UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES

COMPUTATIONAL BIOLOGY DIVISION

MASTERS THESIS

Title:

Analysis of Within-Host Evolution of *Plasmodium falciparum* during Treatment

By

Javan Ochieng Okendo (OKNJAV001)

Supervisor:

Professor Nicola Mulder

(Computational Biology Division, University of Cape Town, South Africa)

Co-Supervisor:

Dr. Ben Andagalu

(Malaria Drug Resistance Lab, US Army Medical Research Directorate - Kenya/KEMRI)

Submitted in fulfilment of the requirements for the degree of MSc(Med) in Bioinformatics by dissertation

Academic Year: 2018

DECLARATION

I, Javan Ochieng Okendo, hereby declare that the work on which this dissertation is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 24/05/2018

Acknowledgements

Special thanks to (DELGEME) under the leadership of Professor Abdoulaye Djimde for funding this research at the University of Cape Town, South Africa. Many thanks go to Unites States Army Medical Research Directorate/Kenya Medical Research Institute, for providing data which were used in the analysis. I would also like to acknowledge my supervisors Professor Nicola Mulder, Head Computational Biology Division, Institute of Infectious Diseases and Molecular Medicine and Dr. Ben Andagalu, Deputy Director, Malaria Drug Resistance Laboratory United States Army Medical Research Directorate/KEMRI for their continued support, encouragement and guidance throughout this project, without you this journey would not have been a success. I am also grateful to my colleagues at the Computation Division for their support and assistance. To my late mother Rose Achieng Okendo even though you never lived to witness this success your spirit lives with us forever. Also, thanks to my Father, Samuel Onyango Okendo, my brothers and all the family members for their encouragement and support. Above many all thanks the almighty God for his abundant grace.

Abstract

Antimalarial drugs impose strong selective pressure on *Plasmodium falciparum* parasite genomes and leave signatures of selection. The evolutionary basis of drug resistant malaria in endemic and epidemic settings continues to remain an ongoing scientific priority whose solution carries a significant effect on treatment outcomes. To understand the evolutionary changes in *P. falciparum* during treatment with ACTs, we used various approaches to test the neutral models of evolution using *P. falciparum* genomic data which were collected from Kombewa and Maseno in Kisumu, Kenya between 2013 and 2015. The Synonymous/Non-synonymous (dN/dS) ratio was used to predict the effect of selection on protein coding loci of the *Pfk13* gene. A logistic regression model was used to test the association between IC_{50s} and the SNPs. mCSM and SDM were used to detect the effects of mutations on the *Pfk13* gene while the PRIMO web server was used to locate the SNPs on the Kelch13 propeller domain. Modeller V9.1 was used to predict the structure of the Kelch 13 propeller domain and the Posview webserver used to predict ACT/kelch 13 interactions. Population differentiation was done using Microsatellite analyzer to calculate F_{ST} and customized R scripts with the relevant population genetics packages were used in the analysis. For samples collected in 2013, Tajima's D genomic summary statistic was 4.53194, Fu & Li D* 2.13380, and Fu &Li F* 3.62142. However, in 2015 Tajima's D was -2.42910, Fu and Li's D* -5.2712, and Fu and Li's F* -5.0045. The dN/dS in 2013 was 1.0299, while in 2015 dN/dS was 2.6884. Kenyan *P. falciparum* SNPs occur on the intra or inter blade domains on the *PfK13* propeller domain. The F_{ST} analysis showed minimal population differentiation of the parasites during treatment. There was no significant association between SNPs and IC_{50} values but SNPs at codon D547E showed association with Artesunate and D559E with AQ and MQ IC_{50} respectively. Even though there is an exponential increase in the number of non-synonymous point mutations in the *Pfk13* gene, the Kenyan *P. falciparum* strains remain sensitive to ACT drugs. Further research needs to be done by deep sequencing this location of chromosome 13 as it will provide more power for finding novel SNPs for further validation.

Key words: *Plasmodium falciparum*, Evolutionary, selective, population, Antimalarial, synonymous, non-synonymous

Table of Contents

Abbreviations

BCE	Before Common Era
BC	Before Christ
CE	Common Era
PCR	Polymerase Chain Reaction
RDP	Rapid Diagnostic Test
WHO	World Health Organization
LAMP	Loop- Mediated Isothermal Amplification
ACT	Artemisinin Combination Therapy
CQ	Chloroquine
SP	Sulphadoxine pyrimethamine
AL	Artemether Lumefantrine
RFLP	Fragment Length polymorphism
DHA	Dihydroartemisinin
crt	Chloroquine resistant transporter
mdr1	Multidrug resistance gene 1
dhfr	Dihydrofolate reductase
dhps	Dihydropteroate synthase
k13	Kelch 13
PfCelTOS	*Plasmodium falciparum* cell traversal protein of ookinetes and sporozoites
PfAMA1	P. falciparum Apical membrane antigen 1
PfMSP1	*Plasmodium falciparum* merozoite surface protein 1
PfRH5	*Plasmodium falciparum* reticulocyte-binding protein homolog 5
PfSERA5	*Plasmodium falciparum* serine repeat antigens 5
PfGLURP	*Plasmodium falciparum* merozoite antigen glutamate-rich protein
PfMSP3	*Plasmodium falciparum* merozoite surface protein 3
Pfs48/45	*Plasmodium falciparum* surface antigen 48/45 kilo Dalton
Pfs25	*Plasmodium falciparum* surface antigen 25kilo Dalton
KEMRI	Kenya Medical Research Institute
WRP	Walter Reed Project
GPS	Global Positioning system
Kg	Kilogram
SDM	Site Directed Mutagenesis
SEA	South East Asia

KE	Kenya
COI	Complexity of Infection
MOI	Multiplicity of Infection
MQ	Mefloquine
dN	Non-synonymous mutation
Ds	Synonymous mutation
CI	Confidence Interval
PI3K	phosphatidylinositol-3-kinase
β- Sheet	Beta sheet
IQR	Interquartile Range
Pfk13	*Plasmodium falciparum* Kelch 13 gene
WRAIR	Walter Reed Army Institute of Research
IRB	Institutional Review Board
UCT	University of Cape Town
HREC	Human Research Ethics Committee
Pop1	Day zero hour zero parasite population
Pop2	Day one hour twenty-four parasite population
Pop3	Day two-hour thirty parasite population
Pop4	Day two hour thirty-six parasite population
Pop5	Day two-hour forty-two parasite population
Pop6	Day two-hour forty-eight parasite population
Pop7	Day twenty-eight parasite population
Pop8	Day thirty-five parasite population
Pop9	Day forty-two parasite population
ATP	Adenosine Triphosphate
SERCA	Sarcoendoplasmic Reticulum Calcium transport ATPase
LAMP	Loop-mediated isothermal amplification
WT_RSA	Wildtype Relative side-chain Solvent Accessibility
MT_RSA	Mutant Relative side-chain Solvent Accessibility
WT_OSP	Wild-type residue-occluded packing density
MT_OSP	mutant residue-occluded packing density
Ddg	delta delta Gibbs free energy score
DELGEME	Developing Excellence in Leadership and Genetic Training for Malaria Elimination in Sub-Saharan Africa

CDC	Centre for Disease Control
SNPs	Single Nucleotide Polymorphisms
#	Number

Table of figures

List of tables

Chapter 1. INTRODUCTION

1.1 Malaria History

The term malaria was coined from the Italian word mal'aria meaning "bad air" and it has been around for more than four millennia before being fully discovered and understood. Malaria is postulated to have influenced human populations and history (Gardner et al. 2011) and caused many deaths in the ancient world (Sallares, Bouwman, and Anderung 2012). Malaria manifestation in ancient times was attributed to the supernatural influences among the people (Neghina et al. 2010). According to Hippocratic medicine, transmission of the infectious disease and the intermittent fevers were caused by "bad air" and "bad" water of marshes which produces Miasmata and was thought to be affecting people residing around those regions (Coluzzi and Corbellini 1995). In 400 BCE, Hippocrates described the features of malaria and discovered its correlation with the season, with the highest incidences in summer (Hempelmann and Krafts 2013). Hippocrates, then drew a correlation between stagnant water and the malaria outbreak in the ancient Roman empire and this prompted the Romans to drain all the stagnant waters to help in the control of malaria (Hempelmann and Krafts 2013).

Historically the spread of malaria was promoted by the movement of migrants, long distance traders and conquering forces (Roper et al. 2004; Tatem, Rogers, and Hay 2006). As late as the sixteenth century to the early nineteenth century, major plantation were being built in Europe and this resulted in the slave trade which saw many Africans being transported to the new world to offer cheap labor (Bianchine and Russo 1992). This resulted in the death of millions of Africans and the spread of malaria to Europe (Bianchine and Russo 1992; Curtin 1968). The infectious diseases that were known to affect the Americans were smallpox, measles, and typhoid. Tropical ecosystems of the world provide a conducive environment for the survival and spread of malaria to other parts of the world and this caused the resurgence of malaria in America due to the transportation of the slaves from Africa to the United States of America (USA) (Bianchine and Russo 1992; Deane 1986).

The understanding of the malaria biology and etiology has evolved over time from theories and myths to a clear scientific understanding due to groundbreaking advances in scientific discoveries and advancement in the field of medicine. In the 19th century, Marcus Terentius Varro, a Roman scholar, and writer discovered that mosquitoes were the vectors for malaria transmission (Hempelmann and Krafts 2013). Varro's finding was later supported by Albert Freeman, an American Physician who later came up with nineteen proofs that showed mosquito to be the vector responsible for malaria transmission (Hempelmann and Krafts 2013). Charles Louis Alphonse, a French officer and

doctor became the first scientist to observe malaria parasite (CDC 2012a; Hempelmann and Krafts 2013). Bartolomeo Camillo Golgi, between 1885-1892 studied the asexual life cycle of malaria and came to the conclusion that there exist two forms of disease; and this was associated with two forms of recurring fevers, tertian and quartan which were caused by different species of *Plasmodium* (CDC 2012a; Hempelmann and Krafts 2013). A groundbreaking discovery was later made in 1898 by Giovanni Battista Grassi, who discovered *Anopheles claviger* as the vector of malaria transmission from infected to uninfected patients (Hempelmann and Krafts 2013). *Plasmodium falciparum* remains the most lethal malarial parasite known to infect human beings and it has been doing so for more than two Millenia, though its origin and evolutionary history remain a controversial debate. Some studies suggest that malaria has its origin from gorillas and not humans, and because of the widespread occurrence of these *Plasmodium* species in Apes, they were spread to human from apes (Liu et al. 2010). Immunochromatographic and immunohistochemical studies on certain body organs of mummies such as the skin and muscles have been used to demonstrate the presence of *Plasmodium falciparum* infection in Egyptian mummy remains dating back to thousands of years ago (Bianucci et al. 2008). The main parasites species in the genus *Plasmodium* that infect human beings are: *P. malariae, P. falciparum, P. vivax, P. ovale*, and *P. knowlesi* with *P. falciparum* being the most virulent malaria parasite and causing the highest malaria cases in the tropical settings of the world (CDC 2012b; Cox 2010). *P. falciparum* it remains a very important parasite in the study of malaria elimination in Africa and in the tropical regions of the world (Hay et al. 2009; Molina-Cruz et al. 2016; Snow 2015).

1.2.0 Malaria Epidemiology

Malaria is a parasitic disease transmitted by the female anopheles mosquito and it remains the most lethal parasitic diseases infecting and affecting humans living in the tropical locations of the world (Oyebola et al. 2017). Those suffering from the disease are mostly children below the age of five years and pregnant women (Molina-Cruz et al. 2016; Talisuna, Bloland, and D'Alessandro 2004). In Africa, 90% of the reported cases were from young children below the age of 5 years (World Health Organization 2015). Malaria transmission is very intense in the tropical region of Africa which has a higher rate of malaria inoculation resulting in high malaria incidences (Figure 1). Approximately 216, million incidences of malaria were reported in the year 2017 and despite the continued efforts, 5 million more deaths were recorded in 2017 compared with 2016 report with about 21.06% of these cases resulting in deaths worldwide (WHO 2017). The highest proportion of malaria infections occurred in resource constrained countries, according to the WHO World Malaria report, 2017.

Malaria infection, being a parasitic disease and its causative agent is known to be *Plasmodium* (Cox 2010; Karuri and Snow 2016).

1.2.1 Evolution of the *Plasmodium falciparum*

The evolutionary time scale shows that human *P. falciparum* is a close relative to the chimpanzee *Plasmodium reichenowi* parasite (Hume, Lyons, and Day 2003). These parasites are thought to have undergone divergence between 5 to7 million years ago, which is the time humans and chimpanzees are believed to have diverged from each other (Molina-Cruz et al. 2016). In phylogenetic analysis, these two parasites always cluster together in space when compared with other human malaria parasite species. Clustering together of these parasites during phylogenetic analysis show that *P. falciparum* has infected *Homo sapiens sapiens* for thousands of years and is as old as human history, however, the origin and spread of malaria parasites is still a debate (Schoepflin *et al*. 2008).

Even though the evolutionary path of *P. falciparum* is still not clear, development of bioinformatics tools enables us to understand its evolutionary dynamics experimentally. Two different theories explain the evolution of ancient *P. falciparum* because of coalescent analysis of *P. falciparum* genomic sequences (Molina-Cruz *et al*. 2016). The first argument asserts that the absence of synonymous mutations in certain genes and absence of SNPs in regularly expressed genes (Hume, *Et al.*2003) shows a recent expansion of *P. falciparum* parasites (Schoepflin *et al*. 2008) and that positive selection can be used to explain diversity. The other camp argues that *P. falciparum* is ancient because of high polymorphism rates observed in certain genes (Hume, *et al.*2003). Thus, *P. falciparum* genetic sequence data provides conflicting explanations in relation to the natural or artificial evolution of the parasite due to drug pressure highlighting the need for further research.

P. falciparum is believed to have originated from Africa and later spread to other continents due to human migration (Molina-Cruz et al. 2016; Tatem, Rogers, and Hay 2006). In the process of movement, *P. falciparum* was confronted by different species of mosquito vectors, which were evolutionary different from the African mosquitos (Hay et al. 2009; Neghina et al. 2010; Tatem, Rogers, and Hay 2006). *P. falciparum* parasites that were fit and had the potential of surviving in the local mosquito vectors continued to infect new hosts and they became fully adapted to new geographic locations (Molina-Cruz et al. 2016; Neghina et al. 2010). Today, approximately 70 anopheline mosquito vector species are known to actively transmit malaria around the globe (Hay *et al*. 2009; Sinka *et al*. 2012; Molina-Cruz *et al*. 2016). Coevolution among organisms has been suggested to have an influence on evolution and adaptation of other organisms in time and space during the evolutionary course (Molina-Cruz *et al*. 2016). The unique life cycles of the *Plasmodium* parasite and its antagonistic association with the human host has been shown by use of molecular markers both in the

host and *Plasmodium* genome (Molina-Cruz *et al*. 2016; Schoepflin *et al*. 2008; Frank Hawking, Wilson, and Gammage 1971).

There is enough evidence showing that the human (host) immune system is the main driver of *P. falciparum* selection as shown by higher diversity in the merozoite surface protein 1 (msp1) and merozoite surface protein 2 (msp2), respectively, which undergo antigenic shift during the life cycle of the *P. falciparum* within the human host, enabling the parasites to evade the host immune system (Figure 2) (Molina-Cruz *et al*. 2016; Frank Hawking, Wilson, and Gammage 1971). The *var* genes in *P. falciparum* are highly diverse and this diversity is due to reorganization and duplication of genes generating multiple domains. Sixty unique *var* genes have been shown to be present in the *P. falciparum* genome (Lusingu and Von Seidlein 2008), and they demonstrate an extremely high level of diversity enabling *P. falciparum* to have a specific repertoire of these gene products even in a population of the same parasites (Molina-Cruz *et al*. 2016; Lusingu and Von Seidlein 2008). Every recombination event of this gene results in the formation of a unique *var* molecule due to a high evolution rate. Differential expression of these *var* genes at different time points during infection also enables *P. falciparum* to evade the host immune system, thus increasing its chance of survival and continuation to infect new erythrocytes within the host (F. Hawking, Worms, and Gammage 19). It is evident that malaria acts as the main driver of selection for certain genetic traits in human populations such as alpha-thalassemia, beta-thalassemia, sickle cell anemia and other haemoglobinopathies in individuals living in high transmission settings (Taylor, Parobek, and Fairhurst 2012), and it is evident that these traits offer protection against severe malaria in these individuals (Molina-Cruz *et al*. 2016). However, it has not been determined whether the *anopheline* immune system imposes selection on *P. falciparum* parasites or whether the *Plasmodium* impose selection in the *anopheline* mosquito. If a parasite vector (*Anopheles gambiae*) from a high transmission region is infected with *P. falciparum* infecting human hosts from the same region (endemic area), the mosquito's immune system does not illicit an optimal antiplasmodial immune response (Taylor, Parobek, and Fairhurst 2012; Molina-Cruz *et al*. 2016). This suggests that transmission of malaria is passive. These findings may be accurate but their interpretation might not be correct. This is because different researched have not come up with a single conclusion and further work need to be done using more advanced bioinformatics tools to further validate their findings.

1.2.2 Single Nucleotide Polymorphisms and adaptive evolution

Most of the molecular markers associated to antimalarial drug resistance are now well understood even though some are yet to be discovered (Picot *et al*. 2009). Non-synonymous single nucleotide polymorphisms are the main molecular markers responsible for the adaptive evolution to antimalarial

drug resistance in *P. falciparum.* For instance, SNPs at codon position: 86, 184, 1034, 1042, and 1246 in the *Pfmdr-1* gene have been shown to modulate the sensitivity of *P. falciparum* to chloroquine and Mefloquine (Foote *et al*. 1990). Discoveries from past studies have shown that Amodiaquine monotherapy selects *P. falciparum* multidrug resistance 1 (*Pfmdr1*) gene SNP N86**Y** (amino acid that is selected is in bold) (Sisowath *et al*. 2005). Resistance to mefloquine, artemisinin and Lumefantrine, have been associated with an increase in the copy number of the *Pfmdr1* gene from experimental studies (Sidhu *et al*. 2006). Point mutations at codons 72 to 76 of the *Pfcrt* gene are known to cause resistance to chloroquine, other quinoline partner drug, and amodiaquine. However, the most prevalent which was associated with the world wide spread of chloroquine resistance is K76**T** (Djimdé 2001). Sulphadoxine-pyrimethamine resistance is caused by SNPs in the genes that encode dihydrofolate reductase and dihydropteroate synthase enzymes which are the main targets of pyrimethamine and Sulphadoxine drugs (Hayton and Su 2008).

Emergence of drug resistant *P. falciparum* parasites in South East Asia and other parts of the world threatens to compromise malaria eradication and elimination efforts using all artemisinin- based combination therapies (ACTs) (Amaratunga *et al*. 2012). *P. falciparum* resistance to ACT is defined by an increase in parasite clearance rate (CR), a significant clinical indicator of drug resistance phenotypes, which is due to increased parasite clearance half-life (Ariey *et al*. 2014). In malaria endemic regions, artemisinin drugs are currently used in combination with other chemical compounds having unrelated chemical properties (Rodrigues *et al*. 2010). The aim is to reduce adaptive evolution to ACT drugs, and increase the chance of their therapeutic efficacy by making use of synergistic actions of their components (Rodrigues *et al*. 2010). The development of an adaptive evolution to the components of ACT in *P. falciparum* treatment is being studied both in Sub-Saharan Africa and around the globe but the discovery of the universal molecular marker for ACT resistance is yet to be made (Eckstein-Ludwig *et al*. 2003). Adaptive evolution to ACT drugs is still debatable but Eckstein-Ludwig *et al*., 2003, proposed that artemisinin exerts its anti- malarial activity by inhibiting the SERCA calcium ion pump and ATPase enzyme in the *P. falciparum* parasite. However, Paloque *et al* (Paloque *et al*. 2016) also observed that mutation in *Pfk13* gene increases the quiescence capacity of *P. falciparum* which wades away when the ACT drug concentration falls below the optimal level in the

blood. SNPs in ATPase have also been associated with the increased Inhibitory Concentrations 50% (IC_{50}) of artemether in South American *P. falciparum* parasites (Rodrigues *et al*. 2010), but experimental evidence of ATPase involvement in ART metabolism is yet to be presented (Eckstein-Ludwig *et al*. 2003). Noedl H, *et al (*2009) found that Western Cambodian parasites did not have point mutation in the ATPase gene of *P. falciparum.* The parasite was confirmed to have reduced

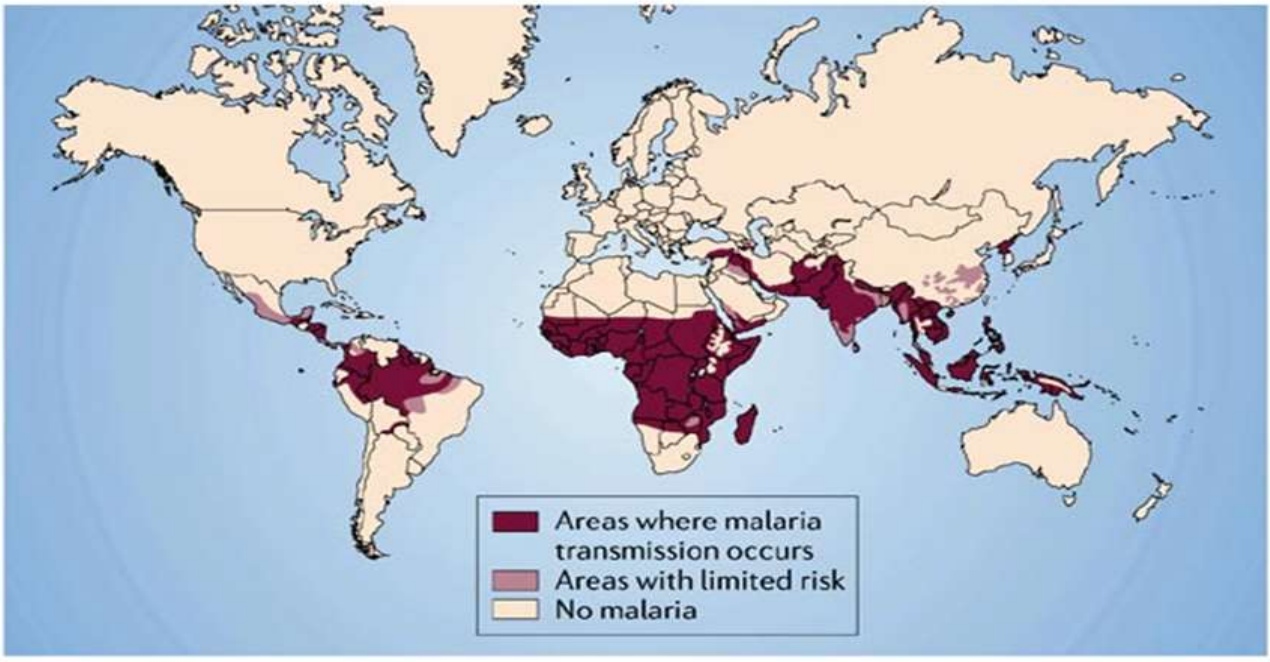
susceptibility to artemisinin drugs, specifically to Artesunate (Noedl, Socheat, and Satimai 2009; Rodrigues *et al*. 2010).

