequences were corrected manually and gaps in the aligned region resolved. The clean alignment was saved in FASTA format.

2.10 Sequenced data analysis

The saved multiple alignment in FASTA formats were imported in DnaSP V 5.10 for nucleotide polymorphisms identification (Rozas *et al*., 2003).

The same FASTA multiple alignment file was exported to CLC® sequence viewer version 7 for further analysis. Sequences were trimmed to the same size. Sequences that had short read-lengths resulting in gaps at the outermost polymorphic sites were excluded. This is because, gaps tend to influence clustering results falsely. The successive sequences were realigned again against the reference and translated to amino acid sequence using the correct reading frame. This was to detect if Single Nucleotide Polymorphisms (SNP) caused a change in the amino acid translated from the codon. The DNA sequence alignment was exported in PHYLIP (Phylogeny inference package) format and the amino acid alignment exported in FASTA format. These file formats were used for haplotype clustering and frequency determination using Phyclust package (Tzeng *et al*., 2005) in R-programming and confirmed by command line in executable USEARCH (unique search) clustering algorithms (Edgar, 2010).

2.10 Ethical consideration

Written informed consent was obtained from the patient's guardian, parent or next of kin before sample collection. The whole study was approved by the Ethics Review Committee of the Kenya Medical Research Institute (Protocol number, SSC 1131).

CHAPTER 3

RESULTS

3.1 MSP1$_{19}$kDa amplification

This study was conducted on 421 DNA samples, and out of this 271 samples were successfully amplified. The remaining 150 samples were not amplified despite repeated PCR amplification attempts using a gradual increase in the template volumes. This could be due to low template concentrations and/or template degradations. The size of the amplified gene fragment was 250bp (Figure 3.1) in relation to the band size distribution of the 1kb DNA hyperladder. The use of controls ensured a quality check for the PCR reagents. There was no band on the negative control lane as opposed to the positive control lane (Figure 3.1.) This was an assurance that the PCR reagents were not contaminated.

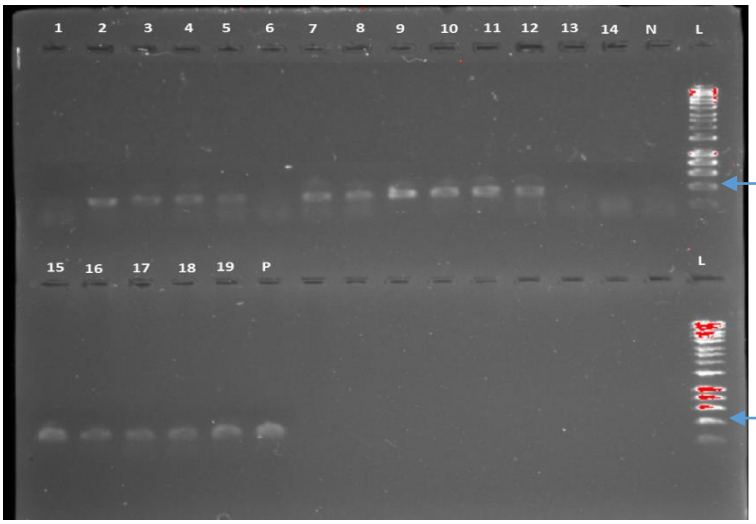


Figure 3.1: Gel image for MSP-1$_{19}$kDa

Bands in lanes 1 to 21 represents amplified products of different samples. The bands in lanes marked N and P represents the negative and positive control samples, respectively. The bands in the lane marked L represents the ladder. Samples in lanes with no bands did not amplify, hence, regarded negative.

3.2 Sequencing results

Nucleotide sequences of *P. falciparum* MSP1 gene block 17 were obtained from 178 samples that were sequenced. These samples yielded a clean chromatogram and comparatively tall peaks. However, only 150 samples assembled correctly against the reference, and had read length which covered the study *Pf* $MSP1_{19}$kDa region from nucleotide positions 4990 to 5208 encoding the C-terminal fragment of the MSP1 protein ($MSP1_{19}$, amino acid positions 1644 to 1716). This was 33.3% of the total 421 parasite isolates collected. The 28 samples (which did not assemble) had short read lengths since one of either of its primer orientation failed to assemble against the other orientation and the reference due to poor peaks, which were quite below the lenient threshold, this resulted in short contigs.

3.3 $MSP1_{19}$kDa polymorphism analysis

Six dimorphic sites at nucleotide positions 4990 G/C, 5132 C/A, 5157 G/A, 5159 G/A, 5161 G/A and 5206 C/T in relation to *P. falciparum* 3D7 as the reference were detected. These nucleotide substitutions resulted in non-synonymous amino acid substitutions at codon positions 1644, 1691, 1699, 1700, 1701 and 1716 as shown in Table 3.1.

Table 3.1: Synonymous and non-synonymous variations detected in $MSP1_{19}$kDa

Position	Ref. Codon	Ref. Amino Acid	Amino Acid Code	Variant Codon	Variant Amino acid	Amino Acid Code
1644	GAA	Glutamate	E	CAA	Glutamine	Q
1691	ACA	Threonine	T	AAA	Lysine	K
1699	AGC	Serine	S	AAC	Asparagine	N
1700	AGC	Serine	S	AAC	Asparagine	N
1701	AGA	Arginine	G	GGA	Glycine	R
1716	CTT	Leucine	L	TTT	Phenylalanine	F

3.3.1 Haplotype frequencies

Based on amino acid replacements, *Pf* MSP1$_{19}$ amino acid sequences for the 150 samples yielded 11 distinct haplotypes as seen in Figure 3.2 all of which have been reported previously.

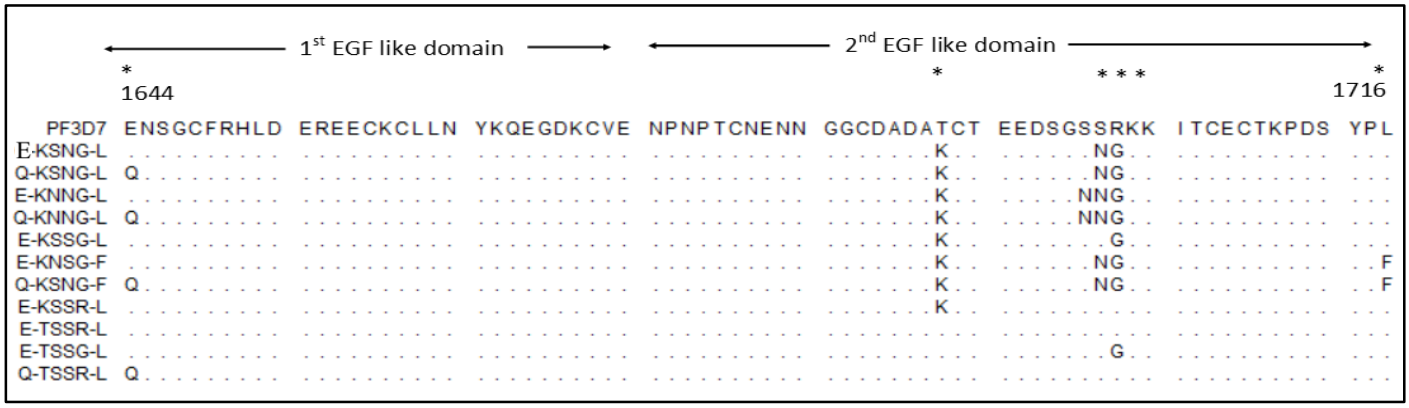


Figure 3.2: Haplotypes of *Pf* MSP1$_{19}$ *from P. falciparum* populations in Kilifi
The arrows indicate the epidermal growth factor (EGF)-like domains 1 and 2 (Kauth *et al.,* 2003) *Asterisks (*) indicate the non-synonymous amino acid substitutions. Numbers above the MSP-1 amino acid sequence of the* P. falciparum *3D7 indicated the positions Dots (.) represent identical amino acid sequences.*

The E-KSNG-L, Q-KSNG-L and Q-KSNG-F haplotypes corresponding to FUP-Uganda PA, FVO Wellcome and Thai (T807) strains of *P. falciparum* respectively, were the dominant circulating haplotypes in the population with frequencies of 0.36, 0.26 and 0.18 respectively. The Q-KNNG-L, E-KSSR-L and E-KNNG-L haplotypes were circulating with frequencies of 0.046, 0.040 and 0.040 respectively. The remaining haplotypes, E-KSSG-L, E-TSSR-L(3D7),Q-TSSR-L, E-TSSG-L and E-KSNG-F had very low frequencies 2.67%, 2.6, 1.3%, 0.6% and 0.6%, respectively, below the truncated horizontal line in the bar plot (Figure 3.3) in and so they were considered as extremely rare circulating haplotypes in this study. The clustering was based on Tzeng's method of evolutionary grouping of haplotypes and then categorises the haplotypes to be retained above the cut-off point (horizontal line in the bar plot Figure 3.3), which is the optimal balance between

the sample size, haplotype number and frequencies, and the net information (Tzeng *et al*., 2005). However, even the rare variants were subjected for further analysis in this study.