To better understand the role of SNPs in the *PF3D7_1343700* gene, Ariey *et al* (2014) managed to model the K13-propeller 6-blade domain consisting of Kelch motifs made up of four anti-parallel beta sheets. Four SNPs including C580Y, R539T, Y493H and I543T (Ariey *et al*. 2014) were associated with an increased parasite clearance time (slope half-life). However, Ariey and colleagues (2014) did not measure the effects of SNPs and the evolutionary pressure due to ACT usage on the Kelch 13 protein by calculating the Gibbs free energy score between the wildtype and mutant Kelch 13 protein structure. Using more advanced bioinformatics tools, the calculation of Gibbs free energy score enables an understanding of whether these SNPs cause a destabilization or a stabilization of the Kelch 13 protein and its evolutionary effects on protein stability (Pires and Ascher 2016). Consequently, a number of research and clinical trials have relied mostly on the parasite clearance rate (CR) to detect the potential drug resistant strains of *P. falciparum* (Hastings, Kay, and Hodel 2015). However, this strategy does not reveal the genetic changes in the *P. falciparum* which arise long before the drug resistance becomes clinically evident.

Hastings *et al* (2015), suggested that human immunity provides differences with which humans are differentially responsive to treatment. Therefore, parasite clearance rate (CR) is not an effective means for surveillance of drug resistance as it may lead to the wrong interpretations (Hastings, Kay, and Hodel 2015). The confounding factors which may lead to inaccurate interpretation of the results in clinical trials may be different pharmacokinetic activities in human, use of faulty machines for measurements and many more. Use of improved bioinformatics tools in the analysis of the *P. falciparum* genome has provided a prime opportunity and enabled us to deeply understand the within-host evolutionary dynamics of *P. falciparum* (Auburn and Barry 2016)**.** To assess the effects of drug resistance, examination of the methodologies used in the study of drug resistance is very important. A SNP barcode, which is made up of 20 to 50 SNPs gives more accurate and reliable results compared to classical microsatellite analysis method. Daniel and his associates (Daniels *et al*. 2008) have developed a panel of 24 SNP barcode that is used globally to detect *P. falciparum* infections (Daniels *et al*. 2008) (Auburn and Barry 2016). These barcodes have been used successfully in Africa, for example in Malawi, Zambia and Senegal, in South East Asia, specifically Cambodia to measure complexity of Infections (CIO) and in the measurement of haplotype diversity (Auburn and Barry 2016). These SNP barcodes were first used in Ethiopia to genotype *P. falciparum* parasite strains collected from two different locations to measure temporal changes in *P. falciparum* transmission pattern (Daniels *et al*. 2008). Baniecki and his associates developed a 42 Single Nucleotide Polymorphism barcode for detection and mapping the geographic

origin of *P. vivax* parasite strains (Baniecki *et al*. 2015). These barcodes provide a universal platform which can be used in inferring parasite transmission patterns internationally (Baniecki *et al*. 2015), but they cannot offer an insight into the potential development of adaptive evolution in Plasmodium parasites during treatment using different ACT regimens. Alexander *et al* (2013) investigated drug resistance SNPs in *P. falciparum* using fragment length polymorphism (RFLP) analysis and real time PCR (Heuchert *et al*. 2015), but this method cannot provide full variant information since adaptive evolution is a stochastic event. Mathematical modelling has been used to study the transmission intensity and the development of drug resistance but this doesn't account for the within host evolution of malaria (Schoepflin *et al*. 2008).

In Kenya, malaria parasite transmission pattern is affected by rainfall, altitude, vector species and the intensity of mosquito biting (Kenya National Bureau of Statistics (KNBS); ORC Macro 2010). Eco-epidemiologically, Kenya is divided into four main zones depending on malaria stability, with the Lake Victoria basins and coastal regions of Kenya having stable malaria distributions with approximated an annual inoculation rate of between 30 to100, epidemic highland areas of western Kenya, seasonal areas of North-Eastern and regions with no to very low transmission risks of malaria (Bejon *et al*. 2010). Approximately US$ 100 million is spent annually in Kenya in the fight against the malaria pandemic and this burden is expected to rise if the proper management and control measures are not put in place (Sicuri *et al*. 2013). The use of substandard drugs which do not satisfy the WHO quality standards are on the rise in the Kenyan markets and this poses a danger of resistance development in parasite populations as they will promote the parasite evolution to ACT drugs (Nayyar *et al*. 2012; Watsierah *et al*. 2011). The low bioavailability of these drugs in the human blood during treatment course is the main driver of within host evolution of parasites enabling the parasites to "acquire" new phenotype status thus becoming non-responsive to antimalarial treatments as reported in South-East Asia and Sub-Saharan Africa (reviewed Nayyar *et al*. 2012). This is a wakeup call for the manufactures and the policy makers who should ensure only drugs which satisfy the WHO quality standards are used in the treatment of malaria.



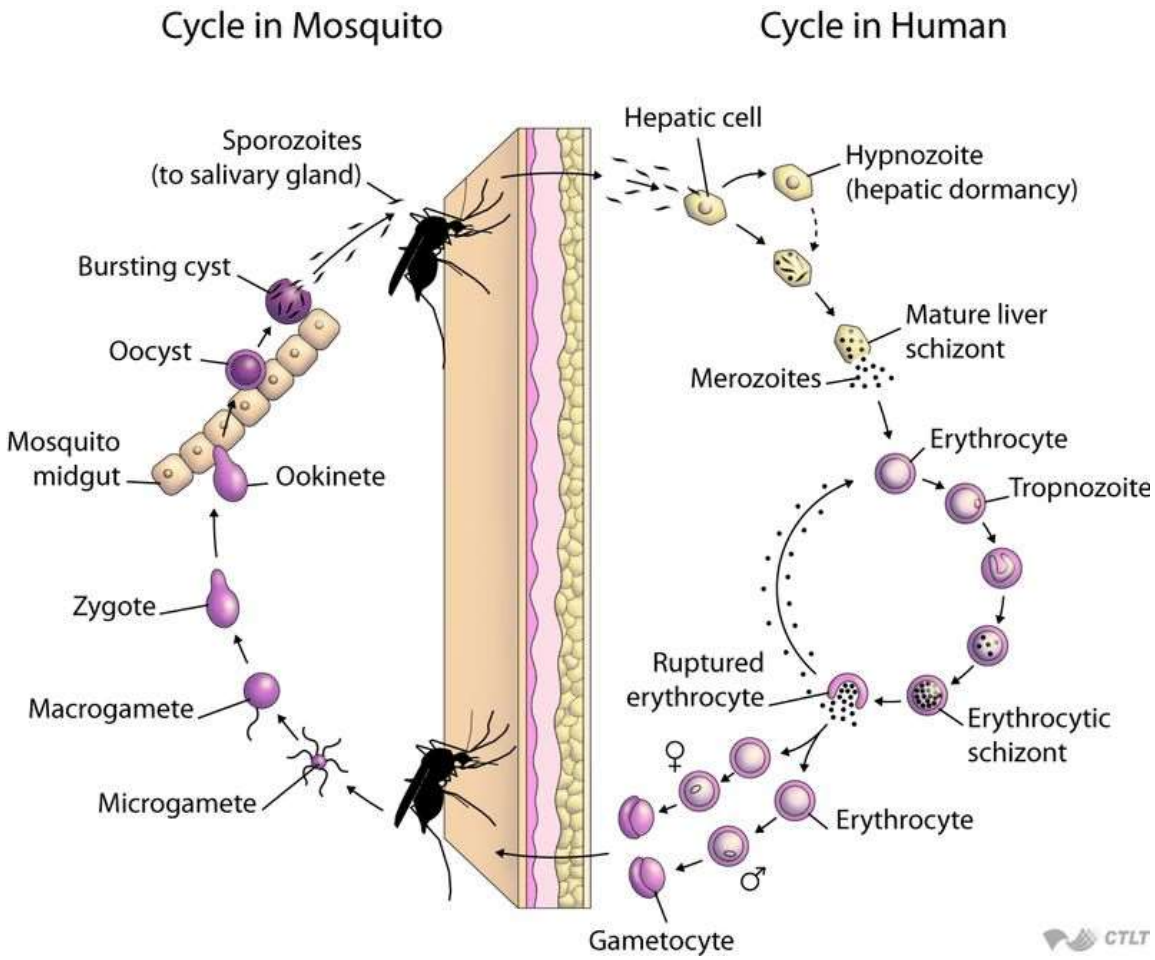
Source: http://go.nature.com/2u68ie8

Figure 1. World malaria map, tropical regions of the world still bare the greatest malaria burden. According to WHO 2017 malaria report, over 95% of malaria cases from Africa are caused by *Plasmodium falciparum* infections.

1.2.3 Malaria parasite life cycle

Malaria parasites are transmitted by female anopheline mosquitoes during a blood feed by depositing motile sporozoites under the skin dermis (White *et al*. 2014). There are more than 400 species of anopheles and almost 20 species are of public health importance with *P. falciparum* parasite being the most lethal red blood cell infecting parasite carried by the vectors (White *et al*. 2014). *P. falciparum* has a complicated life cycle which involves the human host and the mosquito as a vector (Figure 2). The asexual phase is known to take place in human and the sexual reproduction occurs in female anopheline mosquitoes (Hayton and Su 2008). When the mosquito bites a human, sporozoites in the saliva are injected with the bite, and the parasites reach the liver cells through blood. The sporozoite invades the hepatocytes where they multiply (Hsiao-Han Chang and Hartl 2014). The parasite reproduces asexually in the liver hepatocytes forming schizonts which then burst the hepatocytes releasing into the blood hundred thousand of merozoites which infect the peripheral red blood cells and start another asexual cycle (Figure 2). Several asexual reproduction events occur, which results in multiple invasion of red blood cells, with the development from trophozoite to schizont stages every two days, causing an increase in parasite load in the blood (Hayton *et al* 2008). The parasites then switch to a gametocyte stage (Hawking *et al* 19), a female anopheline mosquito takes up the gametocytes during a blood meal. Following a passage through the salivary gland, fertilization occurs between male and female

gametes. This takes place in the mosquito's intestine resulting into the formation of ookinetes and then oocysts in the mosquito gut. The oocysts rupture and release sporozoites, which then move to the salivary glands to await inoculation at the next blood meal (H.-H. Chang *et al*. 2013; White *et al*. 2014; F. Hawking, Worms, and Gammage 19).



Source: http://www.malwest.gr/Portals/0/EID_lec17_slide8-large%5B1%5D.jpg

Figure 2. Malaria parasite life cycle, involving vector and human host.

Malaria parasite infected female anopheles' mosquito bites human being, releasing sporozoites cells into the human's dermis then into the bloodstream. Sporozoites travel through the bloodstream into the liver. In the hepatocyte cells, sporozoites undergoes asexual reproduction, forming merozoite cells. Merozoites then move into the peripheral bloodstream and infect the erythrocytes. In the erythrocytes, merozoites multiply producing thousands of merozoites, which then cause the erythrocytes to burst. Newly formed merozoites infect other erythrocytes. Certain merozoites develop into gametocytes (male and female). When another mosquito bites the infected human being, it is ingesting the female and male cells (gametocytes). The mosquitos gut is where the maturation of the male and female cells occurs. The sexual reproduction occurs between the male and female gamete resulting in the formation of a zygote. The sporozoites are then formed from the zygotes through multiplication of the zygotes, the sporozoites formed then move to the mosquito's salivary glands. Another life cycle is initiated when the mosquito bites and inoculates sporozoites in another human being.

1.2.4 Malaria Clinical Presentation

Malaria clinical presentation can either be severe or uncomplicated, and it remains a major risk factor for children under sixty months of age and pregnant women (White 2004). There is no specific clinical manifestation of malaria infection since it can manifest in many forms: fever, headache, muscle ache, fatigue, gastrointestinal disturbance, joint pain, profuse sweating mostly at night, vomiting, nausea, chills and diarrhea, which may occur singly or in combination (Tahita *et al*. 2013). When a malaria parasite infects an individual, the fever lasts for 12 hours and this results in exhaustion and dehydration of the patient (Collins and Jeffery 2005). Several malaria infections can result in severe anemia, which is due to a decrease in the number of red blood cells in the body (Imbert *et al*. 1997). Malaria clinical presentations vary depending on the infecting *Plasmodium* species. An infection by *P. falciparum, P. ovale* and *P.vivax* results in fever every day to 48 hours following disease onset due to schizont rupture and destruction of red blood cells (Collins and Jeffery 2005). Of all the *Plasmodium* species, *P. falciparum* is known to cause the most severe malaria and highest mortality rate. It causes convulsions, severe headache and high fever (Lalloo *et al*. 2007). *P. falciparum* cause severe malaria resulting in cerebral malaria, acute renal failure and pulmonary edema (Trampuz *et al*. 2003). After treatment of infections resulting from *P. vivax and P. ovale,* some schizonts which "hide" in the hepatocytes, become dormant for a period ranging from 90 days to five years (Lalloo *et al*. 2007). These schizonts may later enter the blood stream periodically causing malaria relapse (Collins and Jeffery 2005). The prolonged stay of schizonts in the host liver may results in a new variant of *Plasmodium* with different phenotypes from the initial infecting parasite due to evolution in the host (Gardner et al. 2011).

1.2.5 Malaria diagnosis

Fast and accurate detection of malaria parasites is essential for the treatment of infected individuals and it also helps in controlling further transmission of malaria in a population (Oriero *et al*. 2015). The microscopy test remains the main diagnostic tool in most resource-limited settings (Musuva *et al*. 2017). In addition to microscopy, rapid diagnostic test (RDT), isothermal amplification, serology, fluorescent microscopy and Polymerase Chain Reaction (PCR) can also be used for the detection of malaria parasite in the blood (Britton, Cheng, and McCarthy 2016; Torrus *et al*. 2015). Before administration of antimalarial therapy in most health facilities, malaria parasite detection is done using microscopy by examining the peripheral parasites in a thick or a thin (used in identifying the species of infecting *Plasmodium*) blood smear or rapid diagnostic test (RDT) which detects the malaria parasite antigen (Histidine rich protein 2, HRP2) in the blood as recommended by the World Health Organization (WHO 2010; Guy 1992). PLDH-based tests has an 85% sensitivity while HRP2 has 92%, and according to a study which was conducted in Kampala, Uganda in children by Hopkin et

la (Hopkins et al. 2007) the pLDH-based tests had 100% in malaria detection and HRP2 had 93%. All these tests can still be used for the detection of *P. falciparum* infection before any ACT drug is prescribed (Hopkins et al. 2007). Accurate microscopy tests can also be used to quantify malaria parasites (thick smear) as well as to detect different species (thin smear) of malaria parasites (Payne 1988;). The molecular diagnosis of malaria parasite is the most accurate diagnostic test to date due to the high sensitivity of PCR with the lower limit of detection of 0.22 parasites per microliter (Britton, Cheng, and McCarthy 2016). It also provides an accurate determination of the *Plasmodium* species within a population of parasites (Thongdee *et al*. 2014).

The loop-mediated isothermal amplification test is a fast and simple nucleic acid amplification tool for malaria diagnosis (Mori *et al*. 2001). The detection of parasite's nucleic acid is made easy by interpreting the turbidity test. A higher turbidity shows the presence of high parasitemia while a low turbidity indicates the presence of low parasitemia (Mori *et al*. 2001; Britton, Cheng, and McCarthy 2016). Loop-mediated isothermal amplification has been used in malaria surveillance in research centers and major malaria reference laboratories to provide a molecular diagnostic tool in the fight against malaria in different regions of the world (Sattabongkot *et al*. 2014). Immunochromatographic tests include RDT, which detects the presence of histidine-rich protein 2, which is specific to *P. falciparum*. Some RDTs detect parasite lactase dehydrogenase (pLDH) or aldolase, which can also be used to distinguish *P. falciparum* infections from non-*P. falciparum* infections, for example *Plasmodium vivax* (WHO 2010).

1.2.6 Malaria Treatment

With accurate and timely diagnosis, malaria is treatable with antimalarial such as Artemisinin Combination Therapy (ACT) and quinoline drugs (Boyce *et al*. 2015; Kazembe, Appleton, and Kleinschmidt 2007). The World Health Organization (WHO, 2015), recommends four main factors to be considered before the administration of antimalarial drugs to malaria patients. It includes whether the patient is suffering from complicated or uncomplicated malaria, the type of infecting species of malaria parasite, high risk groups such as infants and pregnant women, and treatment response, depending on drug resistance patterns in a given geographic location (World Health Organization 2015a). *P. falciparum* resistance to almost all the available antimalarial drugs prompted the World Health Organization to endorse dual or triple dosage of antimalarial drugs to be used in malaria treatment. These antimalarial drugs which are made up of molecules with distinctive mode of action or different target enzymes (Paloque *et al*. 2016). Combination therapy is much more effective and should the parasite develop resistance to one component of the drug the partner drug molecule will clear the resistant *P. falciparum* parasites. The commonly used ACT drugs are: Artemether- Lumefantrine, Artesunate-Amodiaquine, Artesunate-Mefloquine, Dihydroartemisinin-

piperaquine and Sulphadoxine-Pyrimethamine administered over a period of 72 hours (World Health Organization 2015a).

For more than four decades, Chloroquine has been used in the treatment of uncomplicated infections of *P. vivax*, *P. ovale*, *P. malariae*, and *P. falciparum* (Visser *et al*. 2014). Sulphadoxine-pyrimethamine (SP), doxycline, mefloquine, atovaquone, proguanil hydrochloride combination, quinine, and primaquine were also used in the treatment of malaria (Hayton and Su 2008; Lalloo *et al*. 2007). However, Sulphadoxine-pyrimethamine, chloroquine and amodiaquine must not be prescribed to patients who are infected by *P. falciparum* after diagnosis because of the high- level of resistance to these drugs which has developed because of their usage as monotherapy drugs (Paloque *et al*. 2016). In 2015, the World Health Organization adopted the use of ACTs, for the treatment of uncomplicated malaria caused by *P. falciparum* in malaria endemic locations of the world (World Health Organization 2015b). Quinine was recommended for treatment of uncomplicated *P. falciparum* infection in pregnant women in their first trimester as well as in infants, while Artesunate and artemether use were recommended in the second and third trimester. The WHO strongly recommends the use of Sulphadoxine/Pyrimethamine for Intermittent preventive treatment (IPT) in pregnant women in Sub-Saharan Africa and other malaria endemic regions of the world, where a large number of deaths are registered due to malaria in pregnancy (Vinayak *et al*. 2010). Sulphadoxine is an antifolate which inhibits the enzymatic functions of the *dhps* enzyme used in the folate biosynthesis in *P. falciparum* parasite. Patient behavior, such as non-adherence to antimalaria treatment, provides a prime advantage for to undergo adaptive evolution resulting in new species of parasites with a different genetic make-up which become later non-responsive to antimalarial treatment (Vinayak *et al*. 2010).

1.2.7 Malaria Vaccine

Despite efforts to develop malaria vaccines for the past 70 years (CLYDE et al. 1973), no effective malaria vaccine is available on the market (Hoffman et al. 2015). An effective vaccine is needed to reduce disease burden (Keitany, Vignali, and Wang 2014). The complex *P. falciparum* life cycle (see Figure 2) within the mosquito gut and the human host is a great challenge in malaria vaccine development (Draper *et al*. 2015). The differential antigen expression after the inoculation of the sporozoites in the skin dermis, in the hepatocytes before migration to the peripheral blood stage, and the capability of undergoing antigenic variation complicates vaccine development process (Keitany, Vignali, and Wang 2014). Several vaccines have been developed against malaria. The vaccines were based on PfCelTOS, PfMSP1, PfAMA1, PfRH5, PfSERA5, PfGLURP, PfMSP3, Pfs48/45 and Pfs25 (Draper *et al*. 2015). Despite all these attempts there is no licensed malaria vaccine to date. This can be explained in part to the complex malaria parasite's life which involves the human host and mosquito

vector, and the ability of *P. falciparum* to curb and evade the human host immune system, which poses major challenges (Lusingu and Von Seidlein 2008). RTSS has been shown to be effective and is currently undergoing clinical trials in Sub-Saharan Africa specifically in East Africa (Tanzania and Kenya) and Central Africa, where it has been shown to be safe and immunogenic in children below the age of 5 years, but the vaccine protection reduces after six months following its administration (Sacarlal *et al*. 2008; Lusingu and Von Seidlein 2008; Campo *et al*. 2014). Despite the continued efforts and research to discover an effective vaccine, more challenges than gains are still at play. Differential expression of *var*, and *msp1, msp2* surface protein genes during the life cycle of *P. falciparum* complicates the efforts in discovery of the right vaccine (Lusingu and Von Seidlein 2008; Frank Hawking, Wilson, and Gammage 1971). Complex morphological development of *P. falciparum* from stage I-V (See Figure 2) enables the parasite to survive within the host and evade drugs.

1.2.8 Immunity

After several malaria infections, people living in high malaria transmission settings develop partial immunity against malaria (Arama et al. 2014). This partial immunity does not protect an individual from subsequent malaria infections, but it protects them against severe effects of malaria infections (Schwenk et al. 2011; Arama et al. 2014). High malaria mortality rate and severe illness in malaria transmission regions of the world occur in young children below the age of five, pregnant women and infants (Roca-Feltrer, et al. 2008). Infants and children get easily infected with malaria because their immune system has not matured to provide effective protection against the infecting malaria parasite. Immune suppression during pregnancy in women, make them highly susceptible to malaria infection (Roggelin et al. 2014). In the placenta, malaria parasites attach to the blood capillary, which results in the perturbation of blood circulation in the placenta resulting in reduced supply of nutrients to the fetus (Fitri *et al*. 2015). *P. falciparum* infection in pregnant women is the leading cause of stillbirth and miscarriages reducing the survival rate of infants born in malaria endemic areas (Roggelin et al. 2014). The *P. falciparum* gametocyte must evade the human host immune clearance during development for it to mature. This is thought to be due to sequestration of bone marrow for a better part of their development period, and it is after that when they are released to circulation upon maturation (Delves *et al*. 2016). The gametocytes also become insensitive to the antimalarial regimen which is used to treat malaria (Delves *et al*. 2016). Insensitivity problems posed by mature gametocytes to antimalaria drugs results in a situation where an individual is cured of malaria but still harbors gametocytes. This promotes the disease circulation in the population if proper monitoring is not put in place. This explains why transmission blocking vaccines have not come into play.

1.2.9 Motivation

Kisumu County is located along the shores of Lake Victoria and is a stable malaria transmission setting with several malaria hotspots including the Kombewa and Maseno divisions (Bejon *et al*. 2014; Sifuna *et al*. 2014). Until recently, the analysis of *P. falciparum* genomic sequence data has not provided a unifying explanation for the parasite evolution following the introduction of Artemisinin Combination Therapy treatment in Kenya even though some SNPs have shown significant association with elevated IC_{50} values (Ngalah *et al*. 2015). More insight can be obtained by understanding the impact of ACT pressure on the *P. falciparum* genome and their evolutionary effect. A better understanding of the evolutionary dynamics occurring during treatment in *P. falciparum* is needed to help control the disease. In this thesis, we aim to understand the evolutionary changes occurring within-host due to ACT usage and how this could affect the treatment outcome in Kenya. The findings will enable us to identify the priority areas of further research to fully understand the evolutionary causes of drug resistance and management of malaria in the population.

1.3.0 Study Aim

To determine the within-host evolution of the *P. falciparum* parasite during treatment with ACTs and investigate the presence of drug resistance signatures. Main objective

1.3.1 Main objective

To investigate the evolutionary changes in the *P. falciparum* genome due to Artemisinin Combination Therapy intervention in Kenya between 2013 and 2015.

Specific objectives

1. To determine the presence of selection signatures in the *P. falciparum* genome due to ACT usage between 2013 and 2015.
2. To determine the association between Single Nucleotide Polymorphisms and IC_{50s} of ACT drugs, clearance half-life and clearance rate.
3. To investigate the effect of single nucleotide polymorphisms on the Kelch 13 propeller domain.
4. To determine within-host population differentiation as a measure of evolution.

Chapter 2. Materials and Methods

2.1 Study site

Overview

The sample collection was conducted within the Kombewa and Maseno Divisions of the Kisumu West District of Kisumu County in Kenya. Study activities were coordinated from the Kombewa Clinical Research Center located in Kombam village, Kombewa (Figure 3). The study site is located about 40 km west of Kisumu city, the administrative capital of Kisumu County. The study divisions cover an area of about 369 km^2 stretching on the North-Eastern shores of Lake Victoria where malaria is holoendemic (Watsierah *et al*. 2011). The study area has been cartographically mapped by the KEMRI-WRP Kisumu West Health and Demographic Surveillance System (KWHDSS) using Global Positioning System (GPS) technology. "The KWHDSS is a longitudinal population registration system designed to track the evolving demographic and health status of the study populations over time. The KWHDSS population is monitored by field staff visiting each household bi-annually to capture health and demographic information. Various studies nested on this platform take advantage of the sampling frame inherent in the Health Demographics Surveillance System (HDSS), whether at individual, household/compound or regional levels" (Sifuna *et al*. 2014). Kombewa (34°45' E 0°10 S) lies within the Lake Victoria basin (Sifuna *et al*. 2014) (see Figure 3). Malaria in Kombewa is holo- endemic, with *P. falciparum* accounting for more than 95% of single-species infections. Little information is available regarding mixed infections and Malaria transmission occurs throughout the year. The 'long rainy season' from late March to May produces intense transmission from April to August. The 'short rainy season' from October to December produces another less intense transmission season from November to January" (Bejon *et al*. 2010). *Anopheles gambiae* and *Anopheles funestus* are the major vector species with a noted increasing population of *Anopheles arabiensis* in recent years. There is limited published malariometric data from Maseno (34°36' E, 0°0' S). For these two study sites we set out to understand the evolutionary forces conferring drug resistance development in the most lethal *P. falciparum*.

2.1.0 Study design

This was a two-arm randomized open-label trial study which was conducted at the Kombewa clinical Research Centre. Patients presenting with uncomplicated malaria at health facilities within the study area was referred to the research center and enrolled into the study after signing an informed consent form. Those who had parasite density ranging from 2000 to 200000 parasites/µl, willing to comply to the study protocol for the study period, and had a mono-infection of *P. falciparum* was

included in the study. Requirements were adults and children aged between six months and sixty-five years with the minimum weight of 11 kg and having fever with a body temperature greater or equal to 37.5^{0C} or an history of fever within 24 hours prior to presentation to the facility. Detailed information on genotyping of SNPs using Sequinome Mass Array technology has been published by Ngala *et al* and is available online (Ngalah *et al*. 2015). Fifty-nine (59) samples from 2013 and sixty-eight (68) from 2015 were analyzed. Those satisfying the above conditions and were willing to remain in the facility for a period of 72 hours were also included in the study. Subjects showing signs of severe malaria who were unable to take oral medication were excluded from the study. Ethical approval for this study was obtained from Kenya Medical Research Institute Ethics review committee (SSC #2722), and the University of Cape Town, Human Research Ethics Committee (HREC REF:539/2017) gave approval for data analysis. In addition to the above-mentioned IRBs, this study was also approved by Walter Reed Army Institute of Research (WRAIR #1935 and #1935A).

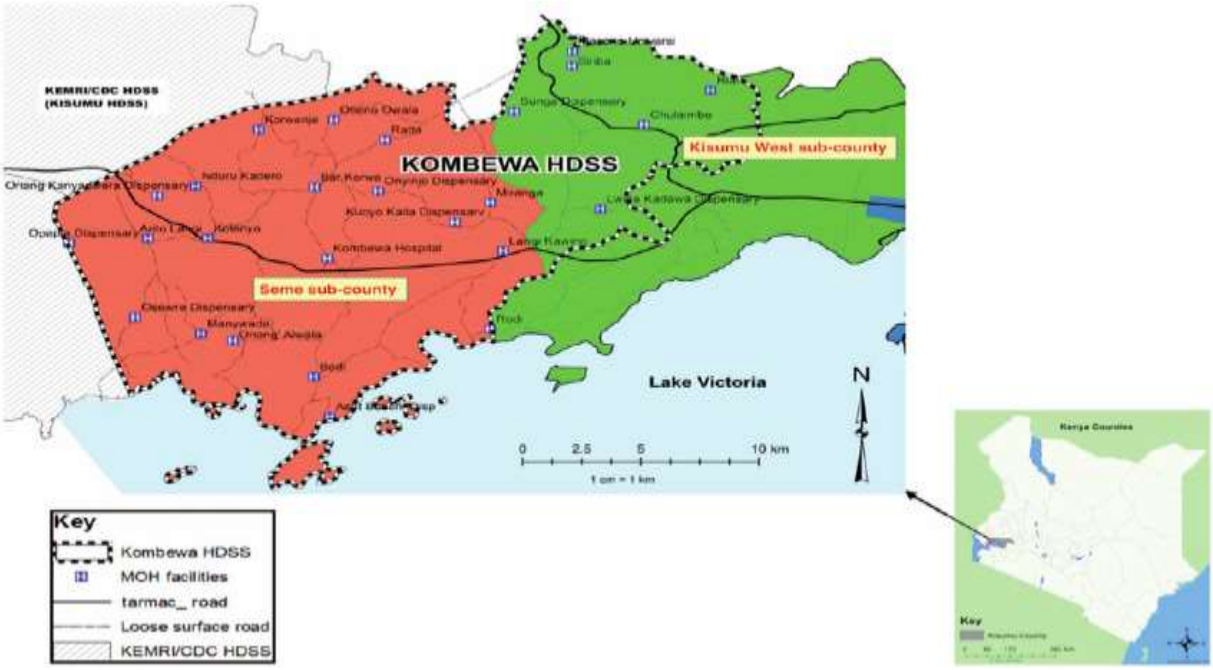


Figure 3. Study site showing the sample collection locations; Source (Sifuna et al. 2014)

2.1.1 Assessing the presence of selection signatures between 2013 and 2015

To identify whether Artemisinin Combination Therapy (ACT) intervention causes selection in *P. falciparum* parasites from Kenya, Tajima's D, Fu & Li's F*, and Fu & Li's D* summary statistics were calculated using the DNA sequence polymorphism analysis software, DnaSP V6 (Librado and Rozas 2009). The analysis was conducted to evaluate whether the *P. falciparum* Kelch 13 propeller domain molecular sequence displays evidence of departure from the neutral model of molecular evolution (Tajima 1989). A total of forty-six sequences from the 2013 sequence data set and seventy-two sequences from the 2015 data set were analyzed. Multiple Sequence Alignment (MSA) was done using the MUSCLE alignment program (Edgar 2004) implemented in the Aliview software (Larsson 2014), which was also used to visualize the alignment of the sequences using the SeaView V4 (Gouy, et al. 2010) visualization program. Aligned sequence data were then exported into DnaSP V6 (Librado et al 2009) for calculations of Tajima's D, Fu and Li D*, and Fu & Li's F* (Fu and Li 1993). In DnaSP the simulation is used to estimate the Tajima's D (Tajima 1989), Fu and Li's D***** (Fu and Li 1993), Fu and Li's F***** (Fu and Li 1993) genomic summary statistics. Information on the allelic variation in the sequence data is used in the test (Ramírez-Soriano *et al*. 2008). These three test statistics are based on the assumption that under the neutral model, there is a correlation between the estimated number of polymorphic sites and the average nucleotide diversity (Tajima 1989). In the calculation of Tajima's D (Tajima 1989), average nucleotide diversity between two sequences (π) and the total population nucleotide diversity are taken into account (Tajima 1989). Haplotype diversity which is the analogue of the allele frequency-based calculation was also conducted. Fu and Li's D* (without an outgroup) tests the difference between total number of singletons and the total number of mutations occurring within the sequence (Fu and Li 1993). Fu and Li's F* (without an outgroup) summary statistic is based on the difference between the number of singletons and the average nucleotide difference between sequence pairs (Fu and Li 1993). The genomic summary statistics values obtained from this analysis were used to interpret the results under the neutral model of evolution (Fu and Li 1993; Tajima 1989; Ramírez-Soriano *et al*. 2008). To infer the directionality and magnitude of adaptive evolution acting on the *Pfk13* protein coding gene (*PF3D7_1343700*) due to ACT intervention, the synonymous mutation per synonymous site and non- synonymous mutation per non-synonymous mutation ratio were calculated. Nucleotide sequences from 2013 and 2015 sample data were first converted into amino acid sequences by EMBOSS TranSeq located at (www.ebi.ac.uk). The amino acid sequences were then aligned using the MUSCLE alignment program (Edgar 2004) implemented in Aliview (Larsson 2014). The aligned sequence data was transferred to Synonymous Non-Synonymous Analysis program (SNAP v2.1.1) webserver for

dN/dS calculation and visualization analysis. The analysis involved forty-six amino acid sequences from the 2013 dataset and seventy-two amino acid sequences from the 2015 data set. The 1st, 2nd and 3rd codon positions were included in the analysis. All the ambiguous positions were removed from each sequence pair. The summary statistics values obtained from these analyses were used to predict the adaptive evolution due to ACT intervention within the *P. falciparum* genome.

2.1.2 Determining the effect of mutations on Kelch 13 propeller domain

2.1.2.0 Using the graph-based signature method

Mutation is the primary driver in evolution as it plays a role in the introduction of diversity in a population of *P. falciparum* parasites (Pires et al. 2014). The non- synonymous mutations in the coding regions of Kelch 13 genes may provide a selective advantage or disadvantage to the *P. falciparum* parasite by affecting Kelch 13 protein stability. To understand the effects of mutations on *P. falciparum* Kelch 13 propeller domains and the development of adaptive evolution against the Artemisinin Combination Therapy drugs, the effects of point mutations (Single Nucleotide Polymorphisms) on the Kelch 13 protein were assessed. The wildtype Kelch13 protein data file (PDB_ID_4yy8) was downloaded from the 3D protein database website (www.rcsb.org). The effect of single nucleotide polymorphisms (SNPs) on the Kelch 13 protein structure was determined using the mCSM (Pires, Ascher, and Blundell 2014) webserver (http://biosig.unimelb.edu.au/mcsm/). The input file (text file) containing the chain name in the first column and the second column having non-synonymous amino acid substitutions information at different loci within the Kelch 13 domain was created from the SNP array data. The wild-type protein data bank (pdb) file together with the mutation list was submitted for the analysis.

2.1.2.1 Using SDM to infer the effects of mutations on Kelch 13 protein structure

To further validate the mCSM results, an environment specific substitution table (ESST) (Pandurangan *et al*. 2017) was used to calculate the energy difference between the wildtype Kelch 13 protein and the mutant protein structure. The Site Directed Mutagenesis (SDM) web server (http://131.111.43.103/) was used in the calculation of the thermodynamic energy difference between wildtype and mutant Kelch 13 propeller domain (Pandurangan *et al*. 2017). The text file containing the list of point mutations at different codons was created as described in the mCSM procedure above. The Kelch 13 protein data file in Protein Data Bank (PDB) format was downloaded from the protein databank (PDB_ID 4yy8). This, together with the text file containing the mutation list were submitted to the server for the thermodynamic difference calculation.

2.1.2.2 Location of point mutations on Kelch 13 protein

Determination of non-synonymous point mutation locations on the Kelch 13 protein structure is fundamental as it enables the visualization of the loci under evolution due to drug pressure. Using the PRotein Interactive Modelling (PRIMO) online homology modelling tool (Hatherley *et al*. 2016) (https://primo.rubi.ru.ac.za), the wildtype structure of *P. falciparum* Kelch 13 protein structure was determined. The Kelch 13 sequence comprising of 726 amino acids was downloaded from the UniProt database (Protein ID: Q8IDQ2). The *PF3D7_1343700* was used as a target protein sequence and using the Basic Local Alignment Search Tools (BLAST) algorithm the best protein template was identified from the Protein Databank. The PDB ID 4yy8 with 100% identity, 53% coverage and a resolution of 1.81 Å (Angstrom) was identified. The target-template alignment was done using the MUSCLE (Edgar 2004) alignment program within PRIMO. The modelling of the target sequence was done using the slow refinement modelling algorithm.

2.1.2.3 Artemisinin Combination Therapy drugs and Kelch 13 protein interaction

The mutations associated with the high Inhibitory concentration 50% are found on chain A as shown in figure 6. Since the Kelch 13 propeller protein is made up of two main chains, we investigated the ACT drug interaction sites with the Kelch 13 protein domain. The drug/ligand was extracted from the wildtype Kelch 13 protein, which was downloaded from the UniProt database in PDB format. Babel was used to convert the wildtype drug/ligand file to .sdf format. Using the Posview web server (https://proteinplus.zbh.uni-hamburg.de/), the wildtype protein file and the ligand .sdf files were uploaded for the analysis.

2.1.3 IC_{50}, Slope-half life and SNP association analysis in 2015

Point mutations (SNPs) have been associated with both Artemisinin and quinoline derivatives treatment failure and increased parasite clearance rate from Sub-Saharan Africa and in South East Asia (Picot *et al*. 2009; Foote *et al*. 1990; Djimdé 2001; Amaratunga *et al*. 2012; Eckstein-Ludwig *et al*. 2003; Ariey *et al*. 2014). The association between SNPs and the half maximal inhibitory concentration 50% of *P. falciparum* isolates were done to determine the relationship between IC_{50s} and SNPs at different loci using SNParray and IC_{50} data. The main goal was to determine if there was an association between these genetic variants and the increased IC_{50} values of different ACT drugs. The following drugs were used for the analysis: amodiaquine, mefloquine, dihydroartemisinin-piperaquine, artemether and lumefantrine. A custom python script was used to clean the data by removing missing data points. The categorical predictors (SNPs) were then converted to binary numbers with zero (0) denoting wild-type variant and one (1) denoting mutant variant using another custom python script. After cleaning the data, R version 3.4.2, was then used in the downstream

analysis of the data. A general linear regression model in R v3.4.2 (R Development Core Team 2016) was used to predict the association between SNPs, slope half-life and IC_{50} values. A Pearson correlation test was done to determine the correlation between different antimalarial drug IC_{50} values as it may provide an insight into the similarity in their therapeutic activities. Sometimes, therapeutic agents with negative correlation values can reveal synergistic partner drugs for rotational deployment in population. A cut-off alpha value of ≤ 0.05, was considered to be statistically significant. The student t-test was used to determine if there was a significant difference in parasite clearance rate between Artemether lumefantrine and Artesunate amodiaquine treatment arms respectively.

2.1.4 Determination of within-host population differentiation as a measure of evolution

P. falciparum parasites at different time points were considered as distinct parasite populations. The hour zero day zero (H0D0) parasites were considered as the first population (population 1), day one hour twenty-four (D1H24) as population 2, day two hour thirty (D2H30) as population 3, day one hour thirty-six (D2H36) as population 4, day one hour forty-two (D2H42) as population 5, day two hour forty-eight (D2H48) as population 6, day twenty-eight as population 7, day thirty-five as population 8 and day forty-two as population 9. Microsatellite data was managed in an Excel spread sheet, from which the data was formatted to form the input for other microsatellite analysis programs. A custom python script was used to remove the missing data points. The value 2 was entered in cell A1 since our data was a two-column format. Parasite population names were entered in the first column, ensuring that no cell remained empty. Since the *Plasmodium* genome is haploid within the human host, the alleles were entered twice to mimic the diploid state of the organism. The input data format was prepared in Excel and then saved in tab delimited format. Using the Microsatellite Analyzer V 4.05 (Chu 2003), *P. falciparum* population differentiation at different time points was estimated by calculating the Wright F_{ST} values. During the analysis, the input file was saved in the same directory where the MSAanalyzerMr.exe executable was, which was then used to specify the input parameters and run the analysis. The Microsatellite Analyzer was also used to generate other data file formats ("data wrangling") such as gene pop, structure and Arlequin input data files. Genepop, the output data format from Microsatellite Analyzer was transferred to R v3.4.2 (R Development Core Team 2016) for the downstream analysis. Various population genetic analysis packages were installed in R v 3.4.2 (R Development Core Team 2016) which include: poppr (Kamvar, et al. 2015), diveRsity (Keenan *et al*. 2013), adegenet (Jombart 2008), vegan (Oksanen *et al*. 2016) ggplot2 (Wickham 2009) and Discriminant Analysis Principle Components (DAPC) (Jombart et al. 2010). This was used to visualize genetic hierarchical clustering of *P. falciparum* during treatment, allele richness and F_{ST}. F_{ST} (Nei and Genetics 1973) was used to measure the

genetic variability within and between *P. falciparum* populations during treatment because it considers differences in multi allele frequencies. For interpretation of $F_{ST} < 0.05$, was defined as low or no genetic differentiation, ≥ 0.05 $F_{ST} < 0.15$, moderate differentiation and finally $F_{ST} \geq 0.15$ denoting large genetic differentiation (Gatei *et al*. 2010).

2.1.5 Phylogenetic tree construction

Multiple sequence alignment (MSA) was done using MUSCLE (Edgar 2004) in Aliview (Larsson 2014). The Aligned sequence data was saved in Newick format which was then transferred to the Recombination Detection Program (Martin et al. 2000) for phylogenetic tree construction. A Maximum likelihood phylogenetic tree was constructed using the General Time Reversible (GTR+ G4) substitution model. The iTol phylogenetic tree editor and annotation program available at (https://itol.embl.de/) was used to edit the tree by adding the parasite clearance slope half-life information (phenotype).

Chapter 3. Results

This research sought to investigate the evolutionary changes in the *P. falciparum* genome due to Artemisinin Combination Therapy intervention in Kenya using samples collected in Kombewa and Maseno divisions between 2013 and 2015. We aimed to determine the presence of selection signatures in the *P. falciparum* genome due to ACT usage between 2013 and 2015. In addition, the study aimed to determine the association between Single Nucleotide Polymorphisms and half maximal inhibition concentration of ACT drugs, clearance half-life and clearance rate. Finally, we investigated the effect of SNPs on the Kelch 13 propeller domain and determined the within-host population differentiation as a measure of evolution.

Wright's F-statistic is very popular since it has been around for a long period of time (Holsinger et al 2009). F_{ST} gives a very important insight which help in the explanation of different evolutionary processes that affects the genetic variations within and among populations of interest. It is the most widely used genomic summary statistic and in evolutionary genetics (Holsinger et al. 2009). The Wright's F-statistic, has been used to identify regions of a genome which are under selection (Chen et al. 2004). The F_{st} values range from 0 to 1, with the zero values indicating no differentiation and 1 indicates that the alleles under study are fixed (Chen et al. 2004; Keenan et al. 2013). D Jost, G_{ST} and G'_{ST} are the analogues of the F_{ST} and their interpretations are the same as those of F_{ST} (Keenan et al. 2013). Tajima's D test is used to compare the average number of pairwise differences and the total number of segregating sites in a sample being analyzed (Ramírez-Soriano et al. 2008). The negative Tajima's D test implies positive selection, whereas positive Tajima's D indicates negative selection while Tajima's D of 1.00 shows that the genomic sequence under study is under balancing selection (Ocan et al. 2016; Oyebola et al. 2017; Tajima 1989).