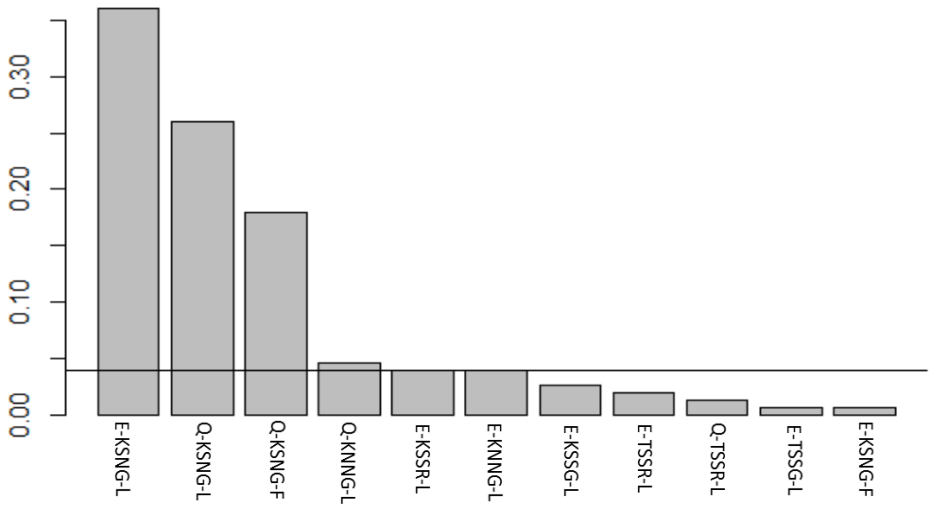


Figure 3.3: MSP1$_{19}$ Haplotype frequencies in decreasing order

Chart generated by Phyclust package for Sequence clustering in R programming language. The Y-axis shows the haplotype frequencies in descending order and the X-axis showing the haplotype name. The horizontal line is the cut-off point of the first 6 haplotypes.

3.3.2 Genetic diversity of *Pf*MSP-1$_{19}$kDa haplotypes

The genetic diversity of the *Pf* MSP1$_{19}$kDa in 150 parasite isolates obtained in 2008 during a malaria season which had a total of 6 segregating sites was defined by its nucleotide and haplotype diversity. The nucleotide diversity (π), mean number of nucleotide substitution of every site in the 150 sequences, was 0.00581 (Standard deviation of 0.00042) While the haplotype diversity (HapDiv), a measure of haplotype uniqueness in a population, for the 11 identified haplotypes was 0.782 (Standard deviation of 0.020) (Table 3.2) Statistical test of neutrality using Tajima's D in DnaSP software were performed to detect if the SNPs were under any selection pressure. The calculated Tajima's D value was 0.58998 (Table 3.2) above the p value of 0.05.

Table 3.2: Genetic diversity at the *Pf* MSP1$_{19}$kDa domain

*Pf*MSP-1$_{19}$-kDa domain							
No. of sequences	**Size(bp)**	**π**	**HapDiv**	**K**	**Tajima's D**	**Fu & Li D**	**Fu & Li F**
150	234	0.00587	0.769	1.37441	0.58998	1.08750	1.08945

π, Nucleotide diversity (standard deviation) , Nhap- Number of haplotypes, HapDiv- Haplotype diversity and K- Average number of nucleotide differences.
For Tajima's D, p>0.10; for Fu and Li's analysis, p >0.05. These values were calculated using DnaSP 5.10.

The same neutrality tests was computed for the target region using a sliding window length of 50 sites long, with a step size of 5 bases (Figure 3.4). This analysis aimed at finding which mutations were influencing the high values within the sequence. The graph showed that the first and sixth segregating sites had Tajima's D values greater than the p value (0.05), which meant the mutation in those sites was neutral, as opposed to segregating sites 2, 3, 4 and 5 which had significant Tajima's values.

3.4 Allelic changes within infections

Out of the 33 children sampled, variant information for 3 children labelled as patient IDs 5, 6 and 7 was not acquired. Out of the 30 children who had at least one variant in their infection episodes,

13 (shown by patient ID with asterisk in Figure 3.5) of them, had E-KSNG-L, Q-KSNG-L and Q-KSNG-F variants, which were dominant alleles in the study. These alleles were observed 2 to 3 times in most of the consecutive infections within the children. The other 8 variants were rare with a prevalence of < 10%. However, the Q-KNNG-L, E-KSSR-L and E-KNNG-L rare variants (retained above the cut-off point based on Tzeng's method of haplotype grouping), were sparsely distributed within the infections in the 30 children.

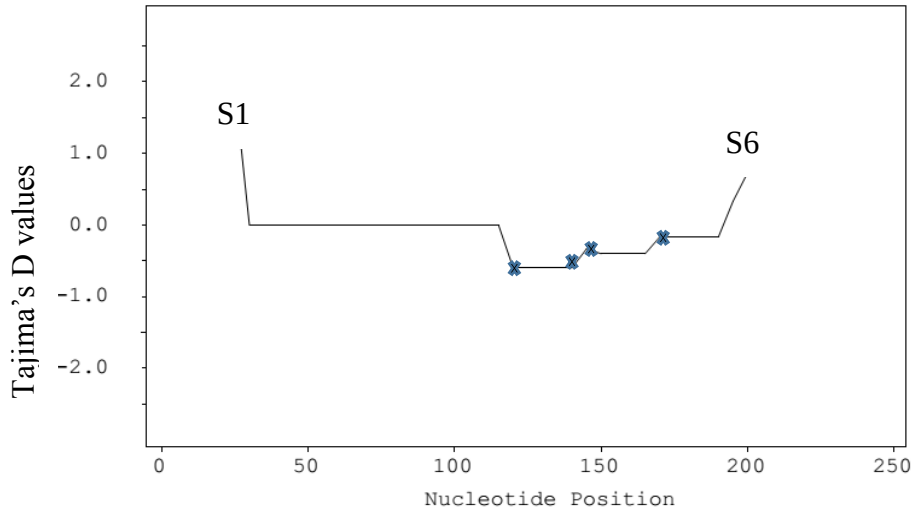


Figure 3.4: Tajima's D window sliding graph

The graph line shows the relationship between the nucleotide position on the x-axis and the Tajima's D values on the y-axis. *The Tajima's value for the region labelled S1 and S6 (segregating site) found between nucleotide positions (1-50) and (170-234) respectively were not significant (p>0.05), While the middle segregating sites marked by X were significant (p>0.05).*

The other rare variants (below the cut-off point), were identified in parasite isolates from 9 children with E-KSSG-L from patient IDs (11, 14, 15 and 17), E-TSSR-L from patient IDs (16, 13 and 29), Q-TSSR-L from patient IDs (14 and 25), E-TSSG-L from patient ID 25 and E-KSNG-F from patient ID 24 (Figure 3.5). In general the allelic changes had no clear trend, pattern or distribution across the infections. However, there were some variant changes in cases where individuals had data of 2 or more consecutive infections.

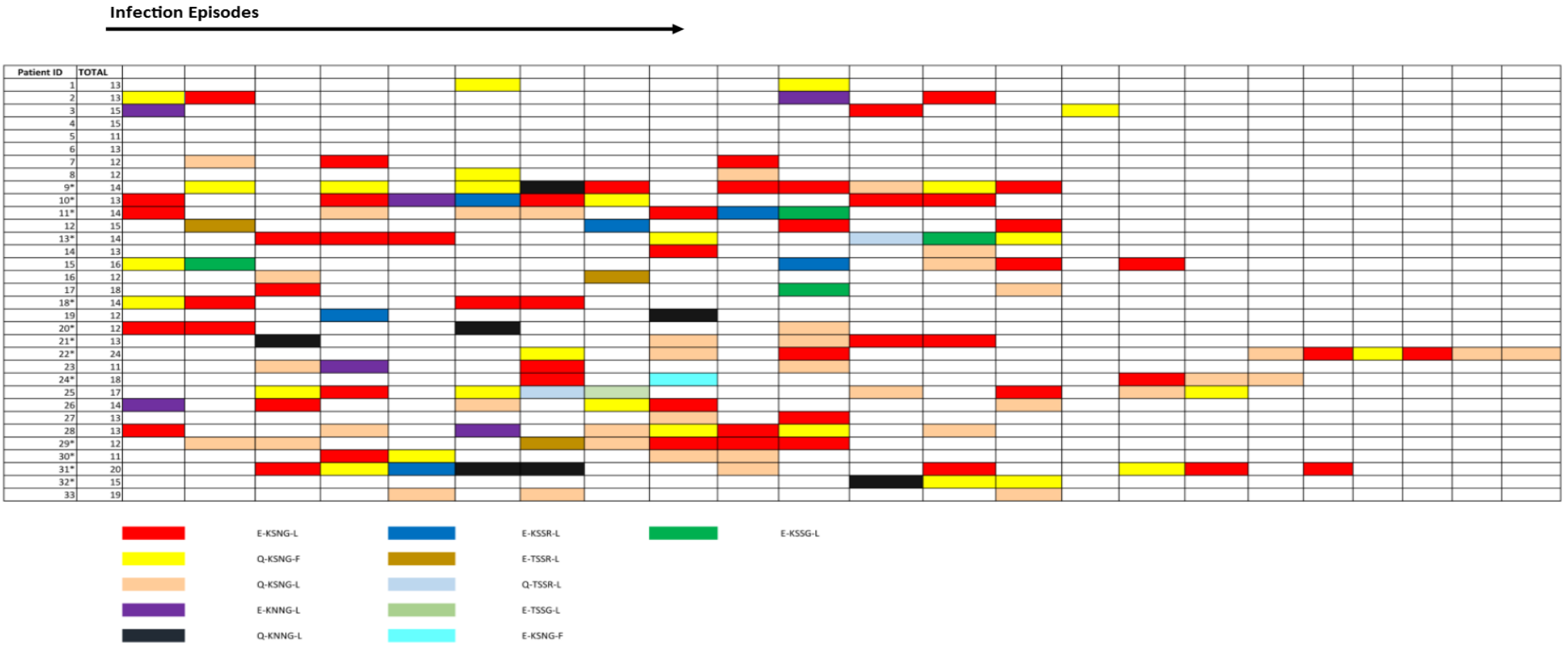


Figure 3.5: Haplotype distribution

Figure shows haplotype distribution within the children and across the infections for all the haplotypes. The empty cells had no variant identified. The first column labelled patient ID is an arbitrary index that identifies the child, while the second column labelled Total, shows the number of infection episodes>9. The colour on every cell represents a variant as described on the key. Patient ID with asteriks(*) show the children who had isolate variant consequtively retained within the infections.