3.1.0 Neutral models of evolution between 2013 and 2015

In 2013, the *P. falciparum* parasite genome was under balancing selection as shown by the significant positive Tajima's D value of 4.53194 and a p-value less than 0.001 (Figure 4). Since the parasite's genome was under balancing selection, the wildtype alleles remained longer in this parasite population since the introduction of ACT drugs in 2006. There was a lack of SNPs as shown by a significant positive Fu & Li D* value of 2.13380 (Figure 4), a p-value of 0.02 and a Fu &Li F* value of 3.62142 and a p-value of 0.02. *P. falciparum* parasites in 2013 showed a low level of variability with the haplotype diversity (Hd) parameter of 0.552 and a nucleotide diversity (π) coefficient of 0.20892. In 2015 *P. falciparum* parasites showed a higher level of diversity (Hd = 0.832) with a nucleotide diversity (π) of 0.00302. There was a significant negative Tajima's D (-2.42910) (Figure 4), and p-value < 0.01, indicating that the *P. falciparum* genome in 2015 was under a selective sweep. There were excess singletons as shown by the significant Fu and Li's D* statistics (-5.2712, p-value

<0.02) and a significant negative Fu and Li's F* (-5.0045, p-value< 0.02).

3.1.1 Synonymous and non-synonymous mutation between 2013 and 2015

The averages of all pairwise comparisons of sequence data in 2013 were dS=0.0039 and dN=0.0050 (see Figure 4). The dN/dS (Li 1993) ratio was 1.0299 (Figure 5), an indication that in 2013, the *P. falciparum* genome was under neutral selection as shown by the very low rate of synonymous and non-synonymous mutations at the protein coding loci. This evidence is supported by significant positive values of Tajima's D, Fu and Li's F* and Fu & Li's D*. The synonymous (dS) mutation rate was 0.0108 and non-synonymous (dN) rate was 0.0052 in the 2015 sequence data set (Figure 4). The dN/dS (Li 1993) ratio was 2.84, an indication that in 2015, the *P. falciparum* genome was under strong positive selection (see Figure 4).

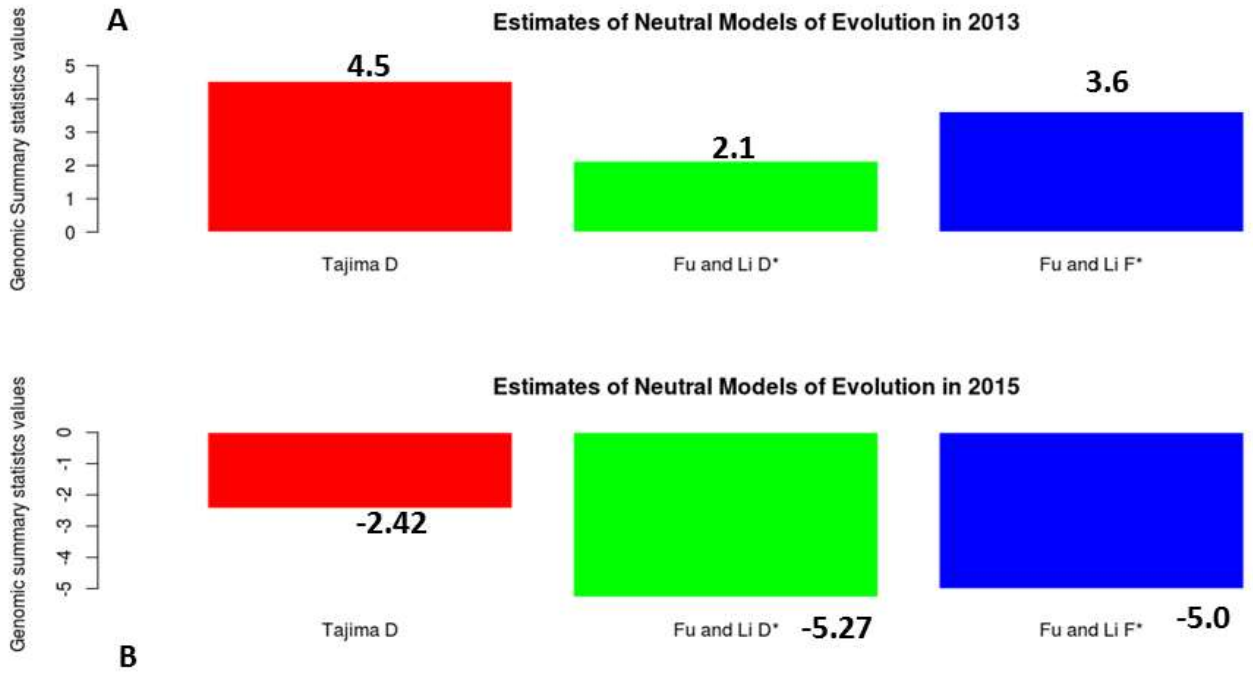


Figure 4. Estimates of neutrality model of evolution between 2013 and 2015. In 2013 Tajima D, Fu and Li D*, Fu and Li F* were all positive, signifying balancing selection and lack of singletons respectively, while in 2015 Tajima D, Fu and Li D*, Fu and Li F* were all negative, an indication of selective sweep and increased number of non-synonymous SNPs.

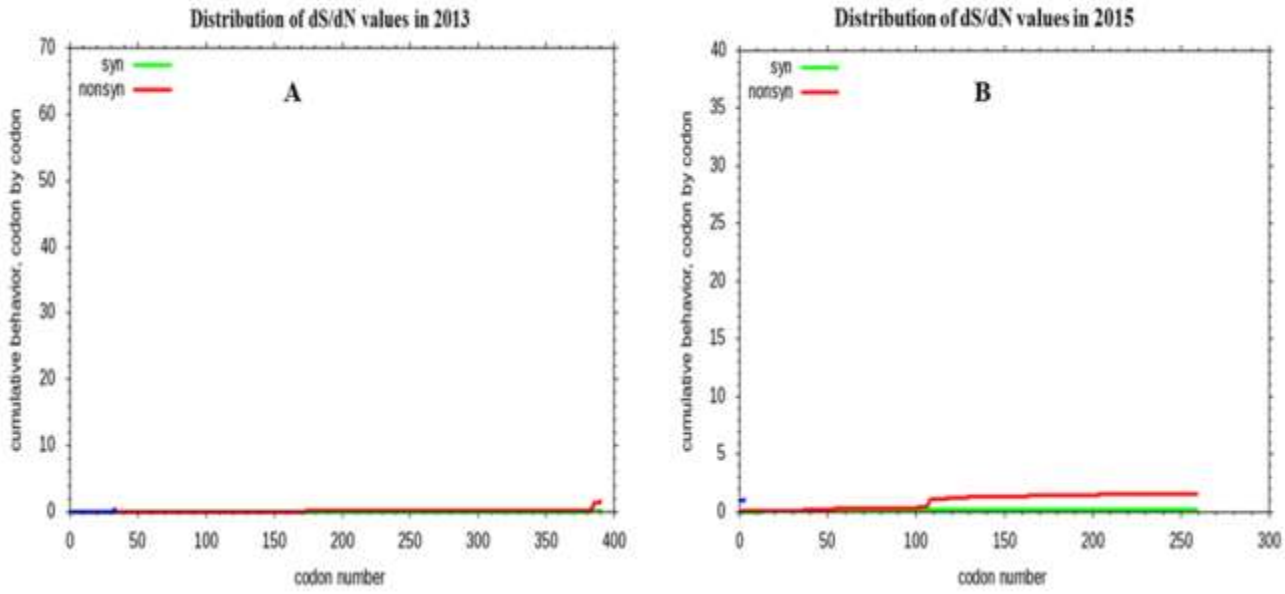


Figure 5. Synonymous and non-synonymous mutations between 2013 and 2015, the synonymous and non-synonymous mutations in 2013 were very low, while in 2015 there was an increase in non-synonymous point mutations, which is consistent with the SNP array data results. Evolutionary effects of mutation on Kelch 13 propeller domain

3.1.2.0 Using the graph-based signature method

The Kelch 13 propeller protein is composed of two chains (chain A and B**)** (Ariey *et al*. 2014). For chain A, the change in Gibbs free energy (ΔΔ D) score was used to infer the effect of different SNPs resulting in amino acid changes (non-synonymous mutations) within the Kelch 13 propeller domain. The non-synonymous mutations at codons N632K and R561P gave a positive Gibbs energy score, indicating the stabilization effects of these mutations in the Kelch 13 protein structure (Pires *et al*. 2014). Negative Gibbs free energy values indicate that the mutation causes destabilization of the protein, with Gibbs free energy score values > -2 showing a high destabilization effect of those SNPs on the protein structure (Pires et al. 2014). From the Gibbs free energy scores, (see Table 1) the non-synonymous mutation at the D547E which occurs at highest frequency, causes minimal destabilization of the protein molecule.

Table 1: SNPs and Gibbs free energy score (mCSM webserver results). Gibbs free energy score can either be negative or positive which a destabilization and stabilization effect on protein.

PDB_FILE	Wildtype	Codon_pos	Mutant type	Gibbs score
Kelch_WT.pdb	V	568	G	-2.14
Kelch_WT.pdb	T	677	A	-1.029
Kelch_WT.pdb	G	639	R	-1.461
Kelch_WT.pdb	I	640	V	-1.564
Kelch_WT.pdb	N	632	K	0.072
Kelch_WT.pdb	L	571	S	-3.566
Kelch_WT.pdb	R	561	P	0.428
Kelch_WT.pdb	D	559	E	-0.862
Kelch_WT.pdb	A	557	T	-1.545
Kelch_WT.pdb	Y	558	N	-3.039
Kelch_WT.pdb	D	547	E	-0.711
Kelch_WT.pdb	G	496	A	-1.629
Kelch_WT.pdb	F	483	L	-2.331
Kelch_WT.pdb	F	451	S	-2.724
Kelch_WT.pdb	F	628	L	-0.912
Kelch_WT.pdb	C	542	G	-1.664
Kelch_WT.pdb	N	489	K	-0.326

As shown in Table 2, these SNPs affect the stability of the Kelch 13 propeller domain and its conformational arrangement in space. To further validate the effects of these SNPs on this protein, site directed mutagenesis which uses the conformationally constrained environment-specific substitution, gave the energy difference between the wildtype Kelch 13 propeller domain and the mutant Kelch 13 protein.

More than one hundred thousand genetic variants are associated with different disease conditions in human beings (Burgess et al. 2015). The main challenge is the identification and characterization of target genetic variants and their functional effects on the protein folding and biochemical activities. Of interest are non-synonymous SNPs that affect the function of the protein by affecting their folding in space and they also interfere with the normal protein interaction with other biomolecules in the cell (Pires *et al*. 2016; Pandurangan *et al*. 2017). To further validate the effects of our SNPs on the Kelch 13 protein (Table 1), we also used the Site Directed Mutagenesis SDM to predict the impact of these mutations on Kelch 13. These SNPs caused either destabilization or stabilization in the Kelch 13 protein. We observed the stochastics effects of these SNPs on WT_RSA (Wildtype Relative side chain Solvent Accessibility), MT_RSA (Mutant Relative side chain Solvent Accessibility), WT_OSP (Wild-type residue-occluded packing density), MT_OSP (mutant residue-occluded packing density). The ddg (ΔΔ Gibbs free energy score) showed that these SNPs cause stabilization and destabilization effects as shown in (Table 2).

Table 2: Using Site Directed Mutagenesis to predict protein stability (SDM webserver results). *WT_RSA (Wildtype Relative side chain Solvent Accessibility), MT_RSA (Mutant Relative side chain Solvent Accessibility), WT_OSP (Wild-type residue-occluded packing density), * MT_OSP (mutant residue-occluded packing density), ddg (ΔΔ Gibbs free energy score)

Mutation	WT_RSA(%)	WT_depth	WT_OSP	MT_RSA(%)	MT_depth(A)	MT_OSP	Predicted
V568G	11.5	4.9	0.4	55.8	4.1	0.26	-1.78
T677A	43.9	3.7	0.26	37.1	3.4	0.26	-0.5
G639R	0	7.7	0.65	0	8.2	0.8	-2.89
I640V	11.7	4.5	0.54	6.8	4.4	0.55	-1.97
N632K	70.4	3.2	0.24	86.7	3.3	0.15	-0.16
L571S	0	8.1	0.55	2.2	7.3	0.43	-3.07
R561P	51.5	3.1	0.13	78.9	3.3	0.17	-0.42
D559E	26.5	3.9	0.44	29.6	3.9	0.41	0.35
A557T	0	6.5	0.57	0	6.3	0.65	-1.73
Y558N	1.6	7.1	0.52	3.7	8	0.46	-2.66
D547E	25	4	0.43	41.6	3.9	0.35	-0.35
G496A	0	8.7	0.53	0	9.1	0.6	-1.7
F483L	0	7.8	0.53	0	8.2	0.46	-2.46
F451S	19	4.1	0.54	15.7	4.6	0.49	-0.64
F628L	2.7	6.9	0.54	2.9	6.8	0.5	-1.87
C542G	0	10.1	0.59	0.8	10.3	0.49	-2.09
N489K	82.4	3.7	0.18	0.18	3.8	0.2	-0.99

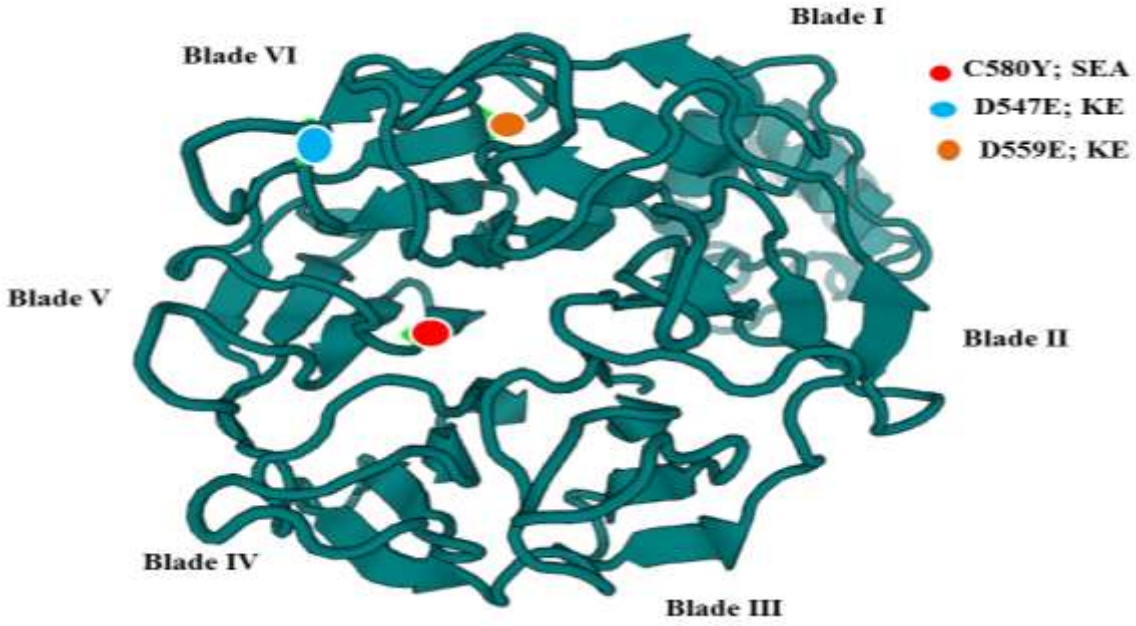


Figure 6. Location of point mutations on the Kelch 13 protein, the red sphere is a C580Y SNP which has been extensively reported in SEA and is associated with reduced parasite clearance rate. The yellow and blue spheres show the Kenyan SNPs which are likely to be associated with treatment failure.

C580Y is associated with the increased parasite clearance rate in the South East Asian (SEA) *P. falciparum* parasites (Ariey et al. 2014). D547E and D559E are associated with high risk of treatment failure in the Kenyan (KE) *P. falciparum* parasites. *PF3D7_1343700* is a homologous structure to human Kelch-like ECH-associating protein 1 (KEAP1) (Ariey et al. 2014). Both SNPs from Kenyan and South-East Asia parasites occur on the inter and intra-blade domains (Figure 6).

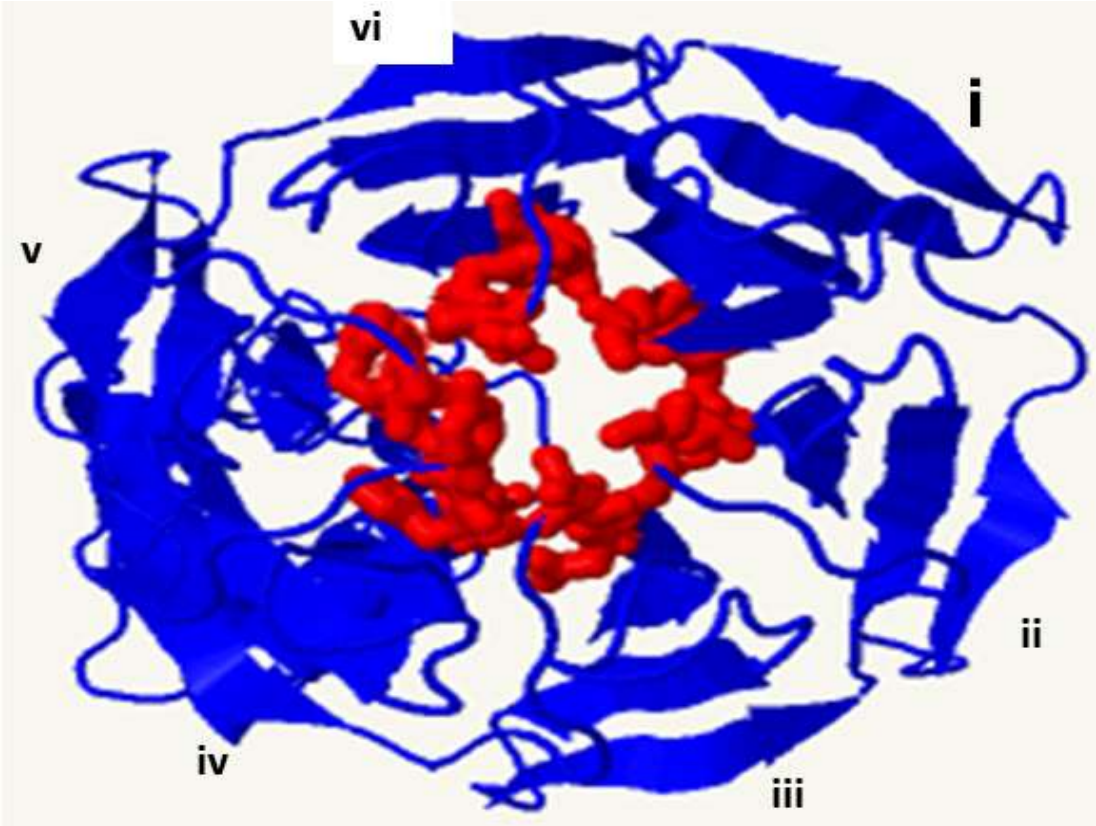


Figure 7. Kelch 13 propeller domain showing the predicted binding pocket of ACT drugs. The red molecules show the drug binding pockets and the blue colors shows the propeller domain.

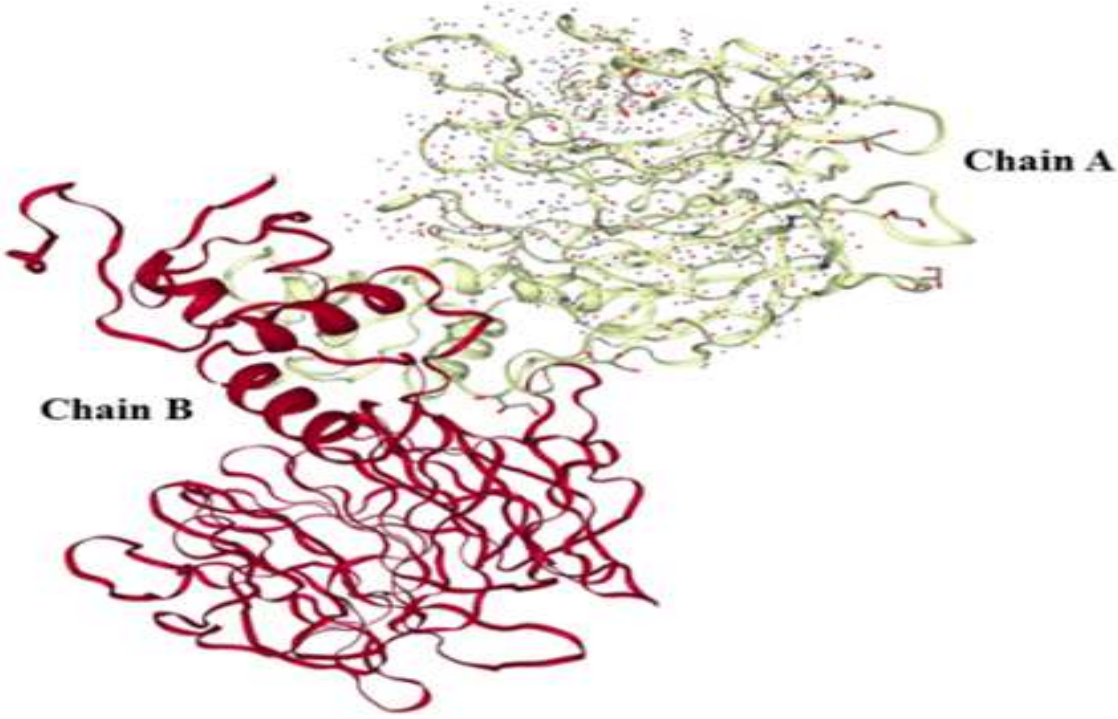


Figure 8 ACT drugs/Kelch 13 protein interactions, drug molecules are shown in red and blue dots and they mostly interact with chain A (Posview drug/ligand interaction prediction tool).