CHAPTER 4

DISCUSSION, CONCLUSION AND RECOMMENDATION

4.1 Discussion

The 19-kDa antigenic domain of *Pf* MSP1 is considered as a major target component for the development of human blood stage malaria vaccines (Wilson *et al*., 2011). Therefore, its sequence diversity has been extensively studied, leading to the identification of its polymorphisms and haplotypes (Simpalipan *et al*., 2014, Takala *et al*., 2007). This, was yet another study whose aim was to define the temporal distribution of variants of this domain in a longitudinal set up in a malaria endemic zone in Kilifi, Kenya using *P. falciparum* parasites obtained from sequential infected children who are less than 5 years.

Pf $MSP1_{19}$ sequences have been previously classified into two allelic groups; one from the 3D7 allelic group (E-TSSR-L haplotype) and the other from Wellcome allelic group (Q-KSNG-L haplotype). Many other haplotypes have been reported from natural *P. falciparum* populations in the world and were proposed to be generated through recombination events between the two allelic groups and due to mutations (Tanabe *et al*., 2007). *Pf* $MSP1_{19}$ analysis in Kenya was previously conducted in the western region and Asembo bay in Nyanza, and at least 14 haplotypes were identified, E-KSNG-L, Q-KSNG-L, Q-KSGN-F, E-KSSR-L, E-TSNG-L, QKNNG-L, E-TSSG-L, E-TSSR-L, Q-TSSR-L, Q-TSNG-L, E-TSNG-F, E-KSNG-F, E-KSSG-L, and E-KNNG-L (Ogutu *et al*., 2009; Qari *et al*., 1998). However, it is not clear whether the extent of *Pf* $MSP1_{19}$ sequence variation in and across infections is common in other *P. falciparum* isolates obtained in other geographical regions of Kenya and if the distribution pattern of these haplotypes is

conserved. This question formed the basis of the present study, which aimed to longitudinally screen for polymorphism and evaluate the persistence of the haplotypes of the MSP1 block 17 in individuals who have multiple malaria episodes from a P. *falciparum* endemic geographical region, Kilifi.

This sequence analysis was performed on *Pf* $MSP1_{19}$ domain from 150 *P. falcipurum* isolates and resulted in the identification 6 SNPs which caused non-synonymous amino acid substitutions. A total of 11 haplotypes of this domain as a result of the 6 SNPs were found to be circulating in the population (Figure 3.2), all of which have been identified before in other areas. These SNPs were shown to be conserved and have been previously reported in field isolates of different studies in Thailand, Cambodia, Mali, Somalia and Tanzania, (Simpalipan *et al*., 2014; Takala *et al*., 2007). In this study, the domain was evolving randomly. This was consistent with previous conducted in Western Kenya (Kariuki *et al*., 2013). The first amino acid substitution (E1664Q) was located between the second and the third Cys residues in the first putative EGF-like motif, while the other substitutions occurred in the second EGF-like motif of *Pf* $MSP1_{19}$ (Figure 3.2).

The high number of haplotypes was expected in Kilifi, since it is an endemic area with moderate to high malaria transmission (Mwangi *et al*., 2005) with an average annual entomological inoculation rate (EIR) of between 20 to 100 infective bites per person in a year (Mbogo *et al*., 2003). This was consistent with other studies done in high malaria transmission areas with high EIRs ranging from 50 to 600 infective bites per person in year resulting 6-15 halpotypes of *Pf* $MSP1_{19}$kDa (Tanabe *et al*., 2007) as opposed to identification of between 1-3 haplotypes in areas of lower transmission and EIRs of less than 10 infective bites per person per year (Sakihama *et al*., 2006). The higher the transmission rate the more recombination occurs during meiosis in the

mosquito, and hence more haplotypes, while the reverse is true in low transmission areas. Therefore, the relative stability of $MSP1_{19}$kDa haplotypes observed in this study suggests that the rates of meiotic recombination in parasite populations in this area remain high despite the reduction in malaria transmission.

The dominant E-KSNG-L (FUP-Uganda PA), Q-KSNG-F (Thai) and E-KNNG-L (K) circulating alleles in this population, have also been reported in Western Kenya, Uganda and Tanzania (Tanabe *et al*., 2007), in such a trend, suggesting that *P. falciprum* populations in East Africa are evolutionary related. Similar distribution patterns of *Pf* $MSP1_{19}$ have also been detected in Thailand, China and Cambodia where E-KSNG-L (FUP-Uganda PA), Q-KSNG-F (Thai) and E-KNNG-L are dominant (Simpalipan *et al*., 2014).

The frequency of the *Pf*MSP-1_{19} haplotype (E-TSSR-L) corresponding to the 3D7 laboratory isolate of *P. falciparum* was detected at a frequency of 2%. Low frequencies of 3D7 *Pf* $MSP1_{19}$ haplotype sequence-types have been reported in previous studies conducted in different malaria transmission areas, in western Kenya, Tanzania and Somalia, and studies conducted recently show a drastic fall in its proportional frequency in a population. It has been suggested that the low frequency of haplotypes corresponding to the 3D7 strain of *P. falciparum* could compromise the efficacy of MSP1 vaccines that are based on the 3D7 clone. The *f*MP1/AS02 vaccine, based on 3D7 parasite (E-TSSR-L) line and target 42kDa of MSP was unfortunately reported to have failed at phase 1 in 2009 at Western Kenya though it elicited greater $MSP1_{42}$-specific immune response (Ogutu *et al*., 2009). It could not provoke immune reaction that would target FUP (E-KSNG-L), FVO (Q-KSNG-L) and Thai (Q-KSNG-L) haplotypes which were prevalent due to high

multiplicity and diversity of the MSP1 gene, hence, it offered no protection against infections caused by other variants other than of 3D7.

There was no clear trend or pattern on how alleles and haplotypes were circulating in the study population and across the infections, but most individuals had multiple malaria episodes with different variants. The dominant E-KSNG-L (FUP-Uganda PA), Q-KSNG-F (Thai) and E-KNNG-L (K) circulating alleles in this population were shown to be sequentially repeated 2 to 3 times in the consecutive infections of 10 children. This could be as a result of slow developing immune responses in these children or non allele-specific immune responses that did not protect against those variants following a prior infection with the same variants. But after this occurence, subsequent isolates from these children in their successive infections had different variants (rare variants). This is in concurrence with other malaria immunity studies that suggests generation and persistence of memory and effector cells for protecting against a variant is dependant on continued exposure to that variant (Langhorne *et al*., 2008). Random variant changes in the consecutive infections of the other children interfered with the immune response development to a variant, hence there was no concrete protection against those variants in succesive infections. This was clear since, all subsequent isolates from these children in their successive infections had the same variants repeating. Most of the rare variants were suppressed towards the end of the infection episodes, this could be due to an initial immune response to the common variants.

4.2 Conclusion and recommendation

Data obtained in this study shows that the $MSP1_{19}$kDa haplotypes change in multiple infections. The changes could be influenced by the host immunity which is gradually developing as a result

of prior infections. This work could support that continued exposure to a malaria antigen plays an important role in production of required antibody titres for protection against that antigen. However, the study did not establish if truly there was an immune response to these variants in the respective sera of the children. Furthermore, we missed variant information for 66.7% of the 421 total parasites isolates collected. Haplotypes identified in this study have remained stable overtime, compared to other previous studies done in Nyanza and western Kenya in 1994, 1996, 1997, 1999 and 2001. This adds current information on the variation of *Pf* MSP1 19kDa in Kenya, which is a potential vaccine candidate.

The information on low frequency of 3D7 *falciparum* clones could be useful in interpreting data on the efficacy of malaria vaccines based on the 3D7 parasite strain in areas of stable malaria transmission. This study essentially suggests that vaccines based on 3D7 laboratory isolates will not work in Kenya, but provides information on the number of MSP1 variants that can be included in multi-allelic vaccine formulations.

A better picture of the infecting haplotypes in these children with multiple malaria episodes and those circulating in the population is required to improve this study. However, we also recommend an immunological evaluation to be done of the sera from the children who we obtained genetic data to determine the association of the immune reactions to these haplotypes. This study suggests the inclusion of common variants (E-KSNG-L, Q-KSNG-F and E-KNNG-L) in MSP1 multi-allelic malaria vaccine formulations.

CHAPTER 5

REFERENCES

Aikawa, M., Miller, L. H., Johnson, J., & Rabbege, J. (1978). Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite. *The Journal of Cell Biology*, *77*(1), 72–82.

Bannister, L. H., & London, C. (2009). Encyclopedia of life sciences, (December).

Bannister, L., & Mitchell, G. (2003). The ins, outs and roundabouts of malaria. *Trends in Parasitology*, *19*(5), 209–213.

Bloland, P. B., Organization, W. H., & others. (2001). *Drug resistance in malaria*. World Health Organization Geneva.

Brown, K. N. (1977). Invasion of erythrocytes and antigenic variation. *Bulletin of the World Health Organization*, *55*(2-3), 187.

Camus, D., & Hadley, T. J. (1985). A Plasmodium falciparum antigen that binds to host erythrocytes and merozoites. *Science*, *230*(4725), 553–556.

Chang, S. P., Case, S. E., Gosnell, W. L., Hashimoto, A., Kramer, K. J., Tam, L. Q., … Hui, G. S. N. (1996). A recombinant baculovirus 42-kilodalton C-terminal fragment of Plasmodium falciparum merozoite surface protein 1 protects Aotus monkeys against malaria. *Infection and Immunity*, *64*(1), 253–261.

Cheong, F., Fong, M., Lau, Y., & Mahmud, R. (2013). Immunogenicity of bacterial-expressed recombinant Plasmodium knowlesi merozoite surface protein-142 (MSP-142). *Malaria Journal*, 1–9.

Chitarra, V., Holm, I., Bentley, G. a., Pêtres, S., & Longacre, S. (1999). The crystal structure of C-terminal merozoite surface protein 1 at 1.8 Å resolution, a highly protective malaria vaccine candidate. *Molecular Cell*, *3*(4), 457–464.

Cowman, a. F., Berry, D., & Baum, J. (2012). The cell biology of disease: The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *The Journal of Cell Biology*, *198*(6), 961–971.

Cowman, A. F., & Crabb, B. S. (2006). Invasion of Red Blood Cells by Malaria Parasites. *Cell*, *124*(4), 755–766.

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasit Vectors*, *3*(1), 5.

Delves, M., Plouffe, D., Scheurer, C., Meister, S., Wittlin, S., Winzeler, E. A., … Leroy, D. (2012). The Activities of Current Antimalarial Drugs on the Life Cycle Stages of Plasmodium: A Comparative Study with Human and Rodent Parasites. *PLoS Medicine*, *9*(2), e1001169.

Duraisingh, M. T., Maier, A. G., Triglia, T., & Cowman, A. F. (2003). Erythrocyte-binding antigen

175 mediates invasion in Plasmodium falciparum utilizing sialic acid-dependent and-independent pathways. *Proceedings of the National Academy of Sciences*, *100*(8), 4796–4801.

Dvorak, J. A., Miller, L. H., Whitehouse, W. C., & Shiroishi, T. (1975). Invasion of erythrocytes by malaria merozoites. *Science*, *187*(4178), 748–750.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461.

Egan, a F., Morris, J., Barnish, G., Allen, S., Greenwood, B. M., Kaslow, D. C., … Riley, E. M. (1996). Clinical immunity to Plasmodium falciparum malaria is associated with serum antibodies to the 19-kDa C-terminal fragment of the merozoite surface antigen, PfMSP-1. *The Journal of Infectious Diseases*, *173*(3), 765–769.

Escalante, A. A., Lal, A. A., & Ayala, F. J. (1998). Genetic polymorphism and natural selection in the malaria parasite Plasmodium falciparum. *Genetics*, *149*(1), 189–202.

Etyang, A. O., Munge, K., Bunyasi, E. W., Matata, L., Ndila, C., Kapesa, S., … Scott, J. A. G. (2014). Burden of disease in adults admitted to hospital in a rural region of coastal Kenya: An analysis of data from linked clinical and demographic surveillance systems. *The Lancet Global Health*, *2*(4), 216–224.

Ford, L., Lobo, C. A., Rodriguez, M., Zalis, M. G., Machado, R. L., Rossit, A. R., … Lustigman, S. (2007). Differential antibody responses to Plasmodium falciparum invasion ligand proteins in individuals living in malaria-endemic areas in Brazil and Cameroon. *The American Journal of Tropical Medicine and Hygiene*, *77*(5), 977–983.

Francia, M. E., & Striepen, B. (2014). Cell division in apicomplexan parasites. *Nature Reviews Microbiology*, *12*(2), 125–136.

Galinski, M. R., Medina, C. C., Ingravallo, P., & Barnwell, J. W. (1992). A reticulocyte-binding protein complex of Plasmodium vivax merozoites. *Cell*, *69*(7), 1213–1226.

Gilberger, T.-W. (2003). The cytoplasmic domain of the Plasmodium falciparum ligand EBA-175 is essential for invasion but not protein trafficking. *The Journal of Cell Biology*, *162*(2), 317–327.

Gimnig, J. E., Vulule, J. M., Lo, T. Q., Kamau, L., Kolczak, M. S., Phillips-Howard, P. a., … Hawley, W. a. (2003). Impact of permethrin-treated bed nets on entomologic indices in an area of intense year-round malaria transmission. *American Journal of Tropical Medicine and Hygiene*, *68*(4 SUPPL.), 16–22.

Guevara Patiño, J. a, Holder, a a, McBride, J. S., & Blackman, M. J. (1997). Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. *The Journal of Experimental Medicine*, *186*(10), 1689–1699.

Haldar, K., & Mohandas, N. (2007). Erythrocyte remodeling by malaria parasites. *Current Opinion in Hematology*, *14*(3), 203–209.

Iyer, J., Grüner, A. C., Rénia, L., Snounou, G., & Preiser, P. R. (2007). Invasion of host cells by malaria parasites: a tale of two protein families. *Molecular Microbiology*, *65*(2), 231–249.

John, C. C., O'Donnell, R. a, Sumba, P. O., Moormann, A. M., de Koning-Ward, T. F., King, C. L., … Crabb, B. S. (2004). Evidence that invasion-inhibitory antibodies specific for the 19-kDa fragment of merozoite surface protein-1 (MSP-1 19) can play a protective role against blood-stage Plasmodium falciparum infection in individuals in a malaria endemic area of Africa. *Journal of Immunology (Baltimore, Md. : 1950)*, *173*(1), 666–672.

Kang, J.-M., Ju, H.-L., Kang, Y.-M., Lee, D.-H., Moon, S.-U., Sohn, W.-M., … Na, B.-K. (2012). Genetic polymorphism and natural selection in the C-terminal 42 kDa region of merozoite surface protein-1 among Plasmodium vivax Korean isolates. *Malaria Journal*, *11*(1), 206.

Kariuki, S. K., Njunge, J., Muia, A., Muluvi, G., Gatei, W., Ter Kuile, F., … Shi, Y. P. (2013). Effect of malaria transmission reduction by insecticide-treated bed nets (ITNs) on the genetic diversity of Plasmodium falciparum merozoite surface protein (MSP-1) and circumsporozoite (CSP) in western Kenya. *Malaria Journal*, *12*(1), 295.

Kauth, C. W., Epp, C., Bujard, H., & Lutz, R. (2003). The Merozoite Surface Protein 1 Complex of Human Malaria Parasite Plasmodium falciparum: interactions and arrangements of subunits. *Journal of Biological Chemistry*, *278*(25), 22257–22264.

Keeley, A., & Soldati, D. (2004). The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. *Trends in Cell Biology*, *14*(10), 528–532.

Kim, T.-S., Sohn, Y., Kim, J.-Y., Lee, W.-J., Na, B., Kang, Y.-J., & Lee, H.-W. (2014). Detection of antibodies against the CB9 to ICB10 region of merozoite surface protein-1 of Plasmodium vivax among the inhabitants in epidemic areas. *Malaria Journal*, *13*(1), 311.

Langhorne, J., Ndungu, F. M., Sponaas, A.-M., & Marsh, K. (2008). Immunity to malaria: more questions than answers. *Nature Immunology*, *9*(7), 725–732.

Lin, C. S., Uboldi, A. D., Epp, C., Bujard, H., Tsuboi, T., Czabotar, P. E., & Cowman, A. F. (2016). Multiple P. falciparum Merozoite Surface Protein 1 Complexes Mediate Merozoite Binding to Human Erythrocytes. *Journal of Biological Chemistry*, jbc.M115.698282.

Liu, W., Li, Y., Learn, G. H., Rudicell, R. S., Robertson, J. D., Keele, B. F., … Hahn, B. H. (2010). Origin of the human malaria parasite Plasmodium falciparum in gorillas. *Nature*, *467*(7314), 420–425.

Lopaticki, S., Maier, a. G., Thompson, J., Wilson, D. W., Tham, W.-H., Triglia, T., … Cowman, a. F. (2011). Reticulocyte and Erythrocyte Binding-Like Proteins Function Cooperatively in Invasion of Human Erythrocytes by Malaria Parasites. *Infection and Immunity*, *79*(3), 1107–1117.

Low, A., Chandrashekaran, I. R., Adda, C. G., Yao, S., Sabo, J. K., Zhang, X., … Norton, R. S. (2007). Merozoite surface protein 2 of *Plasmodium falciparum*: Expression, structure, dynamics, and fibril formation of the conserved N-terminal domain. *Biopolymers*, *87*(1), 12–22.

Maier, A. G., Duraisingh, M. T., Reeder, J. C., Patel, S. S., Kazura, J. W., Zimmerman, P. A., & Cowman, A. F. (2003). Plasmodium falciparum erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nature Medicine*, *9*(1), 87–92.

Marsh, K., & Kinyanjui, S. (2006). Immune effector mechanisms in malaria. *Parasite Immunology*, *28*(1-2), 51–60.

Mayer, D. C. G., Cofie, J., Jiang, L., Hartl, D. L., Tracy, E., Kabat, J., … Miller, L. H. (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.

Mazumdar, S., Mukherjee, P., Yazdani, S. S., Jain, S. K., Mohmmed, A., & Chauhan, V. S. (2010). Plasmodium falciparum merozoite surface protein 1 (MSP-1)-MSP-3 chimeric protein: Immunogenicity determined with human-compatible adjuvants and induction of protective immune response. *Infection and Immunity*, *78*(2), 872–883.

Mbogo, C. M., Mwangangi, J. M., Nzovu, J., Gu, W., Yan, G., Gunter, J. T., … Beier, J. C. (2003). Spatial and temporal heterogeneity of anopheles mosquitoes and *Plasmodium falciparum* transmission along the kenyan coast, *68*(6), 734–742.

Miller, L. H., Baruch, D. I., Marsh, K., & Doumbo, O. K. (2002). The pathogenic basis of malaria. *Nature*, *415*(6872), 673–679.