3.1.2.3 IC_{50}, Slope-half life and SNP association analysis in 2015

There was no significant association between SNPs, Inhibitory Concentration 50% and parasite clearance slope half-life. However, there was a significant association between IC_{50} and codon D559E in Amodiaquine (AQ) and Mefloquine (MQ) IC_{50} (p-value 0.0189, OR=0.9664, 95% Confidence Interval (0.9186-0.9907), and P-value 0.0175, OR=1.044, 95% Confidence Interval (1.01-1.08), respectively). Codon D547E also had a significant association with the Artesunate half maximal concentration 50%, p-value 0.0358, OR= 1.046, 95% CI (0.001 – 1.00). There was a significant negative correlation between Mefloquine (MQ) and Amodiaquine (AQ) half maximal inhibitory concentration r=-0.25, p-value 0.04308, 95% CI (-0.4577- 0.00815) (Figure 10 & Figure 11). There was no significance in the negative correlation between Artesunate and Artemether IC_{50} values, r=-0.16, p-value 0.2, 95% CI (- 0.3821–0.08346) (Figure 10). There was also no significant correlation between Lumefantrine and Dihydroartemisinin drugs, r=-0.22, P-value = 0.08, 95% CI (-0.4317 – 0.02417). In 2013, the median IC_{50} for Lumefantrine was 25.02 with the Interquartile Range (IQR) 40.89. Piperaquine had a median value of 17.185nM, IQR 17.93 (cutoff IC_{50} is still not defined); Dihydroartemisinin had a median of 8.14nM with the IQR 12.07. The Artesunate median IC_{50} value was 7.557, IQR 10.82, while mefloquine had 9.7886nM with IQR of 10.4, and finally the artemether value was 4.10439nM, IQR 3.3067 (Figure 10). In 2015, Lumefantrine had median of 12.515nM and IQR 29.62, Piperaquine 15.48nM with IQR 19.42, Dihydroartemisinin had 3.23 and IQR of 17.14. The Artesunate median IC_{50} was 7.32nM with IQR 11.09, mefloquine 8.496nM with IQR15.01 and finally artemether had 1.90nM and IQR 9.24 (Figure 12). A significant difference between the two treatment arms (Artemether Lumefantrine and Artesunate Mefloquine) was detected, P-value 0.0005574, 95% CI (-0.09837 - 0.02822), mean of 0.2798 and 0.3431 respectively. Using the same t-test model a significant difference between the parasite clearance slope half-life between the two treatment arms was observed, P- value 0.001202, 95% CI (0.1518-0.5999), mean of 2.598 hours and 2.222 hours respectively. The median parasite clearance half-life in the Artemether lumefantrine treatment arm was 2.47 (1.48-4.22) and the Artesunate amodiaquine arm was 2.35 (0.98- 3.56) (Figure 9).

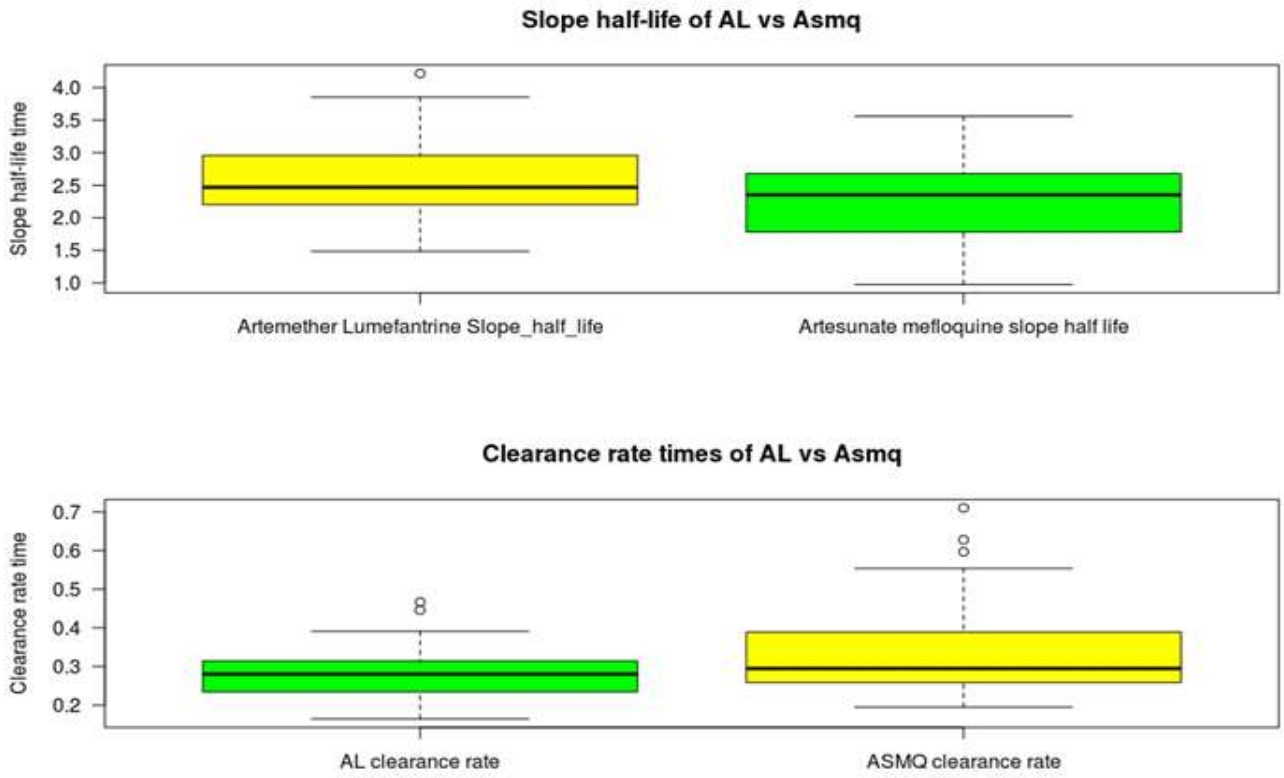


Figure 9. Parasite clearance slope half-life and clearance rate estimation between 2013 and 2015

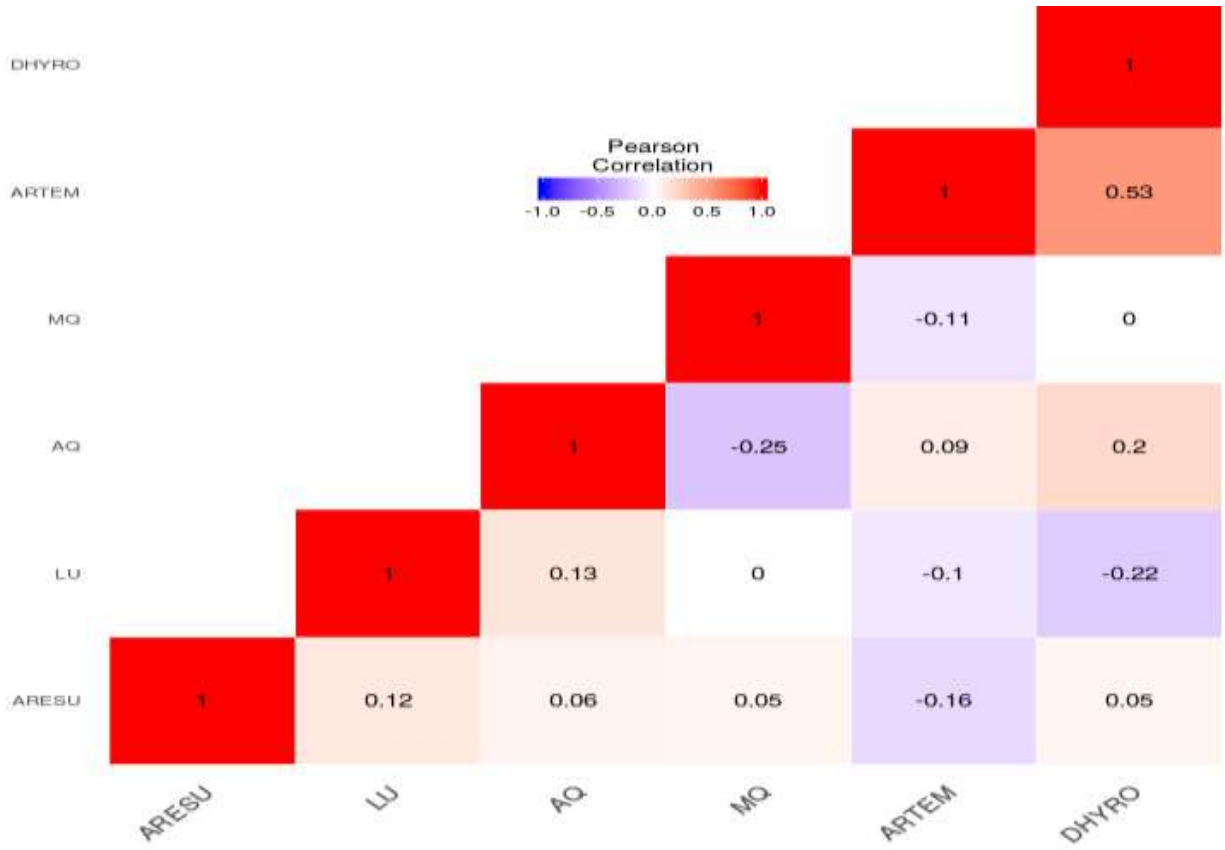


Figure 10. Correlation between different ACT drugs, mefloquine and amodiaquine have a significant negative correlation. Artemether and Artesunate, Artemether and Lumefantrine, Artemether and mefloquine, Dihydroartemisinin and Lumefantrine did not show significant negative correlation (Pearson correlation test).

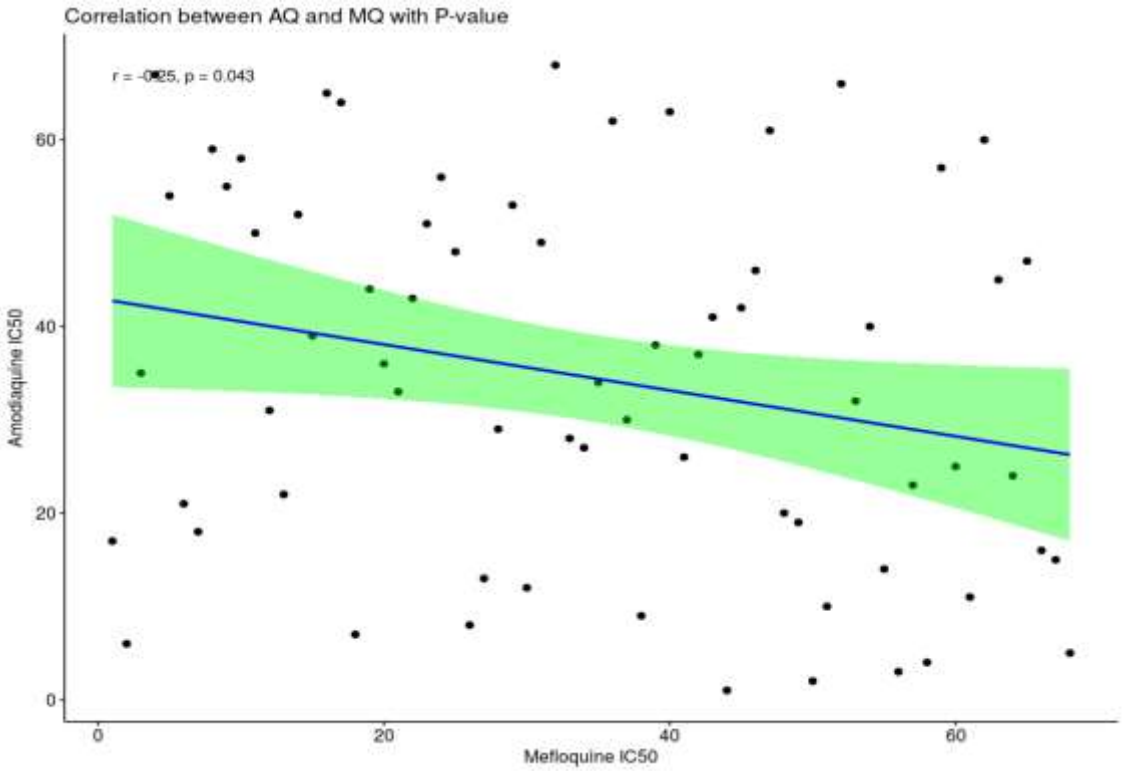


Figure 11. Linear regression plot showing Spearman correlation of Amodiaquine and Mefloquine. The green shade shows the 95 % confidence interval region.

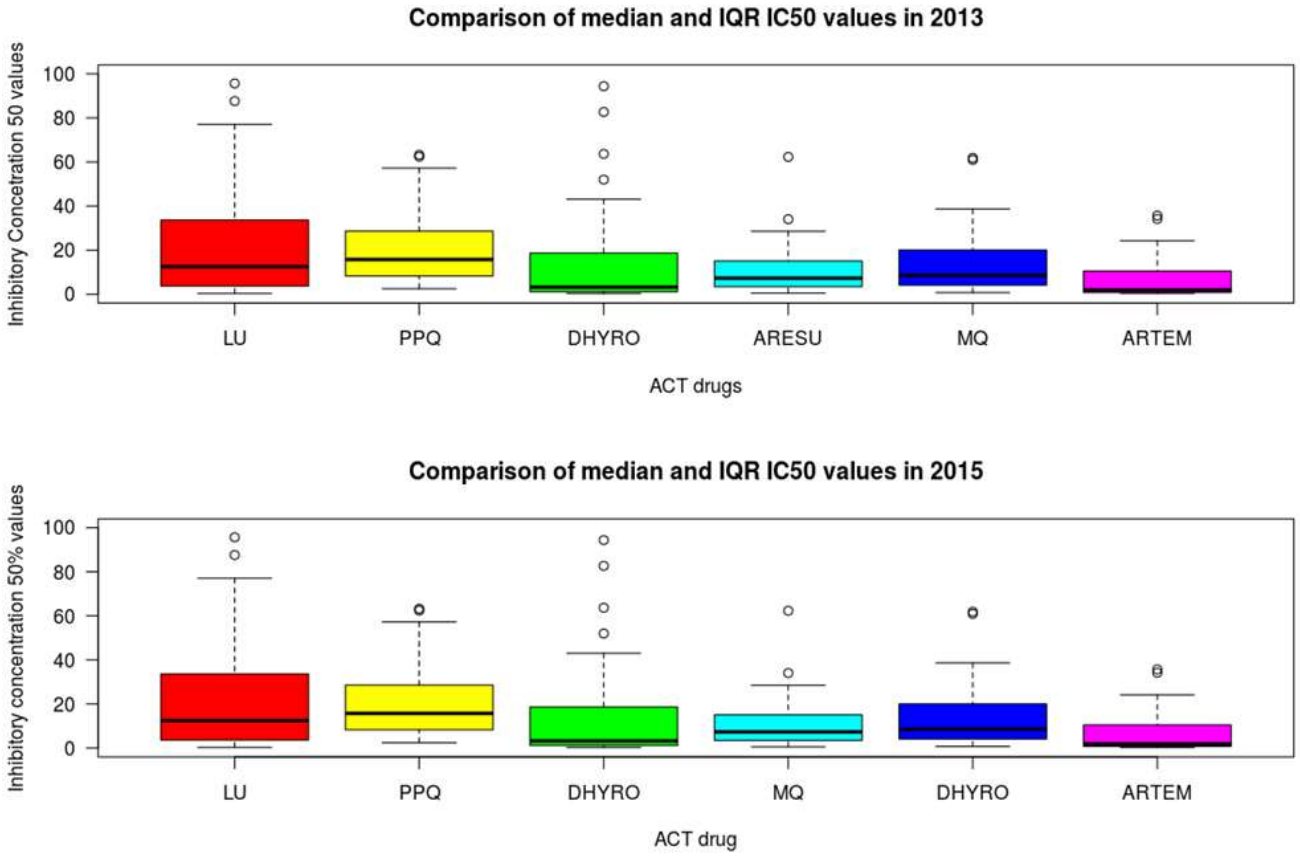


Figure 12. Median inhibitory concentrations 50% for different drugs in 2013 and 2015

3.1.2.4 Within-host population differentiation as a measure of evolution

The parasite population on day zero hour zero (D0H0) and day 1 hour 24 (D1H24) were significantly different from parasites on day 35, P-value 0.0036. The Hedrick (2005) and G_{ST} genomic summary statistics test, which is an analogue of $F_{ST,}$ and Jost (2008) were used to confirm the accuracy of F_{ST} values (see Table 3). The global population differentiation (F_{ST}) over all the loci was 0.0547. Global F_{ST} shows minimal population differentiation within the parasite population at different time points.

Table 3 *Plasmodium falciparum* differentiation at different time points. Parasite population on day thirty-five show high population subdivision (Microsatellite Analyzer)

Fst	pop1	pop2	pop3	pop4	pop5	pop6	pop7	pop8	pop9
pop1	0	0.01275	0.030586	0.043308	0.041928	0.059849	0.038113	0.352825	0.037419
pop2	0.01275	0	0.011893	0.033242	0.036035	0.07678	0.053072	0.375401	0.028065
pop3	0.030586	0.011893	0	0.035672	0.028459	0.081685	0.079302	0.412894	0.029267
pop4	0.043308	0.033242	0.035672	0	0.084945	0.156652	0.082283	0.508292	0.04901
pop5	0.041928	0.036035	0.028459	0.084945	0	0.127946	0.109577	0.548673	0.080283
pop6	0.059849	0.07678	0.081685	0.156652	0.127946	0	0.089128	0.676471	0.147101
pop7	0.038113	0.053072	0.079302	0.082283	0.109577	0.089128	0	0.415424	0.103548
pop8	**0.352825***	**0.375401***	0.412894	0.508292	0.548673	0.676471	0.415424	0	0.50158
pop9	0.037419	0.028065	0.029267	0.04901	0.080283	0.147101	0.103548	0.50158	0

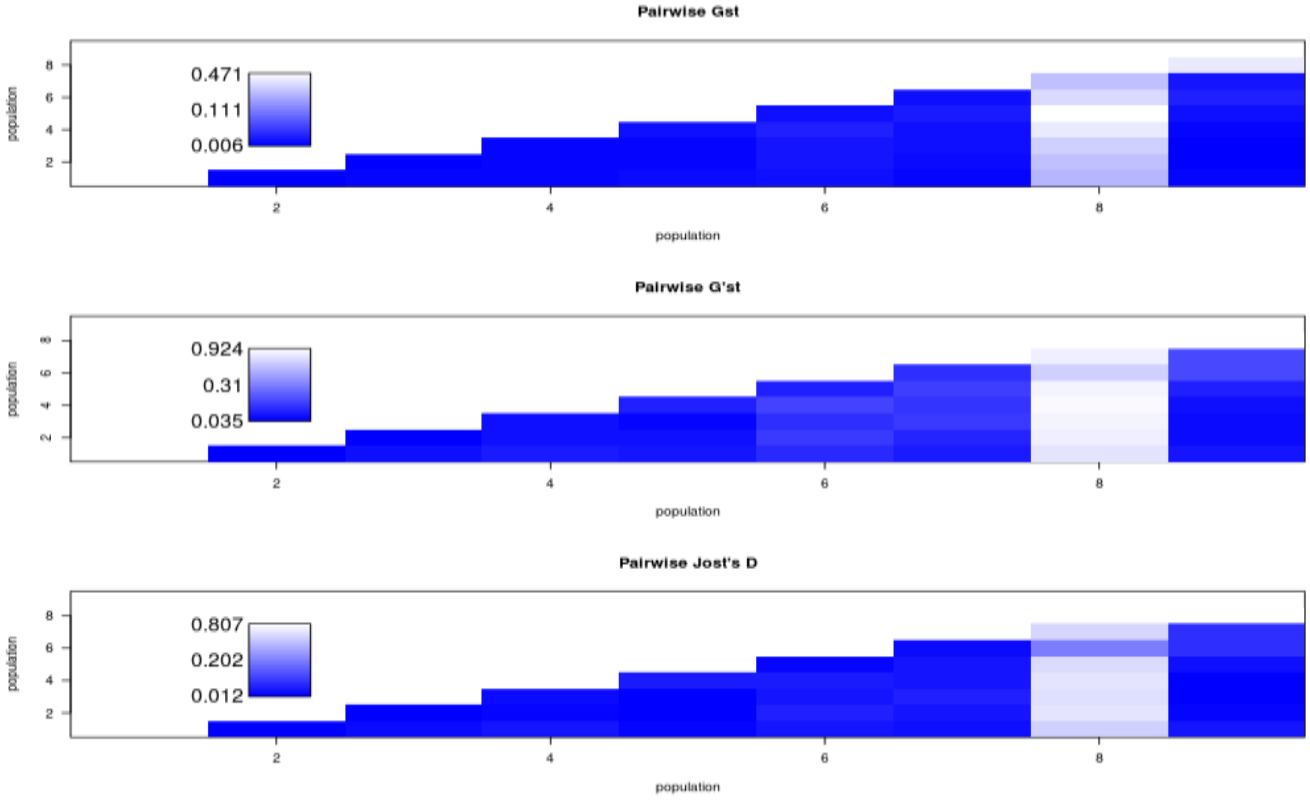


Figure 13. Pairwise Gst, G'st and Jost's D population differentiation estimates.