Moncunill, G., Mayor, A., Jiménez, A., Nhabomba, A., Puyol, L., Manaca, M. N., … Dobaño, C. (2013). Cytokine and Antibody Responses to Plasmodium falciparum in Naïve Individuals during a First Malaria Episode: Effect of Age and Malaria Exposure. *PLoS ONE*, *8*(2), e55756.

Murray, C. J., Rosenfeld, L. C., Lim, S. S., Andrews, K. G., Foreman, K. J., Haring, D., … Lopez, A. D. (2012). Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet*, *379*(9814), 413–431.

Mwangi, T. W., Ross, A., Snow, R. W., & Marsh, K. (2005). Case Definitions of Clinical Malaria under Different Transmission Conditions in Kilifi District , Kenya, *191*.

Mwingira, F., Nkwengulila, G., Schoepflin, S., Sumari, D., Beck, H.-P., Snounou, G., … others. (2011). Plasmodium falciparum msp1, msp2 and glurp allele frequency and diversity in sub-Saharan Africa. *Malar J*, *10*(1), 79–88.

Nahlen, B. L., Korenromp, E. L., Miller, J. M., & Shibuya, K. (2005). Malaria risk: Estimating clinical episodes of malaria. *Nature*, *437*(7056), E3–E3.

Ngoundou-Landji, J., Nwuba, R. I., Anumudu, C. I., Odaibo, A. B., Matondo Maya, W. D., Awobode, H. O., … Ntoumi, F. (2010). Fine specificity of anti-MSP119 antibodies and multiplicity of Plasmodium falciparum merozoite surface protein 1 types in individuals in Nigeria with sub-microscopic infection. *Malaria Journal*, *9*, 287.

Ogutu, B. R., Apollo, O. J., McKinney, D., Okoth, W., Siangla, J., Filip Dubovsky, … for the MSP-1 Malaria Vaccine Working Group. (2009). Blood Stage Malaria Vaccine Eliciting

High Antigen-Specific Antibody Concentrations Confers No Protection to Young Children in Western Kenya. *PLoS ONE*, *4*(3), e4708.

Okiro, E. a, Hay, S. I., Gikandi, P. W., Sharif, S. K., Noor, A. M., Peshu, N., … Snow, R. W. (2007). The decline in paediatric malaria admissions on the coast of Kenya. *Malaria Journal*, *6*, 151.

Osier, F. H. A., Weedall, G. D., Verra, F., Murungi, L., Tetteh, K. K. A., Bull, P., … Conway, D. J. (2010). Allelic Diversity and Naturally Acquired Allele-Specific Antibody Responses to Plasmodium falciparum Apical Membrane Antigen 1 in Kenya. *Infection and Immunity*, *78*(11), 4625–4633.

Pizarro, J. C., Chitarra, V., Verger, D., Holm, I., Pe, S., Dartevelle, S., … Bentley, G. A. (2003). Crystal Structure of a Fab Complex Formed with PfMSP1-19 , the C-terminal Fragment of Merozoite Surface Protein 1 from Plasmodium falciparum : A Malaria Vaccine Candidate, *2836*(03), 1091–1103.

Polley, S. D., Tetteh, K. K., Lloyd, J. M., Akpogheneta, O. J., Greenwood, B. M., Bojang, K. A., & Conway, D. J. (2007). Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. *Journal of Infectious Diseases*, *195*(2), 279–287.

Preiser, P. R., Khan, S., Costa, F. T. M., Jarra, W., Belnoue, E., Ogun, S., … Rénia, L. (2002). Stage-specific transcription of distinct repertoires of a multigene family during Plasmodium life cycle. *Science (New York, N.Y.)*, *295*(5553), 342–345.

Qari, S. H., Shi, Y. P., Goldman, I. F., Nahlen, B. L., Tibayrenc, M., & Lal, A. a. (1998). Predicted and observed alleles of Plasmodium falciparum merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen. *Molecular and Biochemical Parasitology*, *92*(2), 241–252.

Roy, S. W., Ferreira, M. U., & Hartl, D. L. (2008). Evolution of allelic dimorphism in malarial surface antigens. *Heredity*, *100*(2), 103–110.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*(18), 2496–2497.

Sakihama, N., Ohmae, H., Bakote'e, B., Kawabata, M., Hirayama, K., & Tanabe, K. (2006). Limited allelic diversity of Plasmodium falciparum merozoite surface protein 1 gene from populations in the Solomon Islands. *Am J Trop Med Hyg*, *74*(1), 31–40.

Sanders, P. R. (2005). Distinct Protein Classes Including Novel Merozoite Surface Antigens in Raft-like Membranes of Plasmodium falciparum. *Journal of Biological Chemistry*, *280*(48), 40169–40176.

Schatz, M. (2010). Assembly of large genomes using cloud computing. *Illumina Sequencing Panel, Toronto, ON Canada*, 1165–1173.

Simpalipan, P., Pattaradilokrat, S., Siripoon, N., Seugorn, A., Kaewthamasorn, M., Butcher, R. D., & Harnyuttanakorn, P. (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), 54.

Smith, R. C., Colón-López, D. D., & Bosch, J. (2014). Immunization against a merozoite sheddase promotes multiple invasion of red blood cells and attenuates Plasmodium infection in mice. *Malaria Journal*, *13*(1), 313.

Takala, S. L., Smith, D. L., Thera, M. a., Coulibaly, D., Doumbo, O. K., & Plowe, C. V. (2007). Short report: Rare Plasmodium falciparum merozoite surface protein 1 19-KDA (MSP-119) haplotypes identified in Mali using high-throughput genotyping methods. *American Journal of Tropical Medicine and Hygiene*, *76*(5), 855–859.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739.

Tanabe, K., Sakihama, N., Rooth, I., Björkman, A., & Färnert, A. (2007). High frequency of recombination-driven allelic diversity and temporal variation of Plasmodium falciparum msp1 in Tanzania. *American Journal of Tropical Medicine and Hygiene*, *76*(6), 1037–1045.

Tham, W.-H., Wilson, D. W., Lopaticki, S., Schmidt, C. Q., Tetteh-Quarcoo, P. B., Barlow, P. N., … Cowman, A. F. (2010). Complement receptor 1 is the host erythrocyte receptor for Plasmodium falciparum PfRh4 invasion ligand. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(40), 17327–32.

Tonkin, M. L., Arredondo, S. a, Loveless, B. C., Serpa, J. J., Makepeace, K. a T., Sundar, N., … Boulanger, M. J. (2013). Structural and biochemical characterization of Plasmodium falciparum 12 (Pf12) reveals a unique interdomain organization and the potential for an antiparallel arrangement with Pf41. *The Journal of Biological Chemistry*, *288*(18), 12805–17.

Triglia, T., Healer, J., Caruana, S. R., Hodder, A. N., Anders, R. F., Crabb, B. S., & Cowman, A. F. (2000). Apical membrane antigen 1 plays a central role in erythrocyte invasion by Plasmodium species. *Molecular Microbiology*, *38*(4), 706–718.

Tzeng, J. Y., Bourgain, C., Génin, E., & Quesneville, H. (2005). Evolutionary-based grouping of haplotypes in association analysis. *Genetic Epidemiology*, *28*(3), 220–231.

Wilson, D. W., Fowkes, F. J. I., Gilson, P. R., Elliott, S. R., Tavul, L., Michon, P., … Beeson, J. G. (2011). Quantifying the Importance of MSP1-19 as a Target of Growth-Inhibitory and Protective Antibodies against Plasmodium falciparum in Humans. *PLoS ONE*, *6*(11). 05

World Health Organization, Tuschman, M., World Health Organization, & Global Malaria Programme. (2013). *World malaria report 2013*.

Wright, K. E., Hjerrild, K. a, Bartlett, J., Douglas, A. D., Jin, J., Brown, R. E., … Higgins, M. K. (2014). Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature*, *7*(7527), 427–430.

CHAPTER 6

APPENDICES

Appendix 1. Phyclust Script in R programming for haplotype clustering

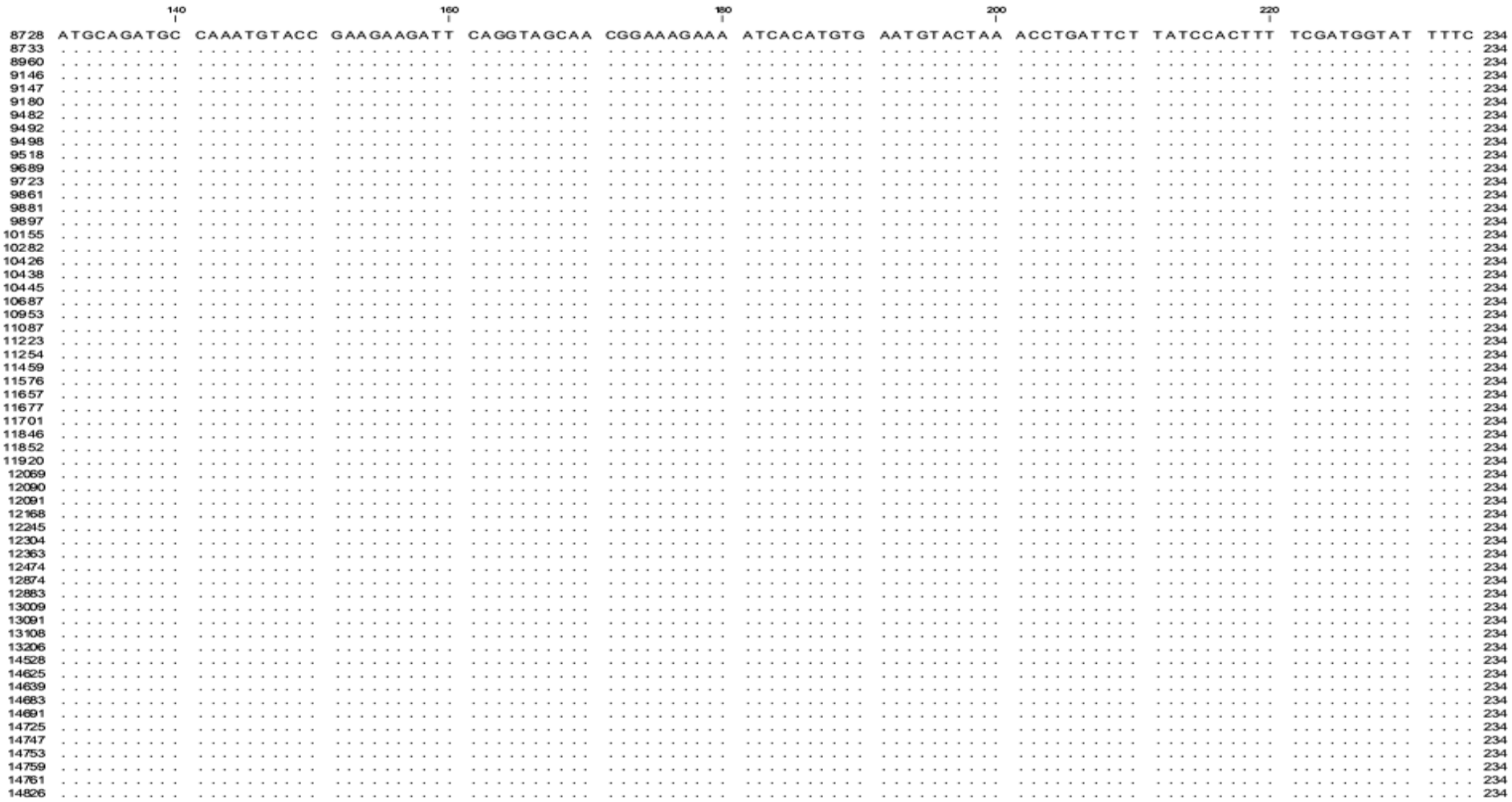
```
R version 3.2.1 (2015-06-18) -- "World-Famous Astronaut"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

        > library(ape)
        > library(phyclust)
        > my.snp <- read.phylip("C:/Users/Reuben/Desktop/msp.phy", code.type = "NUCLEOTIDE")
        > ret <- haplo.post.prob(my.snp$org, ploidy = 1)
        > getcut.fun(sort(ret$haplo$hap.prob, decreasing = TRUE),
        +            nn = my.snp$nseq, plot = 1)
        [1] 6
        > ret$haplo$hap.prob
                [1] 0.360000000 0.260000000 0.040000000 0.046666667 0.180000000 0.026666667
                [7] 0.040000000 0.006666667 0.020000000 0.013333333 0.006666667
        > write.fasta(ret$haplo$haplotype,"C:/Users/Reuben/Desktop/project/haplotype.fa")

# The read.phylip function fetches the sequences from aligned sequence in PHYLIP format guided by the data path
and stores them in my.snp variable in a list of objects.
# The haplo.post.prob function runs the clustering process and stores list of objects in ret.
# The getcut.fun truncates the evolutionary significant haplotypes and then plots a graph.
# ret$haplo$hap.prob, returned the frequencies of the haplotypes
# The write.fasta function, writes the haplotypes and edits the file name in the directory specified for futher
analysis.
```

Appendix 2. Command line sequence clustering confirmation using USEARCH

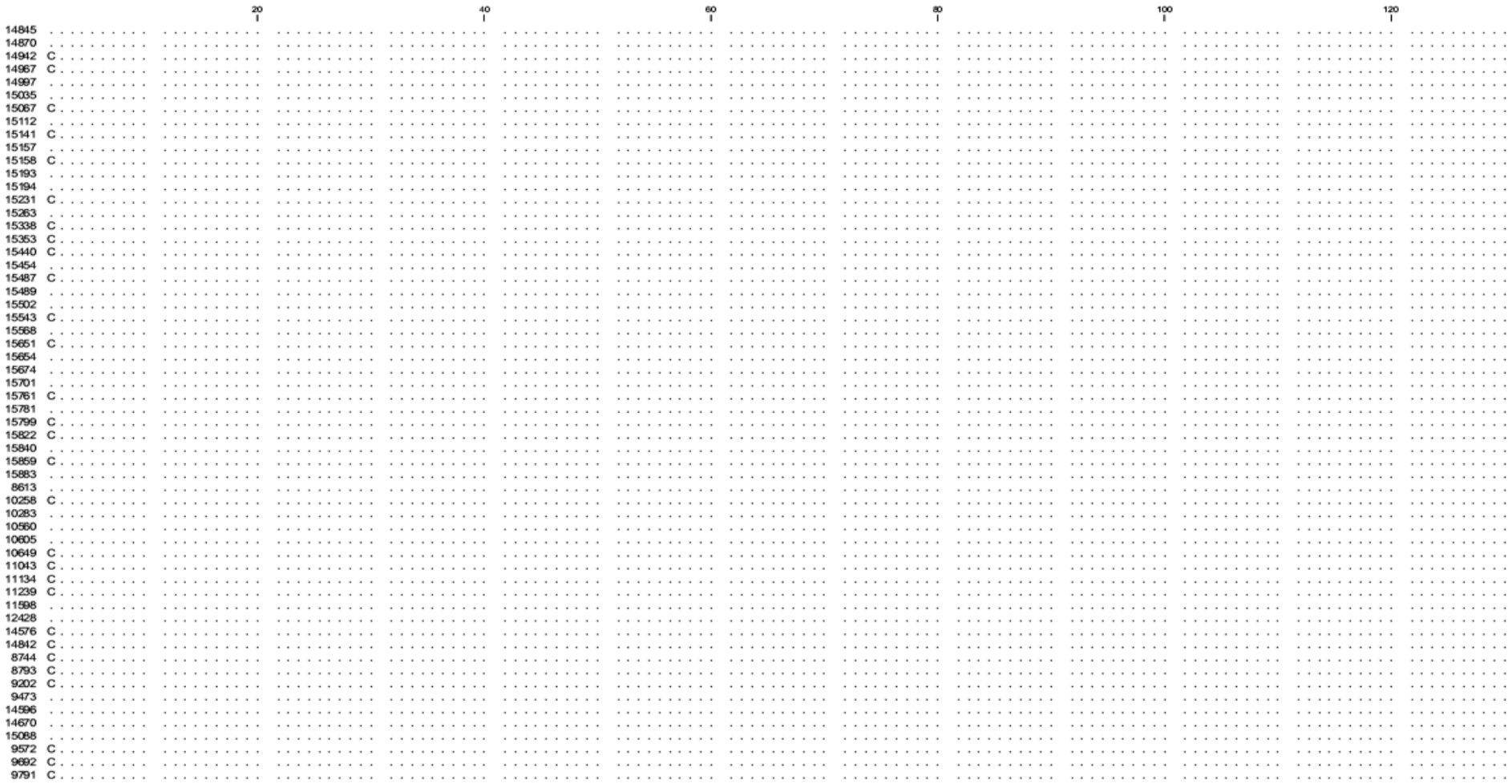
```
                        Clustering command line script
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Reuben>cd

C:\Users\Reuben>cd Desktop

C:\Users\Reuben\Desktop>cd project

C:\Users\Reuben\Desktop\Project>cd Usearch

C:\Users\Reuben\Desktop\Project\Usearch>usearch.exe -cluster_fast msp.fa
-id 1.0
 -centroids nr.fasta -uc clusters.uc -minseqlength 5 -query_cov 1
-target_cov 1
usearch v7.0.1090_win32, 2.0Gb RAM (2.1Gb total), 2 cores
(C) Copyright 2013 Robert C. Edgar, all rights reserved.
http://drive5.com/usearch

Licensed to: mangiruben@gmail.com

00:00 2.1Mb Reading msp.fa, 36.9kb
00:00 2.2Mb 150 seqs, min 234, avg 234, max 234nt
00:00 3.8Mb  100.0% 11 clusters, max size 54, avg 13.6

# This is are results, but detailes are exported as clusters.uc file and
the haplotype sequences in nr.fasta

      Seqs  150
  Clusters  11
  Max size  54
  Avg size  13.6
  Min size  1
Singletons  2, 1.3% of seqs, 18.2% of clusters
   Max mem  3.8Mb
      Time  1.00s
Throughput  150.0 seqs/sec.

# End
```

Appendix 3: Nucleotide Alignment

Appendix 3 cont.