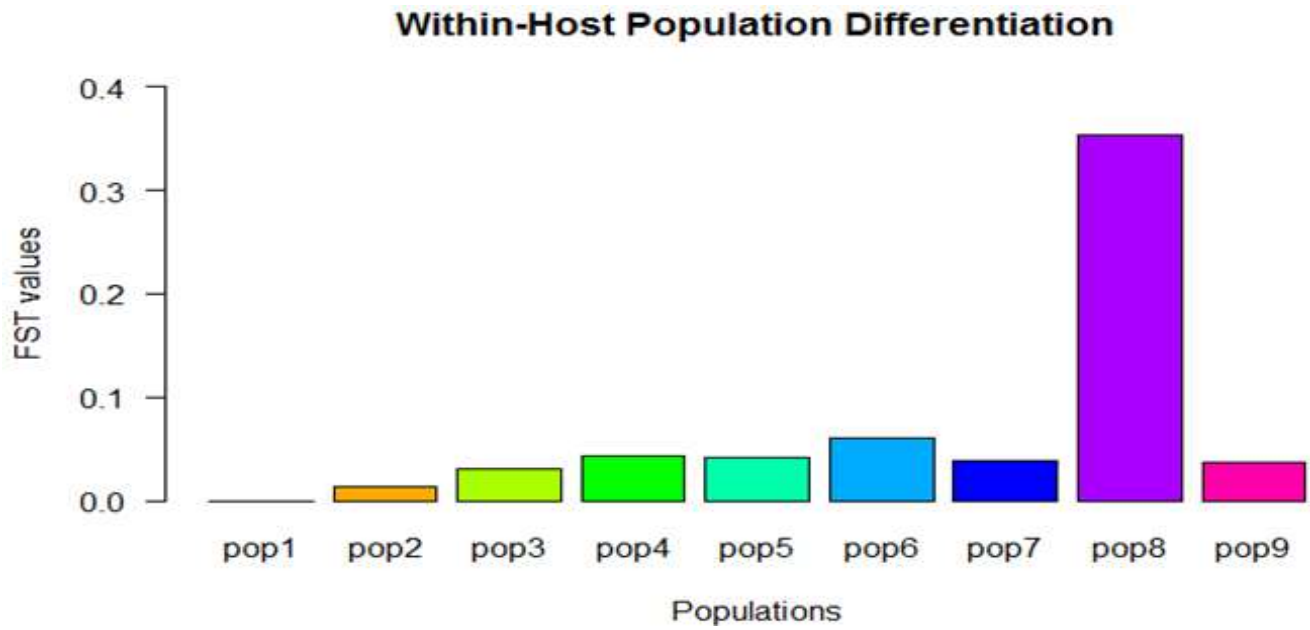


Figure 14. Within-host population differentiation FST values. Pop1 (H0D0), pop2 (D1H24), pop3 (D2H30), pop4 (D2H36), pop5 (D2H42), pop6 (D2H48), pop7 (D28), pop8 (D35), pop9 (D42).

3.1.2.3.0 Within-host allele frequency calculation

A total of one hundred and twenty-five *P. falciparum* parasites were used in the analysis. Parasites were grouped according to the treatment time points from day zero, hour zero to day forty-two, giving a total of nine populations (pop1-pop9). A total of 12 microsatellites were analyzed (microsatellite chromosome location in parenthesis) and they include the following: 2490 (10), ARA2 (11), PFG377 and Pfk2 (12), poly alpha (4), TA 81 (5), TA1 and TA 87 (6), TA60 (13), TA40 (10), TA108 (6), TA42(5). The range of alleles per locus was between 6 and 28. Locus poly alpha had the highest Simpson diversity (0.91) and ARA2 had the most evenly distributed alleles (0.86). TA40 had the lowest expected heterozygosity (0.42) while poly alpha locus had the highest expected heterozygosity (0.91) as shown in Table 4.

Table 4: Allele frequency distribution across 12 microsatellite loci.

Locus	Allele	Genetic Diversity	He	Evenness
2490	10.00	0.66	0.67	0.55
ARA2	6.00	0.73	0.73	0.86
PFG377	8.00	0.65	0.65	0.67
PFpk2	10.00	0.82	0.82	0.73
Poly alpha	28.00	0.91	0.91	0.66
TA109	14.00	0.87	0.88	0.83
TA40	16.00	0.43	0.43	0.34
TA42	12.00	0.52	0.52	0.45
TA60	13.00	0.86	0.86	0.81
TA87	14.00	0.58	0.59	0.43
TA1	18.00	0.90	0.90	0.83
TA81	13.58	0.73	0.73	0.65

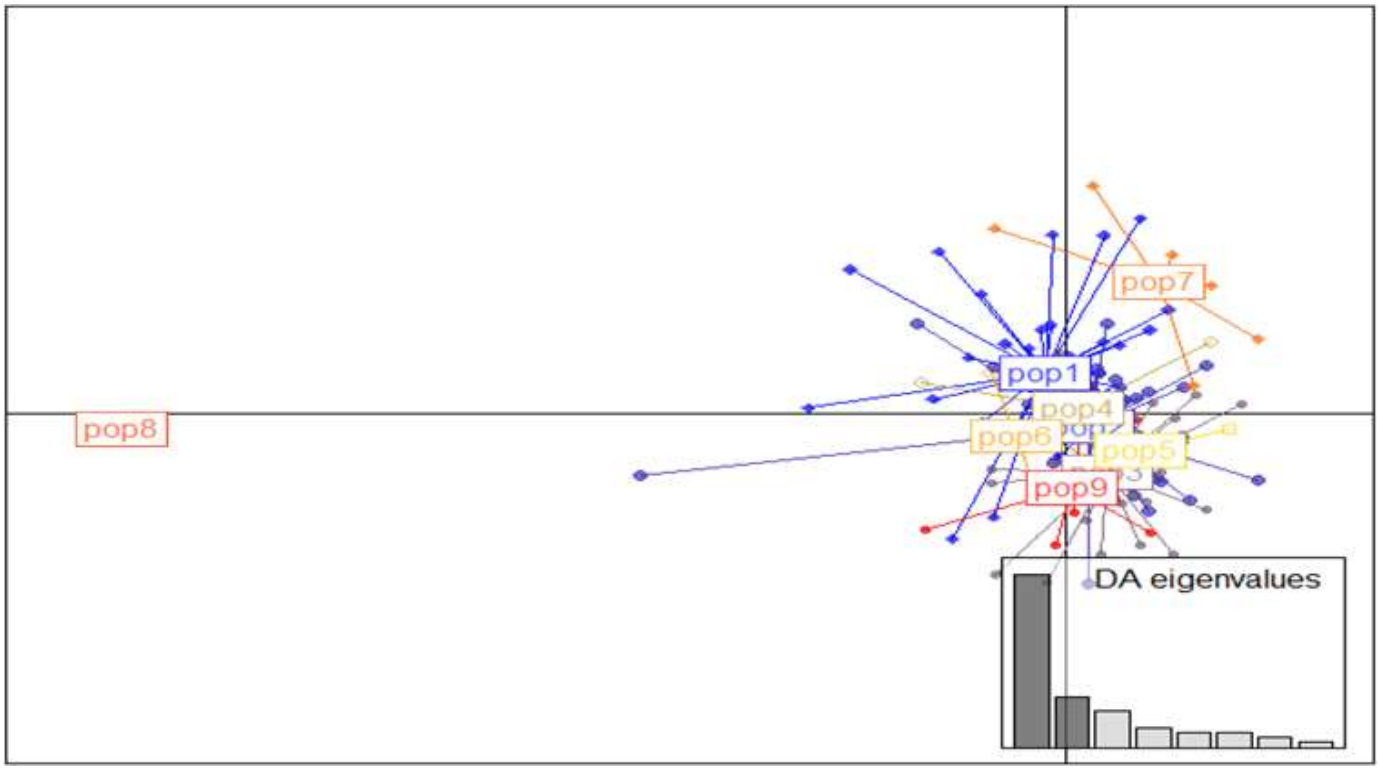


Figure 15. Discriminant Analysis of *Plasmodium falciparum* parasites at different time points

We then did the discriminant analysis of the principle components (Jombart et al. 2010), to visualize how different parasite populations cluster in space during the treatment course as shown in (Figure 15, Figure 18 and Figure 19). It was evident that the first two discriminant analysis eigenvalues explained the data better and they are labelled in black (Figure 15). From the results (see Figure 15) it is evident that the parasite population in day 35 (pop8) is more distinct genetically from other parasite populations because they cluster separately. It is imperative to look at the alleles that most differentiate these parasite population from the other parasites. From Figure 16, it can be seen that alleles at the microsatellite loci poly alpha (chromosome 4) and TA81(chromosome 5) are involved in differentiating the parasite populations at day 35 (pop8).In Figure 16, we determined the microsatellite loci which were highly polymorphic and how they contribute to the parasite clustering in space. We found that, Poly alpha and TA81 loci are highly polymorphic. The allele richness analysis (Figure 20 and Figure 21) confirmed that the *P. falciparum* from Kombewa and Maseno are still sensitive to ACT drugs. This is due to a rapid decrease in the parasite population from day one to day forty-two during the treatment period (see Figure 20 and Figure 21). There were variations in SNPs over the twelve microsatellite loci during the treatment course and this did not follow a clear pattern (Figure 17). This could be due to differences in the immune response during treatment. It was evident from our analysis that the clearance slope half-life does not correlate with the evolutionary rate of the parasite (see Figure 22) and that it is a random process.

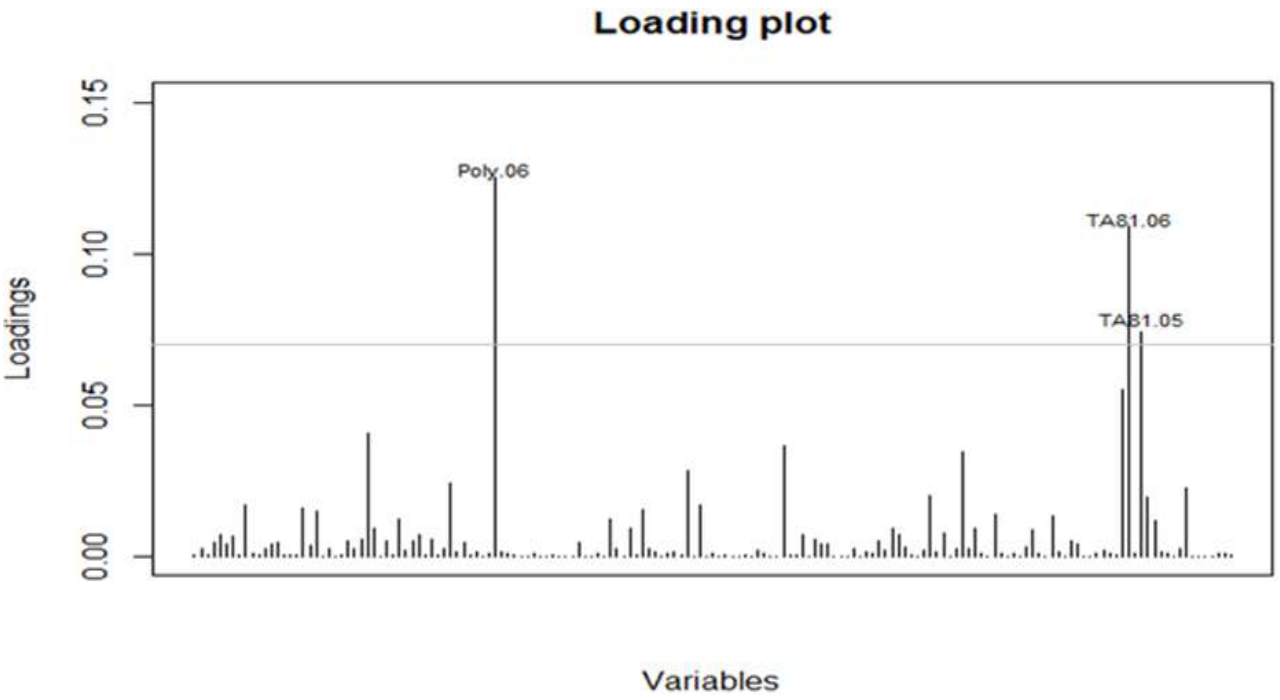


Figure 16. Loading plot showing why parasites on day thirty-five (pop8) are distinct from the other parasite populations within the host.

The evolutionary effect of the two loci, Poly alpha and TA81 across the entire *Plasmodium falciparum* population during treatment was determined (Figure 17).

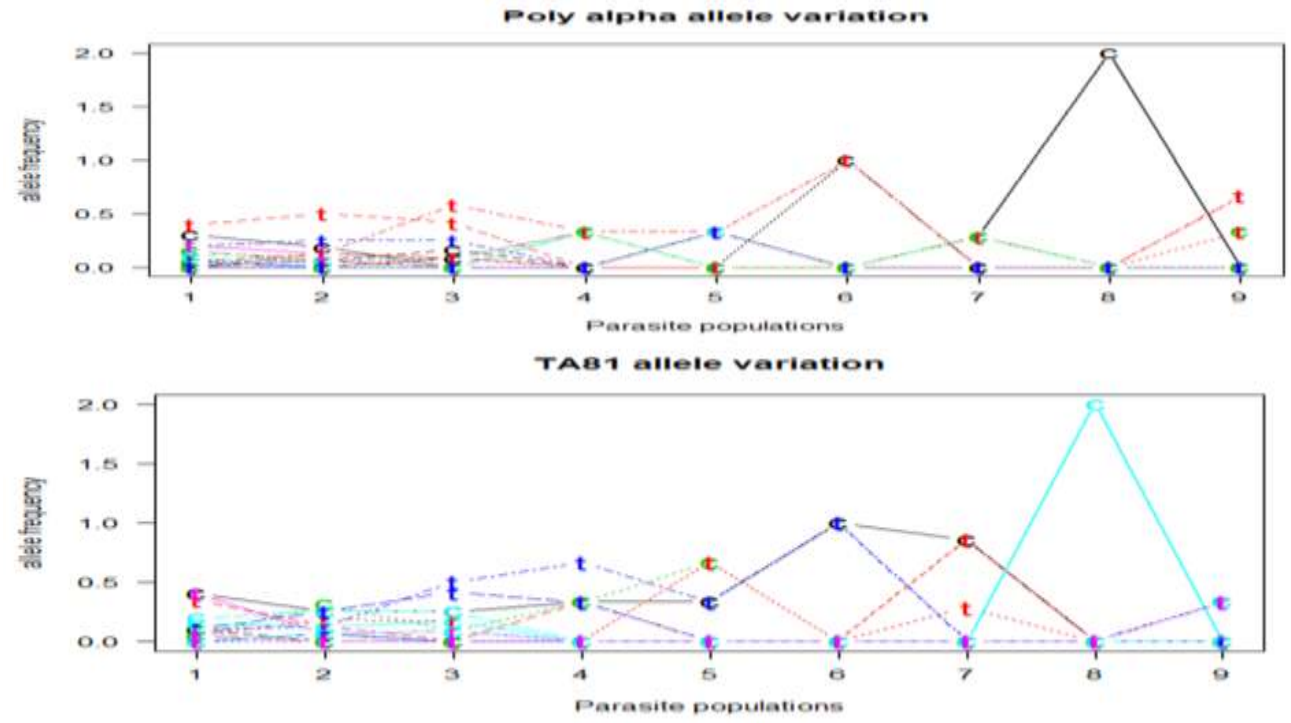


Figure 17. Evolutionary effect of poly alpha and TA81 loci. Allele variation over the two loci, Poly alpha and TA81 from day zero to day forty-two. The parasite populations at different time points were treated as distinct parasite populations and the allele variations does not follow a clear pattern across the populations.

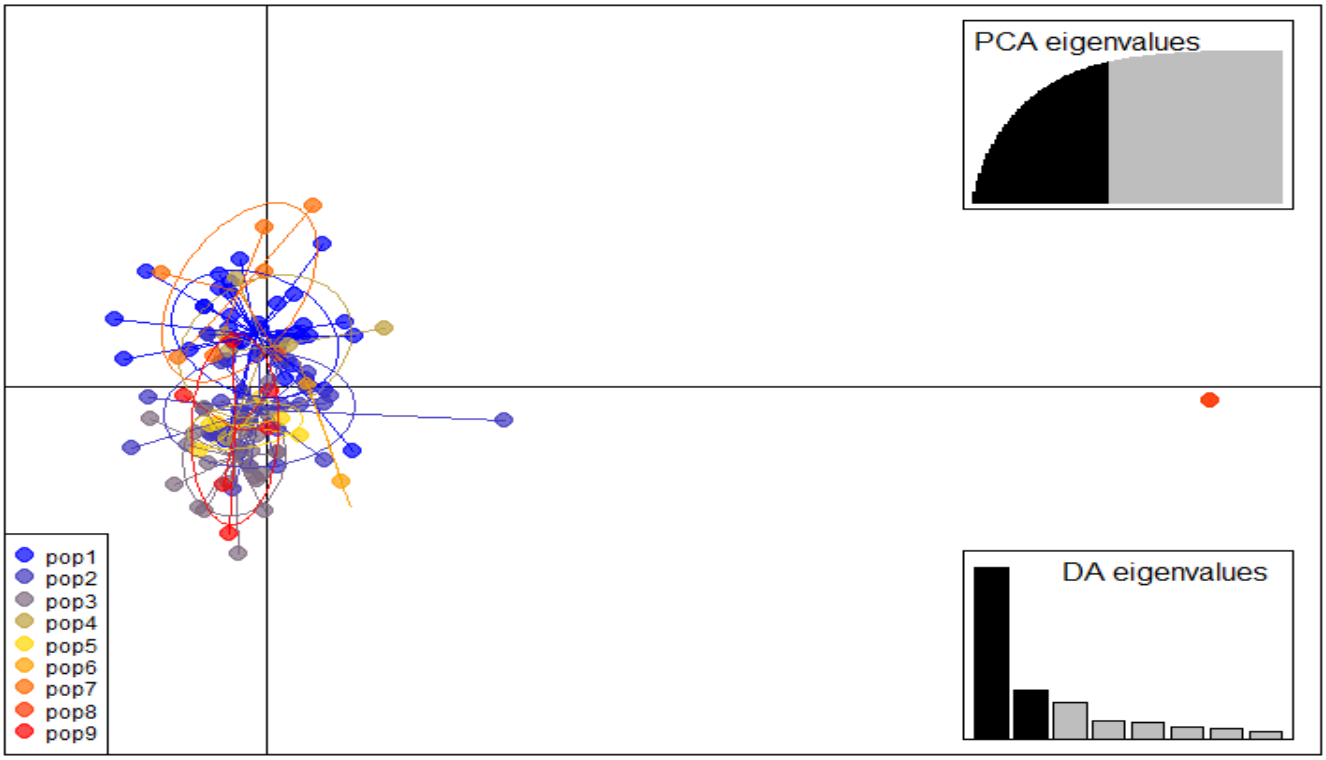


Figure 18. Discriminant Analysis of Principle Components (DAPC) plot of within-host parasite populations at different time points.

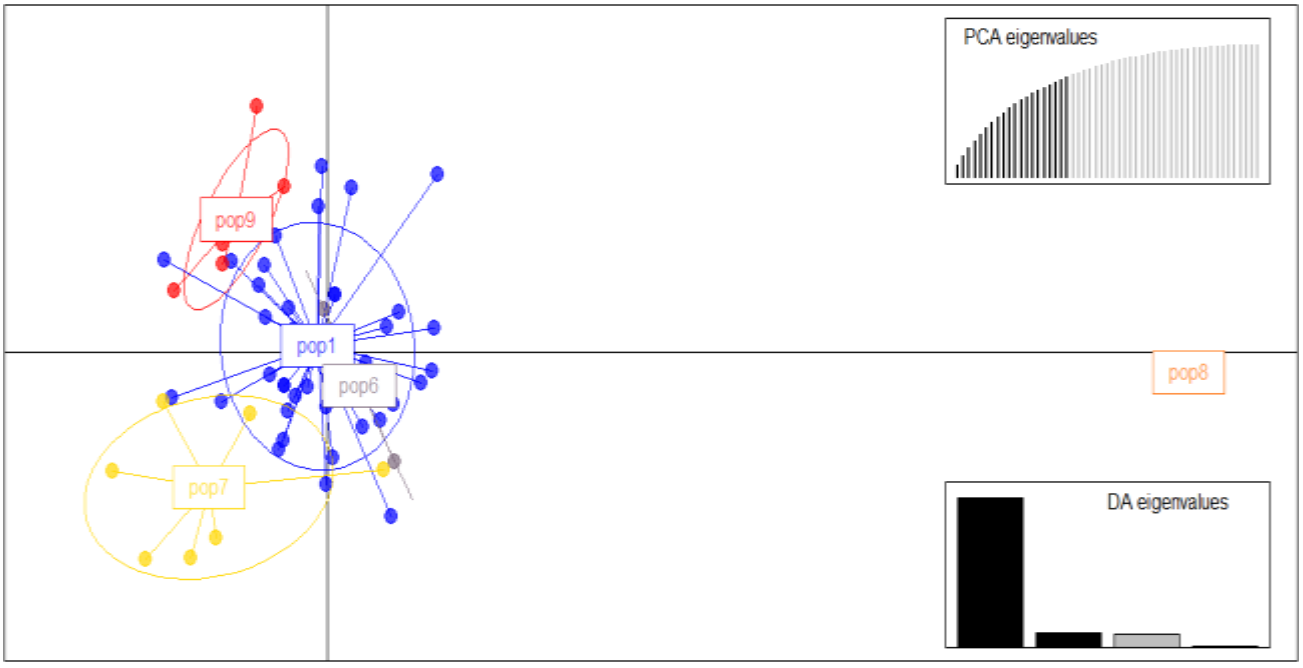


Figure 19. Discriminant Analysis Principle Components analysis. This analysis shows distinct clustering of pop1, pop6, pop9 with pop8 still clustering differently with other parasite populations in space.