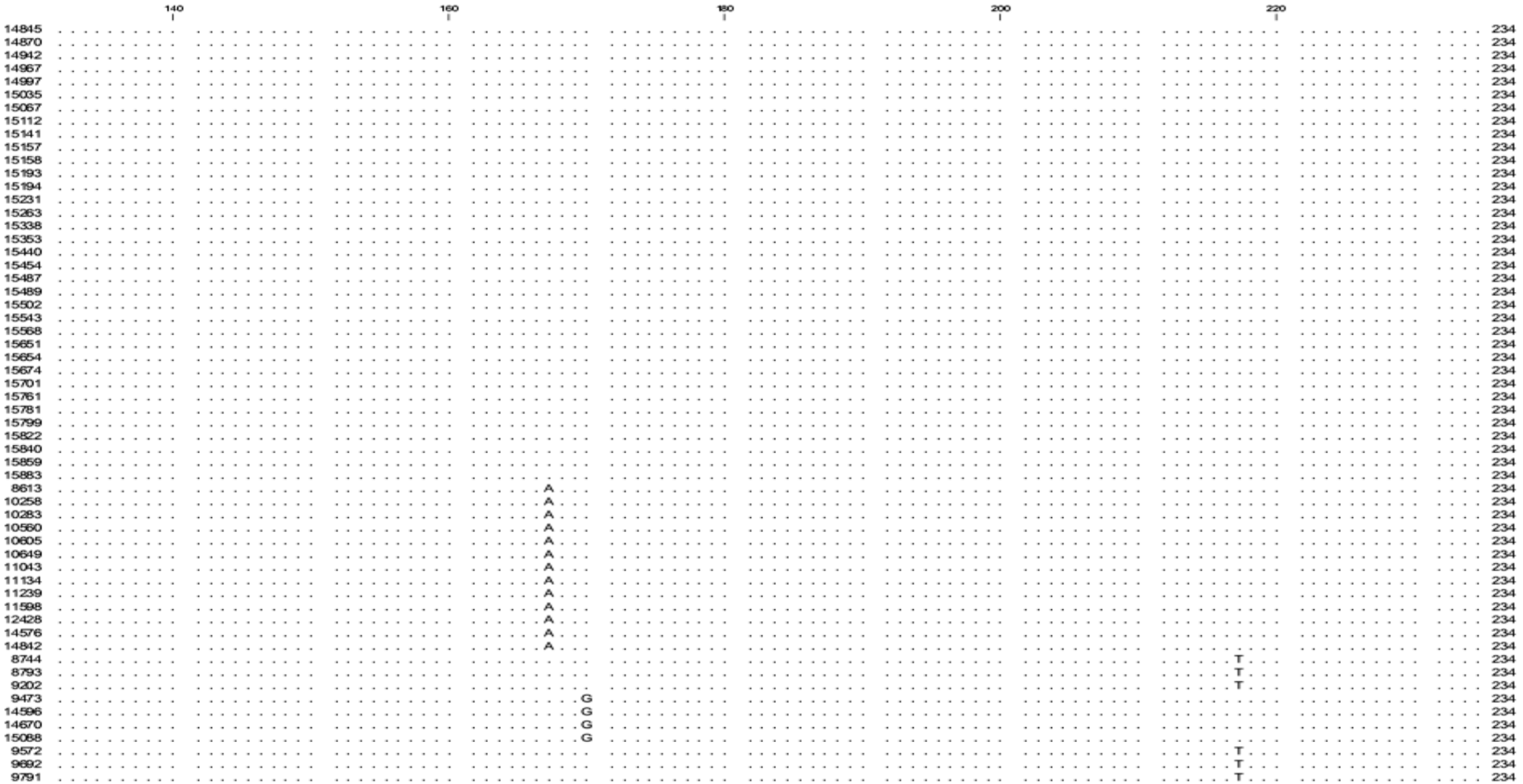


Appendix 3 cont.



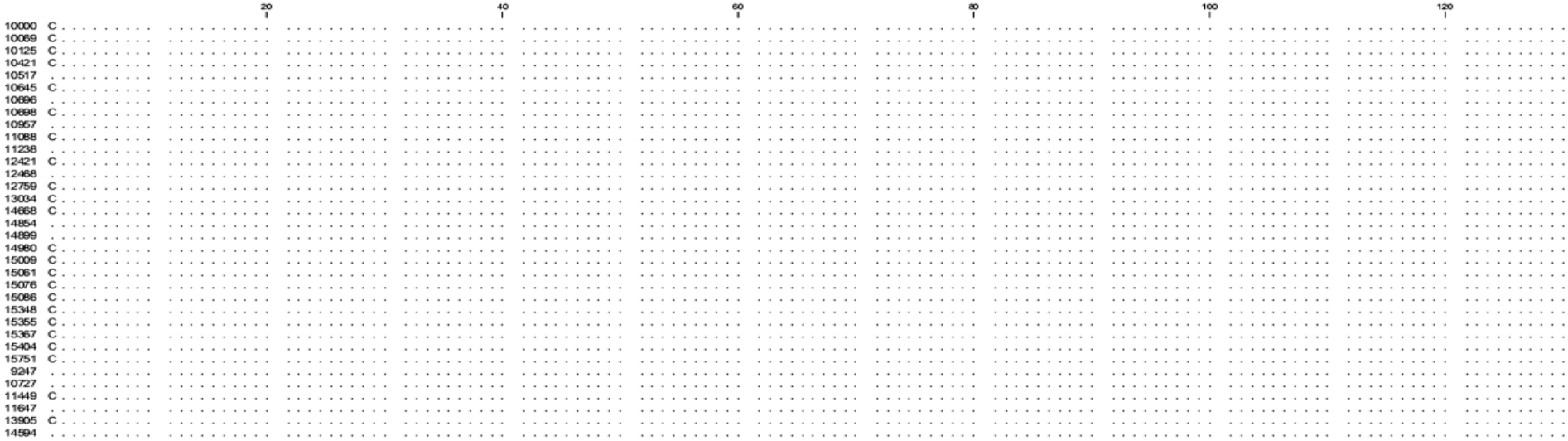
3

Appendix 3 Cont.



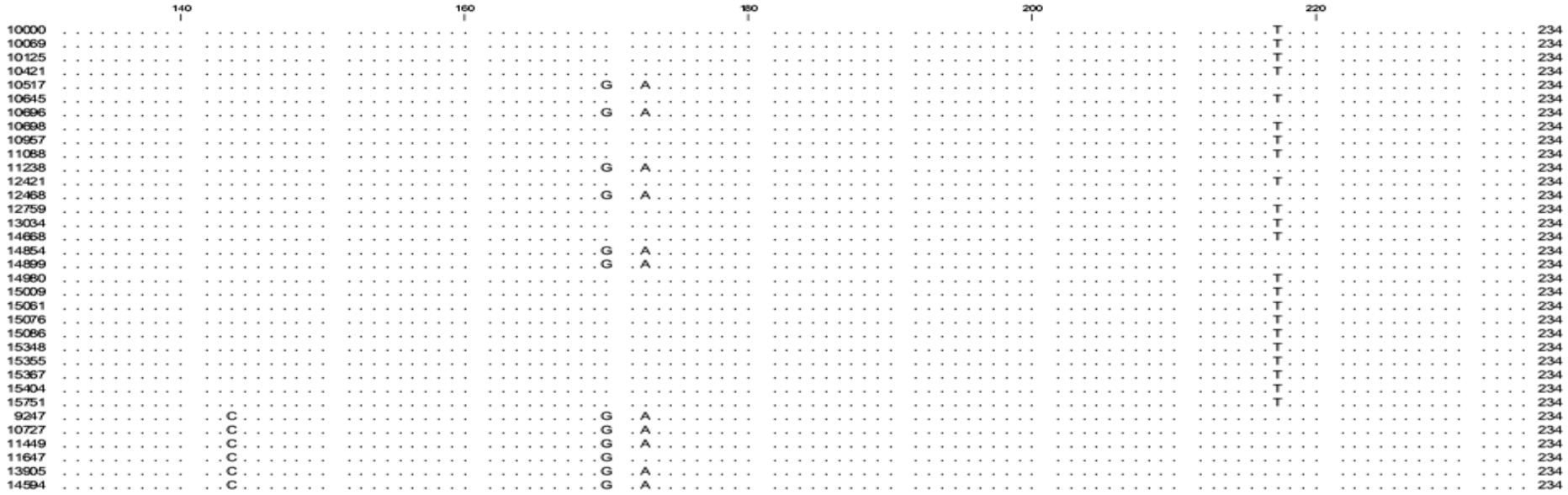
4

Appendix3 Cont.



5

Appendix 3 Cont



6

Appendix 4. Clustering information in .UC file

C:\Users\Reuben\Desktop\Project\Usearch\clusters.uc Jumamosi, Agosti 29, 2015 5:56 PM

```
S   0   234 *   *   *   *   *   8728    *
H   0   234 100.0   .   0   0   =   8733    8728
H   0   234 100.0   .   0   0   =   9147    8728
H   0   234 100.0   .   0   0   =   9482    8728
H   0   234 100.0   .   0   0   =   9492    8728
H   0   234 100.0   .   0   0   =   9498    8728
H   0   234 100.0   .   0   0   =   9518    8728
H   0   234 100.0   .   0   0   =   9689    8728
H   0   234 100.0   .   0   0   =   9723    8728
H   0   234 100.0   .   0   0   =   9861    8728
H   0   234 100.0   .   0   0   =   9881    8728
H   0   234 100.0   .   0   0   =   9897    8728
H   0   234 100.0   .   0   0   =   10155   8728
H   0   234 100.0   .   0   0   =   10426   8728
H   0   234 100.0   .   0   0   =   10438   8728
H   0   234 100.0   .   0   0   =   10687   8728
H   0   234 100.0   .   0   0   =   11223   8728
H   0   234 100.0   .   0   0   =   11254   8728
H   0   234 100.0   .   0   0   =   11657   8728
H   0   234 100.0   .   0   0   =   11677   8728
H   0   234 100.0   .   0   0   =   11701   8728
H   0   234 100.0   .   0   0   =   11920   8728
H   0   234 100.0   .   0   0   =   12069   8728
H   0   234 100.0   .   0   0   =   12090   8728
H   0   234 100.0   .   0   0   =   12091   8728
H   0   234 100.0   .   0   0   =   12168   8728
H   0   234 100.0   .   0   0   =   12304   8728
H   0   234 100.0   .   0   0   =   12474   8728
H   0   234 100.0   .   0   0   =   13091   8728
H   0   234 100.0   .   0   0   =   13206   8728
H   0   234 100.0   .   0   0   =   14625   8728
H   0   234 100.0   .   0   0   =   14725   8728
H   0   234 100.0   .   0   0   =   14759   8728
H   0   234 100.0   .   0   0   =   14761   8728
H   0   234 100.0   .   0   0   =   14826   8728
H   0   234 100.0   .   0   0   =   14845   8728
H   0   234 100.0   .   0   0   =   14870   8728
H   0   234 100.0   .   0   0   =   14997   8728
H   0   234 100.0   .   0   0   =   15035   8728
H   0   234 100.0   .   0   0   =   15112   8728
H   0   234 100.0   .   0   0   =   15157   8728
H   0   234 100.0   .   0   0   =   15193   8728
H   0   234 100.0   .   0   0   =   15194   8728
H   0   234 100.0   .   0   0   =   15263   8728
H   0   234 100.0   .   0   0   =   15454   8728
H   0   234 100.0   .   0   0   =   15489   8728
H   0   234 100.0   .   0   0   =   15502   8728
H   0   234 100.0   .   0   0   =   15568   8728
H   0   234 100.0   .   0   0   =   15654   8728
H   0   234 100.0   .   0   0   =   15674   8728
H   0   234 100.0   .   0   0   =   15701   8728
H   0   234 100.0   .   0   0   =   15781   8728
H   0   234 100.0   .   0   0   =   15840   8728
H   0   234 100.0   .   0   0   =   15883   8728
S   1   234 *   *   *   *   *   8960    *
H   1   234 100.0   .   0   0   =   9146    8960
H   1   234 100.0   .   0   0   =   9180    8960
```

Appendix 4 Cont.

C:\Users\Reuben\Desktop\Project\Usearch\clusters.uc Jumamosi, Agosti 29, 2015 5:56 PI

```
H   1   234 100.0   .   0   0   =   10282   8960
H   1   234 100.0   .   0   0   =   10445   8960
H   1   234 100.0   .   0   0   =   10953   8960
H   1   234 100.0   .   0   0   =   11087   8960
H   1   234 100.0   .   0   0   =   11459   8960
H   1   234 100.0   .   0   0   =   11576   8960
H   1   234 100.0   .   0   0   =   11846   8960
H   1   234 100.0   .   0   0   =   11852   8960
H   1   234 100.0   .   0   0   =   12245   8960
H   1   234 100.0   .   0   0   =   12363   8960
H   1   234 100.0   .   0   0   =   12874   8960
H   1   234 100.0   .   0   0   =   12883   8960
H   1   234 100.0   .   0   0   =   13009   8960
H   1   234 100.0   .   0   0   =   13108   8960
H   1   234 100.0   .   0   0   =   14528   8960
H   1   234 100.0   .   0   0   =   14639   8960
H   1   234 100.0   .   0   0   =   14683   8960
H   1   234 100.0   .   0   0   =   14691   8960
H   1   234 100.0   .   0   0   =   14747   8960
H   1   234 100.0   .   0   0   =   14753   8960
H   1   234 100.0   .   0   0   =   14942   8960
H   1   234 100.0   .   0   0   =   14967   8960
H   1   234 100.0   .   0   0   =   15067   8960
H   1   234 100.0   .   0   0   =   15141   8960
H   1   234 100.0   .   0   0   =   15158   8960
H   1   234 100.0   .   0   0   =   15231   8960
H   1   234 100.0   .   0   0   =   15338   8960
H   1   234 100.0   .   0   0   =   15353   8960
H   1   234 100.0   .   0   0   =   15440   8960
H   1   234 100.0   .   0   0   =   15487   8960
H   1   234 100.0   .   0   0   =   15543   8960
H   1   234 100.0   .   0   0   =   15651   8960
H   1   234 100.0   .   0   0   =   15761   8960
H   1   234 100.0   .   0   0   =   15799   8960
H   1   234 100.0   .   0   0   =   15822   8960
H   1   234 100.0   .   0   0   =   15859   8960
S   2   234 *   *   *   *   *   8744    *
H   2   234 100.0   .   0   0   =   8793    8744
H   2   234 100.0   .   0   0   =   9202    8744
H   2   234 100.0   .   0   0   =   9572    8744
H   2   234 100.0   .   0   0   =   9692    8744
H   2   234 100.0   .   0   0   =   9791    8744
H   2   234 100.0   .   0   0   =   10000   8744
H   2   234 100.0   .   0   0   =   10069   8744
H   2   234 100.0   .   0   0   =   10125   8744
H   2   234 100.0   .   0   0   =   10421   8744
H   2   234 100.0   .   0   0   =   10645   8744
H   2   234 100.0   .   0   0   =   10698   8744
H   2   234 100.0   .   0   0   =   11088   8744
H   2   234 100.0   .   0   0   =   12421   8744
H   2   234 100.0   .   0   0   =   12759   8744
H   2   234 100.0   .   0   0   =   13034   8744
H   2   234 100.0   .   0   0   =   14668   8744
H   2   234 100.0   .   0   0   =   14980   8744
H   2   234 100.0   .   0   0   =   15009   8744
H   2   234 100.0   .   0   0   =   15061   8744
H   2   234 100.0   .   0   0   =   15076   8744
```

Appendix 4. Cont.

C:\Users\Reuben\Desktop\Project\Usearch\clusters.uc Jumamosi, Agosti 29, 2015 5:56 PM

```
H   2   234 100.0   .   0   0   =   15086   8744
H   2   234 100.0   .   0   0   =   15348   8744
H   2   234 100.0   .   0   0   =   15355   8744
H   2   234 100.0   .   0   0   =   15367   8744
H   2   234 100.0   .   0   0   =   15404   8744
H   2   234 100.0   .   0   0   =   15751   8744
S   3   234 *   *   *   *   *   10258   *
H   3   234 100.0   .   0   0   =   10649   10258
H   3   234 100.0   .   0   0   =   11043   10258
H   3   234 100.0   .   0   0   =   11134   10258
H   3   234 100.0   .   0   0   =   11239   10258
H   3   234 100.0   .   0   0   =   14576   10258
H   3   234 100.0   .   0   0   =   14842   10258
S   4   234 *   *   *   *   *   10517   *
H   4   234 100.0   .   0   0   =   10696   10517
H   4   234 100.0   .   0   0   =   11238   10517
H   4   234 100.0   .   0   0   =   12468   10517
H   4   234 100.0   .   0   0   =   14854   10517
H   4   234 100.0   .   0   0   =   14899   10517
S   5   234 *   *   *   *   *   8613    *
H   5   234 100.0   .   0   0   =   10283   8613
H   5   234 100.0   .   0   0   =   10560   8613
H   5   234 100.0   .   0   0   =   10605   8613
H   5   234 100.0   .   0   0   =   11598   8613
H   5   234 100.0   .   0   0   =   12428   8613
S   6   234 *   *   *   *   *   9473    *
H   6   234 100.0   .   0   0   =   14596   9473
H   6   234 100.0   .   0   0   =   14670   9473
H   6   234 100.0   .   0   0   =   15088   9473
S   7   234 *   *   *   *   *   9247    *
H   7   234 100.0   .   0   0   =   10727   9247
H   7   234 100.0   .   0   0   =   14594   9247
S   8   234 *   *   *   *   *   11449   *
H   8   234 100.0   .   0   0   =   13905   11449
S   9   234 *   *   *   *   *   11647   *
S   10  234 *   *   *   *   *   10957   *
C   0   54  *   *   *   *   *   8728    *
C   1   39  *   *   *   *   *   8960    *
C   2   27  *   *   *   *   *   8744    *
C   3   7   *   *   *   *   *   10258   *
C   4   6   *   *   *   *   *   10517   *
C   5   6   *   *   *   *   *   8613    *
C   6   4   *   *   *   *   *   9473    *
C   7   3   *   *   *   *   *   9247    *
C   8   2   *   *   *   *   *   11449   *
C   9   1   *   *   *   *   *   11647   *
C   10  1   *   *   *   *   *   10957   *
```

Appendix 5. Multiple Sequence Alignment Grid

Alignment file: C:/Users/Reuben/Desktop/MSP_1_19 Amino_acid Alignment.fa@Multiple alignment

Table content: symbol counts

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Consensus	**-**	**N**	**S**	**G**	**C**	**F**	**R**	**H**	**L**	**D**	**E**	**R**	**E**	**E**	**C**	**K**	**C**	**L**	**L**	**N**	**Y**	**K**	**Q**	**E**	**G**	**D**
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	150	0	0	0	0	0	0	0	0	0	150	0	150	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150
E	75	0	0	0	0	0	0	0	0	0	150	0	150	150	0	0	0	0	0	0	0	0	0	150	0	0
F	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0
H	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	150	0	0	0	0
L	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	150	150	0	0	0	0	0	0	0
N	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0
R	0	0	0	0	0	0	150	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0

	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
Consensus	**K**	**C**	**V**	**E**	**N**	**P**	**N**	**P**	**T**	**C**	**N**	**E**	**N**	**N**	**G**	**G**	**C**	**D**	**A**	**D**	**A**	**K**	**C**	**T**	**E**	**E**
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	150	0	0	0	0	0
C	0	150	0	0	0	0	0	0	0	150	0	0	0	0	0	0	150	0	0	0	0	0	150	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	150	0	0	0	0	0	0
E	0	0	0	150	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	150	150
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	150	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	144	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	150	0	150	0	0	0	150	0	150	150	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	150	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	6	0	150	0	0
V	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
Consensus	**D**	**S**	**G**	**S**	**N**	**G**	**K**	**K**	**I**	**T**	**C**	**E**	**C**	**T**	**K**	**P**	**D**	**S**	**Y**	**P**	**L**	**F**	**D**	**G**	**I**	**F**
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	150	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0
D	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	150	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	150	0	0	0	150
G	0	0	150	0	0	139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0
K	0	0	0	0	0	0	150	150	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	122	0	0	0	0	0
N	0	0	0	13	134	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	150	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	150	0	137	16	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	150	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0

Legend: 10% 25% 50% 70% 90%

This profile grid is generated by UGENE genetic analysis software, and is used for computing substitution frequencies.