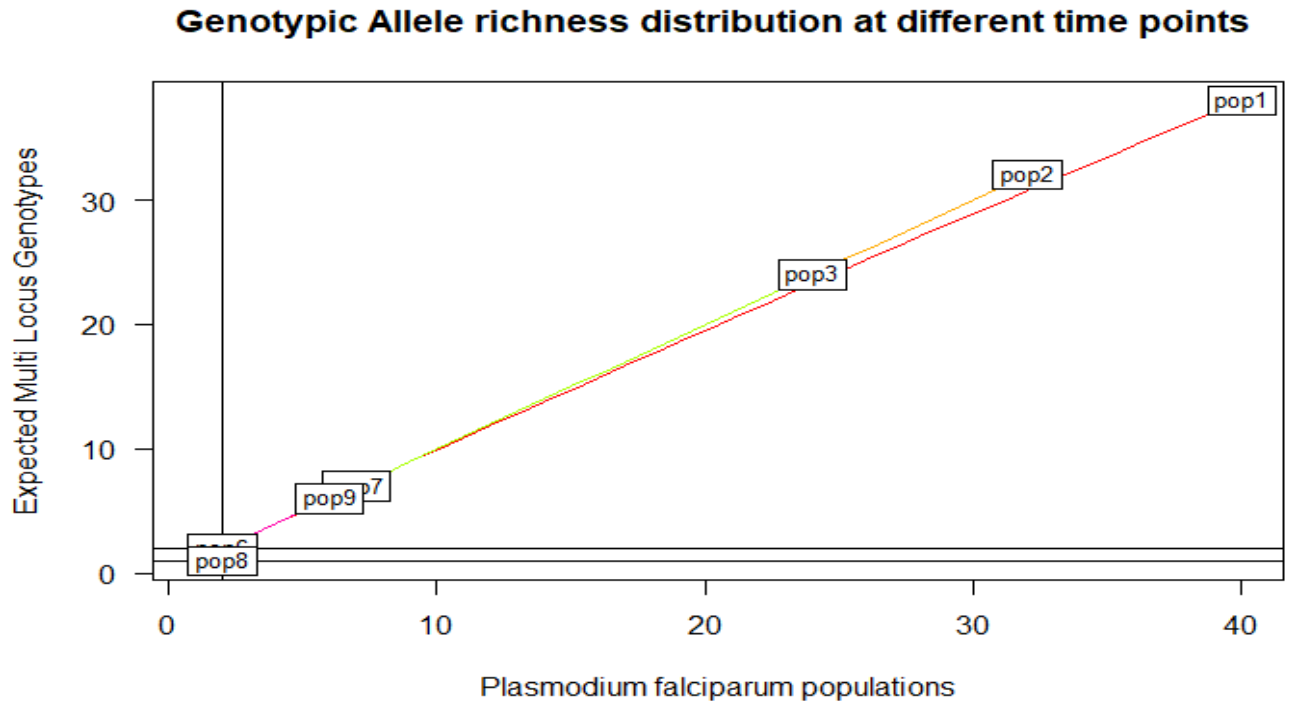


Figure 20. *Plasmodium falciparum* allele richness distribution over the treatment course from pop1 to pop9.

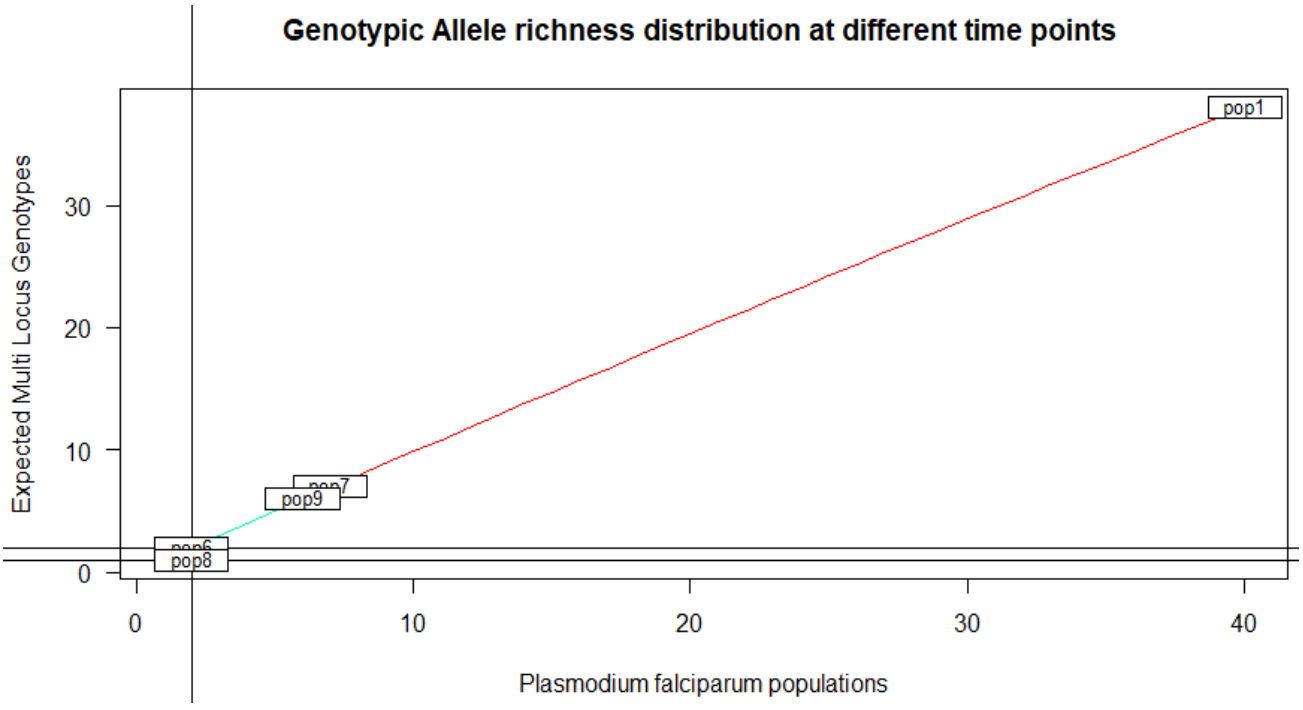


Figure 21. *Plasmodium falciparum* allele richness variation over time. Parasites on day zero hour zero have the highest allele richness, meaning they have high evolvability potential compared with the other parasite populations during treatment.

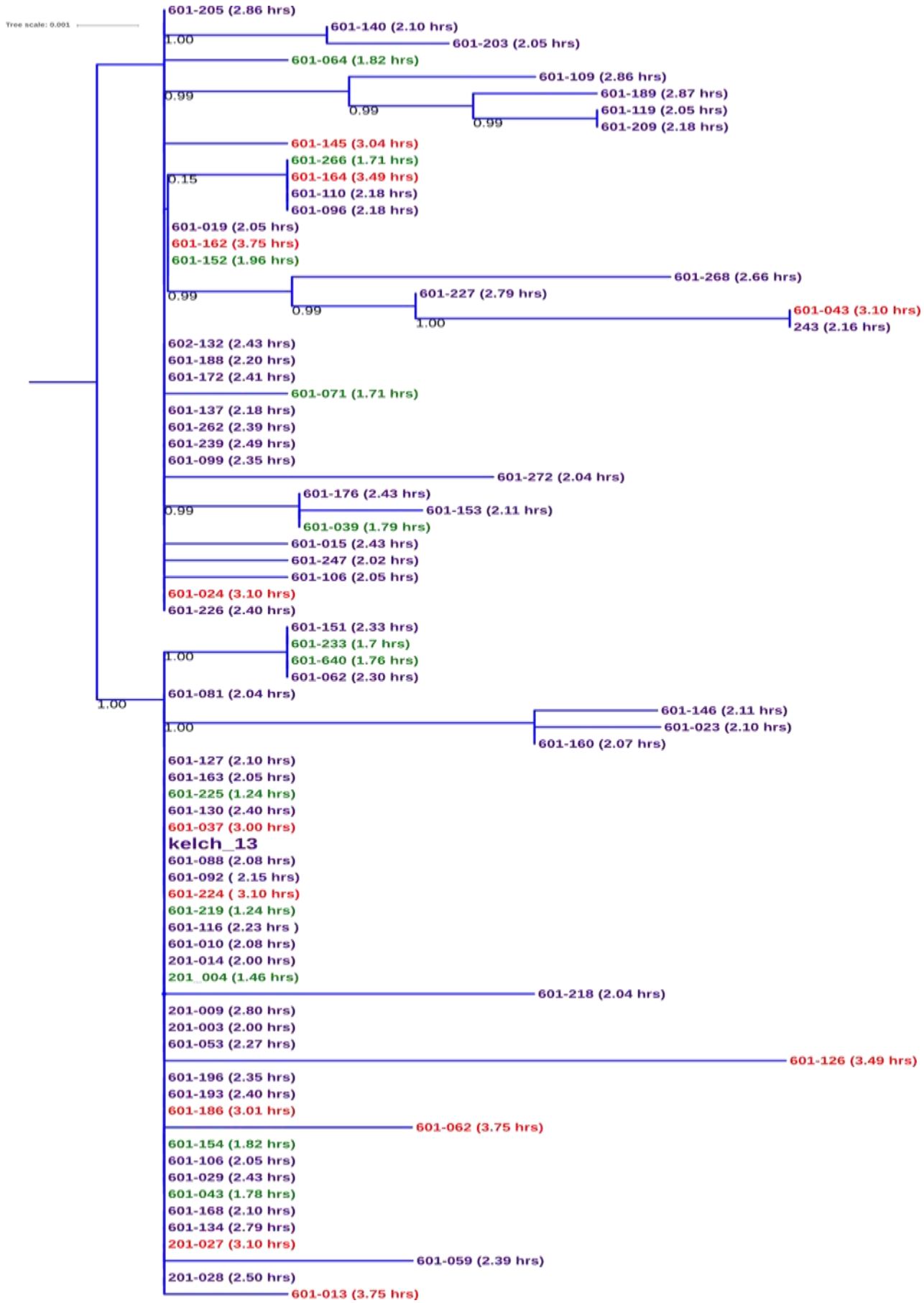


Figure 22. Kelch 13 maximum likelihood phylogenetic tree, the parasite evolution clustering is stochastic since the clearance slope half-life of parasites during treatment does not show a distinct clustering.

Phylogenetic analysis (see Figure 22) was done and the main motivation was to understand the evolutionary relationship of the parasite population within the human host.

Chapter 4. Discussion

4.1 Neutral models of evolution

Adaptive evolution to artemisinin combination therapy regimen can be studied using genotypic and Ring Stage Survival Assay ($RSA_{0\text{-}3hr}$) methodologies. ACT resistance is evident when there is an increase in parasite clearance slope half-life which is associated with the reduced susceptibility to ACT drugs *in vitro* (Ariey *et al*. 2014), resulting in ACT treatment failure. In Kombewa and Maseno, which are high transmission settings in Kisumu County, Kenya, there is a large usage of artemisinin drugs throughout the year. A study by Watsierah *et al* (Watsierah *et al*. 2011) in Kisumu peri-urban settings found that only 29.4 % of the study population used the right ACT drugs to treat malaria infection (Watsierah *et al*. 2011). Patient behavior such as non-adherence to ACT drug and individual variations may provide room for the parasites to evolve and become non-responsive to treatment. When poor-quality drugs are prescribed to the patients, it results in low bioavailability, making the drugs unable to exert their therapeutic effects, hence promoting the development of drug resistance (Maponga and Ondari 2003). A study carried out by Akulayi *et al* in Africa found that 20% of non-quality assured Artemisinin Combination Therapy drugs (non-QAACT) are in circulation in the Kenyan markets, contrary to the World Health Organization quality assurance recommendations (Akulayi *et al*. 2017; Maponga and Ondari 2003). Musuva *et al* reported that only 58.2% of the drugs in circulation satisfy WHO quality assurance standards (Musuva *et al*. 2017). These statistics are of great concern noting that the use of substandard drugs in South East Asian countries is the main driver of drug resistance development. Kenyan *P. falciparum* populations may also evolve and become non-responsive to these counterfeit ACT drugs.

Kenya adopted ACT in treatment of non-severe malaria in 2006 and it has remained efficacious since its adoption (Amin *et al*. 2007). The World Health Organization report indicates an increase usage of ACT in Sub-Saharan Africa and tropical regions of the world between 2006 and 2015 (World Health Organization 2016, 2015c). As a malaria holoendemic setting in Kenya (Bejon *et al*. 2010, 2014), an increase usage of ACT drugs has a possibility of raising drug pressure and thus the risk of adaptive evolution of *P. falciparum* parasites to these drugs (Ocan *et al*. 2016). Therefore, the neutrality model of evolution was used to determine the departure of Kelch 13 nucleotide sequences from neutral evolution using Tajima's D, Fu & Li F*, and Fu & Li D* genomic summary statistics tests (Tajima 1989; Fu and Li 1993). In 2013, the *P. falciparum Kelch 13* gene was found to be under balancing selection (see Figure 4) as shown by significant positive values of Tajima's D. These 2013 finding were consistent with results from Oyebola et *al study* conducted in Nigeria in 2013, which showed

that *P. falciparum* isolates were under balancing selection. The findings are an indication that African parasites might be behaving the same in the presence of ACT drugs (Oyebola *et al*. 2017). However, there was also an evidence of selective sweeps (positive selection) as shown by significant negative test values observed from 2015 Kelch 13 nucleotide sequences (Figure 4).

Our findings using the 2015 Kelch 13 sequence data is consistent with results of a previous studies conducted in Uganda by Ocan *et al* (Ocan *et al*. 2016) and in the Gambia by Amambua- Ngwa *et al* (Amambua-Ngwa *et al*. 2012), in which *P. falciparum* was shown to be under selective sweep (Ocan *et al*. 2016; Amambua-Ngwa *et al*. 2012). This is an indication that the African *P. falciparum* parasites are gaining fitness potential to be able to survive under ACT treatment nine years after ACTs were adopted as the first line treatment of uncomplicated malaria in most African countries, including Kenya (Ministry of health 2012).

The prevalence of non-synonymous single nucleotide mutations which are beneficial to *P. falciparum* were on the rise during our study and this may be attributed to the effects of an increase drug pressure. The lack of adherence and an improper usage of antimalarial drugs are thought to be the main causes of the adaptive evolution to ACT drugs in malaria parasite population (Watsierah *et al*. 2011). For decades, *P. falciparum* proteins have been shown to be enriched in non-globular domains whose functions remain unknown. These protein domains, including the Kelch 13 propeller domain, have a high potential for undergoing genetic drift and thus provide a big challenge in the attempts to determine the molecular dynamics of adaptive evolution in *P. falciparum* parasites (Gardner *et al*. 2011). Our finding using nucleotide sequences was consistent with the amino acid sequence analysis of Li *et al*. conducted in Lira and Gulu, in Northern Uganda (Li 1993), which showed that in 2013 *P. falciparum* was under balancing selection. The codon-by-codon evolutionary pattern along the *Kelch 13* (*pfk13)* gene changed exponentially between 2013 and 2015, which increased the evolutionary potential of these parasites to resist ACT drugs. This research could not explicitly explain whether mutations in the *PF3D7_1343700* gene in *P. falciparum* from Kisumu were due to transfer through global human travel or local emergence due to parasite population level processes.

A total of seventeen non-synonymous single nucleotide polymorphisms were detected in *P. falciparum* parasites from Kisumu (Table 2). Some of these seventeen SNPs in the Kelch 13 propeller domain may be associated with drug resistance development as was demonstrated by Njoka *et al* (Njokah *et al*. 2016). The two predominant non-synonymous SNPs in our data were D559E which was present in 42/ (61.76%) of the samples and D547E which existed in 37/ (54.41%) of the samples. These non-synonymous mutations are specific to Kenyan parasites and have not been reported in any other African country or South-East Asia countries (Ocan *et al*. 2016; Ariey *et al*. 2014). The Kelch 13 propeller domain is highly conserved in *Plasmodium* and even though its exact function has not

been determined, it is thought to be involved in protein-protein interactions. The World Health Organization officially included the following non-synonymous single nucleotide mutations: Y493H, I543T, R539T, C580Y and M476I in the *PF3D7_1343700* gene in tracking ACT drug resistance in August 2014 (Njokah *et al*. 2016). In our research, the mutations reported elsewhere were not detected and there was no signal of reduced susceptibility to ACT in the Kenyan *P. falciparum* populations. Artemisinin resistance due to ACT drug pressure has not been reported in Africa, but SNPs which are not associated with treatment failure are present as observed in several studies (Daniels *et al*. 2008; Sisowath *et al*. 2005). Since Myanmar is centrally located between Africa and South East-Asia, and resistance to ACTs has been reported in this region, it is feared that these resistant parasites may gain entry to Africa following the chloroquine resistance path of the 1960s (Roper *et al*. 2004).

Periodic tracking of these genetic markers of ACT drug resistance is of great importance in controlling the spread of drug resistance to Sub-Saharan Africa. *P. falciparum* parasites in 2013 showed low level of variability with the haplotype diversity (Hd) and nucleotide diversity (π). In 2015, *P. falciparum* parasites showed a higher level of diversity and the nucleotide diversity. The high variability between the two-time points was consistent with the findings in a study conducted in Malawi by Ocholla *et al* 2014, where high genetic diversity of the parasite population was detected during treatment, a characteristic associated with the African parasites, especially in the regions of high malaria transmission (Ocholla *et al*. 2014). A study conducted by Feng *et al* (Feng *et al*., 2015) found SNPs at codons 537 and 574 in the Kelch 13 gene in *P. falciparum* strains which infected Ghanaian migrants to the Republic of China. These mutations have also been reported in Cambodian *P. falciparum* parasite strains (Ocan *et al*. 2016). The findings in this research, in addition to Feng *et al's* results in Ghana, and other Sub-Saharan countries (Feng *et al*. 2015), is an indication that ACT drug resistance is likely to be reported among *P. falciparum* parasites from Africa soon.

Recent studies carried out in Kenya and Uganda (Achieng *et al*. 2015; Cooper *et al*. 2015), in addition to other regions of Sub-Saharan Africa, specifically in the Republic of Mali, West Africa (Ouattara *et al*. 2015), and Myanmar (Tun *et al*. 2015), found non-synonymous point mutations in *P. falciparum* which have not been reported in Cambodia (Ariey *et al*. 2014). Geographic differences in Kelch 13 propeller gene point mutations indicate the presence of a global reservoir for the Kelch 13 propeller gene (Ocan *et al*. 2016). However, whether these SNPs are associated with the increased parasite clearance slope half-life needs to be investigated further, as it will help in understanding ACT resistance in high transmission areas outside of Cambodia.

4.2 IC_{50}, Slope-half life and SNP association analysis

In this research, we determined that in 2013 there were no non-synonymous point mutations after mapping and visualizing the mapped reads to the reference Kelch 13 reference gene sequence using Artemis (Carver *et al*. 2012). Due to a lack of non-synonymous point mutations from the 2013 SNPArray data, we did not carry out the association analysis of parasite clearance slope half-life, and IC_{50s} with the SNPs. There was no significant association between SNPs and parasite clearance slope half-life which was consistent with the report by Ngala *et al* in 2013 (Ngalah *et al*. 2015) in samples from Kisumu County. Seventeen non-Synonymous SNPs were present in the *P. falciparum* genome in 2015 samples. There was a significant association between IC_{50} and the SNP at codon D559E with Amodiaquine (AQ) and Mefloquine (MQ). Codon D547E also had a significant association with the Artesunate half maximal concentration.

These data reveal a sharp increase in non-synonymous point mutations between 2013 and 2015 in Kenyan *P. falciparum* parasites. Previous studies using *P. falciparum* parasites from Africa and South-East Asia have found a 35 kilobase locus on chromosome 13 to be associated with the reduced parasite clearance rate, an indication that it plays a role in ACT drug resistance (Cheeseman *et al*. 2012; Ngalah *et al*. 2015; Amato *et al*. 2017). Lack of non-synonymous point mutations from the Kenyan *P. falciparum* parasites may be explained in part to the reduction of pressure following the withdrawal of chloroquine and Sulphadoxine-pyrimethamine as the first line treatment of uncomplicated malaria in 2006 (Watsierah *et al*. 2011). Candidate genes on chromosome 13 such as the Kelch 13 gene, lipoate synthase, and hsp70 and many more conserved genes whose functions are not well understood have been implicated in the development of drug resistance in *P. falciparum* (Ouattara *et al*. 2015). Resistant parasites from SEA have a high frequency of mutations in *pfk13* and these mutations have shown significant association with the delayed parasite clearance during treatment (Cheeseman *et al*. 2012).

The two SNPs, D557E and D559E, as Y493H, I543T, R539T, C580Y and M476I which were reported by Ariey *et al* (Ariey *et al*. 2014), may be considered as the potential treatment failure markers in Kenyan parasites. Since these two SNPs, D557E and D559E occur at a higher frequency in our samples, we alluded that they may be the potential markers for drug resistance in the Kenyan *P. falciparum*. Further research needs to be done to validate this finding. It is interesting that a significant association in *in vitro* drug susceptibility and SNPs was detected in the Kenyan *P. falciparum* parasite in our study, and it may be a wakeup call for surveillance of ACT drug resistance, since such molecular findings have been used previously to determine drug administration policy in Tanzania (Mugittu *et al*. 2004).

The results in (Figure 12) from two different time points show that the Kenyan parasites are still

sensitive to ACT drugs even though the non-synonymous point mutations are on the rise due to adaptive evolution. The same observations have been reported in Mali by Ouattara *et al* (Ouattara *et al*. 2015), in Uganda by Cooper *et al* (Cooper *et al*. 2015) and in Kenya by Ngala *et al* (Ngalah *et al*. 2015).

Drugs with correlated modes of therapeutic effects indicates related mode of action. Most importantly, drugs with significant negative correlations may reveal synergistic partner drugs which can be deployed in population on a rotational basis (Wendler *et al*. 2014). There was a significant negative correlation between Mefloquine (MQ) and Amodiaquine (AQ) half maximal inhibitory 50% concentration. Artesunate and Artemether modes of action were negatively correlated without reaching while there was no significant correlation between Lumefantrine and Dihydroartemisinin. Interestingly, this was consistent with a study done in Kenya by Wendler *et al* 2014, in which they found that piperaquine was significantly correlated with the amodiaquine. The mean parasite clearance slope half-life was less than three hours compared with Cheeseman *et al* (Cheeseman *et al*. 2012) report, where the parasite clearance slope half-life was more than five hours. These findings are expected in part because Kenyan parasites are still sensitive to ACT drugs. The finding also mirrors the Ngalah *et al* report which showed that *P. falciparum* from Western Kenya are still sensitive to ACT.

4.3 Within-host population differentiation as a measure of evolution

The change in malaria treatment and control policies in Kenya in 2006 (Watsierah *et al*. 2011), where artemisinin derivatives were adopted for the treatment of uncomplicated malaria, was a major achievement in the fight against malaria in Kenya. Interestingly, when these drugs were being introduced in Kenya, resistance to these drugs was already evident in Cambodia (Noedl *et al*. 2008). ACT drugs impose a strong selective pressure on the *P. falciparum* genome, altering parasite genomic structure. This analysis was aimed at evaluating the changes in *P. falciparum* parasite population genetic diversity and differentiation at different time points during ACT treatment. The analysis further looked at the temporal changes in *P. falciparum* parasites from day zero to forty-two within the human host. There was moderate parasite population structure differentiation (Figure 13) between the parasite populations within host, except for day thirty- five parasites, which were significantly different from day one parasite populations (Table 3).

The lack of significant genetic differentiation is consistent with findings (Ouattara *et al*. 2015) on African parasites which are still sensitive to ACT drugs. These drugs still neutralized *P. falciparum* within the human host as expected, since no ACT resistance has been reported from Kisumu *P. falciparum* parasites. The (unique parasite population from day 35 may be explained in part to superinfection. Patients on day thirty-five might have been infected by parasites with different genetic

background, and since we did not have the meta data on the migration history of these patients but hypothesis that these unique parasites having a different genetic background may have been introduced in this population due to human migration into the Kombewa. Druilhe and his colleagues (Druilhe *et al*. 1998) demonstrated that a single mosquito bite (Figure 2) may deposit a large amount of malaria parasites in the human blood stream which may have different genetic diversity. In addition, Zhong *et al* in their study using the same microsatellite data also found that the parasite population from Western Kenya had a significantly high genetic differentiation which was attributed to human migration (Zhong *et al*. 2007). This may explain the distinct parasite difference on day thirty-five, although it does not explain why this dropped again after day 35.

Selective pressure due to ACT usage results in unique genetic signatures of *P. falciparum* population structure within the host. The expected heterozygosity in our data was much higher compared with data from Columbia, Brazil and Bolivia with expected heterozygosity ranging between 0.30-0.40. There was high genetic diversity of *P. falciparum* during the treatment period, which was attributed to high malaria transmission in Kombewa and Maseno. Within host genetic differentiation was moderate, an indication of gene flux in this parasite population. This observation is consistent with Nabet *et al's* finding in Mali, where the parasite populations displayed high genetic diversity and moderate population differentiation during ACT treatment (Nabet *et al*. 2016). In Thailand, *P. falciparum* parasites had an expected heterozygosity value of 0.51 (Anderson *et al*. 2000), but our finding was consistent with the joint studies carried out in African Countries: Cameroon, Ghana, Uganda, Gambia and Benin, where expected heterozygosity was 0.69 (Durand *et al*. 2003). Genetic diversity in our *P. falciparum* data was much higher during the treatment with only locus TA40 harboring a genetic diversity parameter below 50% (Table 4). Zhong *et al* (Zhong *et al*. 2007) found that the poly alpha loci across the entire parasite population had the highest number of alleles in their study and our results also show the same pattern. In Congo and Zimbabwe, the studies show that the expected heterozygosity of *P. falciparum* higher (0.76-0.80) during ACT treatment, but the parasites remain sensitive to ACT drugs. These findings are comparable with our result (Table 4). Interestingly, Vardo-Zalik and colleagues (Vardo-Zalik *et al*. 2013) also observed high genetic diversity (He = 0.74) in *P. falciparum* from Kombewa after the ACT intervention and the scale up of insecticide treated bed nets (ITNs) usage in western Kenya. Our data also shows the same pattern with the high genetic diversity values seen across all the twelve microsatellite loci (Table 4). This may have a large implication as it may result in an increase in parasites which may soon become non-responsive to ACT drugs in our population. The two microsatellite loci: Poly alpha and TA81 (Figure 16) are the main molecular markers that contribute most to the differentiation of the *P falciparum* parasites and specifically on day 35 (pop8). Day thirty-five parasites were expected to be similar to day zero-hour zero parasite populations which was not observed in our data. Since we did not have these patients'

metadata, we hypothesized that these patients may have been infected with a parasite with a different genetic background. This is indicative of high polymorphism rate in those two microsatellite loci. The allele variation over these loci at different treatment time points from day zero, hour zero pop1 to pop_9 (day 42) is considerable and does not follow a clear pattern (Figure 17). A study conducted in Guinea by Murray and colleagues to identify the genetic diversity using *P. falciparum* clinical isolates found that the polymorphisms on these microsatellite loci are not uniform (Murray *et al*. 2016). These findings are comparable with our result where there was a stochastic variation of alleles over these two loci at different time points during treatment with the highest frequency reached in population 8 (Figure 17). Irrespective of the process underlying these diversities (selection or genetic drift), it is an illustration that in *P. falciparum*, specific non-synonymous point mutations may occur in the genome within a short period of time during treatment.

Hierarchical clustering of parasite populations is an important tool in the study of parasite evolution within the human host over time. As revealed by the Discriminant Analysis of Principle Component (Jombart et al. 2010), *P. falciparum* "populations" demonstrated distinct clustering from population 1 to population 9 (Figure 18 & Figure 19), the parasite populations were primarily clustered in nine and five groups, during the treatment course. In Figure 19, *P. falciparum* showed distinct clustering according to the treatment time points. The *P. falciparum* in our study have been shown to have a high parasite genetic diversity (see Table 4), this explains why some parasites did not undergo evolution and their genome contents did not change over the treatment course. This is evidenced by the overlapping of the points from one population to the other in the Discriminant Analysis of Principle component plots (Figure 19). Another plausible explanation for the overlapping of these parasite strains in space can be attributed to the moderate parasite population differentiation during treatment course owing to the fact that Kisumu is a high malaria transmission zone (Sifuna *et al*. 2014). *P. falciparum* from Kisumu are still sensitive to artemisinin derivate (Ngalah *et al*. 2015) drugs and this is shown by the rapid parasite population decline during ACT treatment. Analysis of the microsatellite data provides a prime basis on which we can discriminate parasite populations at different time points, but a lack of power for discriminating clonal from non-clonal infections and distinction of different parasite strains from each other is very difficult. The rapid neutralization and the sensitivity of *P. falciparum* to ACT drugs was evident due to a sharp decrease in the allele richness during the treatment (Figure 20 & Figure 21). Allele richness has a direct correlation with the population under study and the sharp decrease in the allele richness in this study confirmed that these drugs are still efficacious. A rapid decrease in the allele richness over time reduces the evolvability potential of these parasite strains to ACT drugs as they get neutralized within the expected time.

4.4 Evolutionary effects of mutations on Kelch 13 propeller domain

The Kelch 13 propeller domain is made up of six blades (Figure 6) consisting of protein-protein interaction sites. Mutations occurring in the β-sheet sites may hinder effective K13 protein-protein interactions with other protein molecules. Mutations in the Kelch like proteins have been linked with the development of hypertension in humans (Boyden *et al*. 2012). The *P. falciparum* PI3K is assumed to be the binding partner of Kelch 13 and it is the main protein molecule being targeted by ACT drugs (Boussaroque *et al*. 2016). When ACT drugs bind to the phosphatidylinositol 3 Kinase (PI3K) protein, its kinase activity is reduced significantly, thus reducing phosphatidylinositol-3-phosphate (PI3P) biochemical activity, and parasite growth is inhibited. Non-mutant Kelch 13 binds PI3K and marks it for ubiquitin ligase, which facilitates the ubiquitination process of Kelch 13, a marker for degradation by proteasome enzymes (Boussaroque *et al*. 2016). This biochemical process cannot be facilitated by mutant Kelch 13 protein because it does not bind PI3K. Our predicted Kelch 13 propeller domain model (Figure 6) was consistent with findings in Senegal by Boussaroque *et al* (Boussaroque *et al*. 2016), results from South East Asia by Ariey *et al* (Ariey *et al*. 2014) and data from Boyden *et al* (Boyden *et al*. 2012) model of human kelch like propeller domain.

Mutations occurring in structural genes are the main drivers of evolution as they may introduce diversity into the organism's genome and proteome, and this occurs mostly through non-synonymous point mutations (Pires et al. 2014). Non-synonymous mutations result in changes of amino acids and are of interest because they interfere with the natural conformation of the protein, hence potentially reducing stability or interaction of the protein with other protein molecules. These mutations can favor selection during the evolutionary course and render functional proteins non-functional by either reducing or increasing their stability (Pandurangan *et al*. 2017).

Intra and inter-blade hydrogen bonding in a mouse Kelch-like ECH associated protein 1 (keap1) domain maintains the structure and ensures effective interaction with keap1 and Nrf2 "regulator of the cellular oxidative stress response". Non-synonymous mutations are associated with development of human cancer because these mutations, change the native structure of keap1, hence reducing the keap1 association with the Nuclear factor erythroid 2-related factor 2 (Nrf2) (Padmanabhan *et al*. 2006). In *P. falciparum*, evolutionary forces affecting the protein targeted with drugs may result in the development of non-responsive parasites to ACTs drugs. We found that most of the non-synonymous mutations occurring in *P. falciparum* in 2015 were associated with reduced stability of the Kelch 13 propeller domain (Table 1). The reduced stability of the Kelch 13 propeller domain may be causing a change in the local conformation of the protein hence affecting its activity, resulting in the development of parasites which are insensitive to ACT drugs but they remain responsive to treatment. This finding is consistent with a study done by Gardner *et al* (Gardner *et al*. 2011), which showed

that non-synonymous point mutations in *P. falciparum* structural genes cause adaptive evolution (resistance) to ACT drugs. The mutation at codon R561P causes a stabilization of the Kelch 13 propeller domain and this can be explained in part to the introduction of a kink by a proline amino acid, which increases the folding of the domain.

The change in the Gibbs free energy score infers the effect of non-synonymous point mutations by predicting the energy difference between wildtype and the mutant protein molecule which predicts the functional consequence of the mutation (Pandurangan *et al*. 2017). These mutations, just like other point mutations occurring on structural proteins, are thought to modulate the Kelch 13 protein function and affect its interactions with other proteins as well as drug molecules. Even though the "Pan African" SNP (M476I) (Torrentino-Madamet *et al*. 2014) selected in parasites *in vitro* was not detected in our samples, it is also thought to affect the Kelch 13 propeller domain local conformation. *P. falciparum* and its disease progression within-host is the product of protein interactions between *P. falciparum* and the human host.

P. falciparum is a threat to life in the tropics (Figure 1) this is due to its high potential for developing resistance to already available antimalarial drugs and its evolutionary dynamics in nature (Saha *et al*. 2017). Paloque *et al*. (Paloque *et al*. 2016) also found that *P. falciparum* adaptive evolution to ACT drugs is due to escalated counts of young ring forms of parasites which enter quiescence when they come into contact with artemisinin and immediately continue growing when the artemisinin drug pressure is withdrawn. The ability to undergo quiescence is made possible to a non-synonymous point mutation in the Kelch 13 propeller domain which is the main target of the ACT drugs. Figure 7 and Figure 8 show the Kelch 13 proper ACT binding pocket and ACT/k13 interaction respectively. Interestingly, these predictions reveal the hypothetical active site of Kelch 13 which is predicted to be on chain A. This explains why the non-synonymous point mutations are mainly located on the intra or inter-blade domains on the Kelch 13 protein (Amato *et al*. 2017; Feng *et al*. 2015). Kelch 13 propeller domain modelling data indicates that most of these mutations increase the relative side chain solvent accessibility in the mutant protein structure (Table 2). This result is in agreement with the Mohon *et al* (Mohon *et al*. 2014) findings in Bangladesh where non-synonymous mutations were shown to change the local conformation of the Kelch 13 propeller domain. L571S, which is associated with the highest destabilization effect in our data (Table 1) on the *pfk13* propeller, was closest to the SNP at codon C580Y (Figure 6), the main point mutation occurring at high frequency in the Cambodian *P falciparum* strains (Ariey *et al*. 2014). Of great importance is the position of these mutations, all of them lie after SNP position 400 and this may have a significant effect on Kelch 13 structure and hence may affect the ACT drug activities. Mutations in our data all occurred after position 400 on Kelch 13 propeller domain and this was consistent with the Paloque *et al's* findings

which showed that most mutations occurring after position 400 are non-synonymous and some of them are strongly associated with ACT resistance (Paloque *et al*. 2016). Measures such as parasite clearance slope half-life and *P. falciparum* clearance rate are multifactorial "phenotypes" which may be affected by the host immunity, drug metabolism or pharmacokinetics and the patient adherence to ACT drugs during treatment (Mohon *et al*. 2014). Increased parasite clearance slope half-life has been documented in Bangladesh even though mutations in the Kelch 13 propeller domain have not been reported which may be explained by the multifactorial effects of within host dynamics mentioned (Mohon *et al*. 2014). In Africa, mutations have been reported in the Kelch 13 propeller domain but the ACT drugs remain efficacious in treatment of malaria in Sub-Saharan Africa (Ouattara *et al*. 2015).

4.5 *Plasmodium falciparum* Kelch 13 phylogenetic tree construction

Evolution of *P. falciparum* within the human host can be described as a "tree evolving inside another tree" (Sennblad *et al*. 2007). As humans evolve, the parasite also undergoes evolution which, in this case, is promoted by high ACT drug pressure within the human host. Sporadic adaptation of organisms to new environmental conditions or stress has been described in a variety of proteins, including stomach proteins in primates, sperm proteins of invertebrates and the variation of bacterial surface proteins (McInerney et al. 2003). These adaptive changes have led to the lifestyle changes and speciation among these organisms from which these changes have been described (McInerney et al. 2003). From our findings, it is evident that a positive selection has occurred in the *P. falciparum kelch 13 gene* (Figure 4 & Figure 5) showed the identified amino acids at codon D559E and D547E which may be the main drivers of adaptive evolution in our *P. falciparum* population in response to ACT drugs. From the Maximum likelihood (Yang 2000) phylogenetic tree (Figure 22), it is interesting to note that over 25.32 % of our sequences are showing evidence of positive selection which drives change in *P. falciparum* to become non-responsive to ACT treatment. This is because they are evolving faster due to intense drug pressure. The positive selection observed may be explained in terms of their location on the Kelch 13 domain; where they are on chain A or Chain B. Amino acids sequences found on chain A are expected to undergo positive selection faster compared with the amino acid sequences on chain B (Figure 8). The reason being that, it is chain A which is predicted to contain the active site where ACT drugs are thought to be interacting with the Kelch 13 propeller domain. Based on the phylogenetic analysis of the Kelch 13 sequences, it is not easy to explain the differential appearances of selective pressure in different *P. falciparum* isolates within the human host, however this may be attributed to differential interaction of *P. falciparum* with the human host immune system. The human immune system works differently and individuals may illicit different

immune responses during infections as was observed by Hastings *et al* (Hastings, Kay, and Hodel 2015).

However, in a previous study conducted by Maeno *et al* (Maeno *et al*. 2017) in Vietnam, they demonstrated that the evolution of *P. falciparum* isolated from the mosquito salivary gland is much slower compared with the human host. This could be because within the mosquito host the parasites are free from the immune attack hence no immune pressure unlike in human where they face both the drug and the immune pressure. Parasite clearance slope half-life is a stochastic event as it varies from parasite to parasite (see Figure 22) and parasites with the same clearance slope half-life may have different evolutionary rates. Most of non-synonymous polymorphisms were present at low frequency and they were specific parasite strains, an indication that they might be transitory point mutations which may be attributed to ACT drug pressure or natural evolution. This explains why most parasite strains have a lower evolution rate and they cluster together in the phylogenetic tree (Figure 22). This result is consistent with Talundzic *et al's* finding in Senegal where they attributed the transient occurrence of non-synonymous mutation to natural evolution due to ACT usage on Senegalese *P. falciparum* parasite populations (Talundzic *et al*. 2017).

Chapter 5. Conclusions and Recommendations

5.0 Conclusions

In this dissertation, we set out to investigate the evolutionary changes in the *P. falciparum* genome due to Artemisinin Combination Therapy (ACT) intervention in Kenya between 2013 and 2015. Our data suggests that *P. falciparum* in Kenya still lacks previously known non- synonymous point mutations such as Y493H, I543T, R539T, C580Y and M476I, which have been associated with the increased parasite clearance slope half-life in Cambodia and other locations in South East Asia. The "Pan African SNP" M476I which was selected in parasites *in vitro* across Africa was not detected in our samples. However, non-synonymous point mutations at codon D559E and D547E are present at high frequency in our samples and affect the local conformation and the native structure of the Kelch 13 propeller domain by destabilizing the protein structure. These two SNPs showed significant associations with the IC_{50s} of AQ, MQ and Artesunate but these ACT drugs were still efficacious in Kenya. The samples from the Kombewa and Maseno divisions, which are in a malaria holoendemic region in Western Kenya showed a high variation in parasite clearance slope half-life, which was expected in part since different *P. falciparum* parasites have different genetic makeup.

Parasites from Kisumu are under positive selection an indication that these parasites are gaining fitness in the presence of ACT and soon they may become non-responsive to these drugs as happened with CQ. In this study a total of seventeen non-synonymous single nucleotide polymorphisms were detected and none showed significant association with the increased parasite clearance slope half-life.

P. falciparum clearance slope half-life time in Kenya is still in the low range (2.22-2.60 hours), compared with Cambodian and Myanmar parasites which take more than 6.2 hours to be cleared. We also found significant negative association between mefloquine and amodiaquine which may be used in our population on a rotational basis to avoid the simultaneous development of adaptive evolution to resistance against these drugs. Population subdivision of *P. falciparum* within the human host showed minimal population differentiation of *P. falciparum* during treatment. In a span of forty-two days, we found parasites with a distinct genetic make-up which is due to the adaptive evolution of these parasites. There was a limitation in the understanding of the evolutionary effects of the *P. falciparum Kelch 13* gene due to ACT usage. In this dissertation we demonstrated clearly that the Kenyan parasites are becoming more fit in the presence of ACT even though they are still susceptible to ACT treatment. Microsatellite data shows that these parasite populations are highly polymorphic having a high potential for adapting within the host to impede ACT intervention. Low population

differentiation implies that resistant parasites may spread easily from one region to the other, which makes malaria control at the local level inefficient.

5.1 Recommendations

The adaptive evolution on the *P. falciparum* Kelch 13 propeller domain is very high due to a rapid increase in non-synonymous mutations, which increases the potential of Kenyan *P. falciparum* to increase its fitness in the presence of ACT drugs. We recommend regular molecular surveillance of this lethal parasite to determine novel SNPs which will point us in the right direction in the fight against malaria in Kenya. An accurate diagnosis of malaria infection is needed before the right dosage prescription of ACT drugs is delivered.

The statistics on the supply and use of counterfeit drugs from both the private and public sector is worrying and quality assurance checks need to be enhanced to avoid the used of these low-quality drugs (Musuva *et al*. 2017; Akulayi *et al*. 2017). We recommend that the Kenya Bureau of standards should ensure that both public and private health facilities prescribe drugs which satisfy the World Health Organization quality standards (Maponga and Ondari 2003) for use in the treatment of *P. falciparum* malaria in the population. The pharmaceutical industries should be under surveillance to ensure that they adhere to the Good Manufacturing Practice (GMP) guidelines when making these antimalarial drugs.

We recommend the continuous surveillance of the molecular markers that might correlate with the ACT drug resistance development by tracking the travel history of the patients visiting health facilities with *P. falciparum* malaria infections. This will inform the treatment and intervention strategies as resistant *P. falciparum* malaria may follow the path of chloroquine resistance (Ashley *et al*. 2014; Molina-Cruz *et al*. 2016). Effective use of ACT drugs in our populations should be encouraged as it will lead to a fast neutralization of the *P. falciparum* parasites within the host, hence derailing the development of drug resistance.

Accurate information must be made available to the consumers as it will ensure the current ACT usage and cure rate is maintained in Kenya (Watsierah *et al*. 2011). The policy makers should ensure that ACT drugs of high quality are in circulation to avoid the use of low quality drugs in the population.

References

(WHO), World health organization. 2017. *World Malaria Report 2017*.

Ariey, Frédéric et al. 2014. "A Molecular Marker of Artemisinin-Resistant Plasmodium Falciparum Malaria." *Nature* 505(7481): 50–55. http://www.ncbi.nlm.nih.gov/pubmed/24352242%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5007947.

Bianchine, P J, and T a Russo. 1992. "The Role of Epidemic Infectious Diseases in the Discovery of America." *Allergy and Asthma Proceedings* 13(5): 225–32. http://www.ncbi.nlm.nih.gov/pubmed/1483570.

Bianucci, Raffaella et al. 2008. "Immunological Evidence of Plasmodium Falciparum Infection in an Egyptian Child Mummy from the Early Dynastic Period." *Journal of Archaeological Science* 35(7): 1880–85.

CDC. 2012a. "About Malaria - History." *Centers for Disease Control and Prevention*. http://www.cdc.gov/malaria/about/history/.

———. 2012b. "CDC - Malaria - About Malaria - Biology - Malaria Parasites." *USA Government*. http://www.cdc.gov/malaria/about/biology/parasites.html.

Chen, Hong, Noboru Minakawa, John Beier, and Guiyun Yan. 2004. "Population Genetic Structure of Anopheles Gambiae Mosquitoes on Lake Victoria Islands, West Kenya." *Malaria journal* 3: 48.

CLYDE, DAVID F., HARRY MOST, VINCENT C. McCARTHY, and JEROME P. VANDERBERG. 1973. "Immunization of Man against Sporozite-Induced Falciparum Malaria." *The American Journal of the Medical Sciences* 266(3): 169–77. http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000441-197309000-00002.

Coluzzi, M, and G Corbellini. 1995. "[The Places of Mal'aria and the Causes of Malaria]." *Medicina nei secoli* 7(3): 575–98. http://www.ncbi.nlm.nih.gov/pubmed/11623488.

Cox, Francis Eg. 2010. "History of the Discovery of the Malaria Parasites and Their Vectors." *Parasites & vectors* 3(1): 5.

Curtin, P D. 1968. "Epidemiology and the Slave Trade." *Political science quarterly* 83(2): 190–216.

Deane, Leonidas M. 1986. "Malaria Vectors in Brazil." *Memórias do Instituto Oswaldo Cruz* 81: 5–14.

Gardner, Kate B et al. 2011. "Protein-Based Signatures of Functional Evolution in Plasmodium Falciparum."

Hay, Simon I. et al. 2009. "A World Malaria Map: Plasmodium Falciparum Endemicity in 2007." *PLoS Medicine* 6(3): 0286–0302.

Hempelmann, Ernst, and Kristine Krafts. 2013. "Bad Air, Amulets and Mosquitoes: 2,000 Years of Changing Perspectives on Malaria." *Malaria journal* 12(1): 232. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3723432&tool=pmcentrez&rendertype=abstract.

Hoffman, Stephen L., Johan Vekemans, Thomas L. Richie, and Patrick E. Duffy. 2015. "The March Toward Malaria Vaccines." *American Journal of Preventive Medicine* 49(6): S319–33.

Holsinger, Kent E., and Bruce S. Weir. 2009. "Genetics in Geographically Structured Populations:

Defining, Estimating and Interpreting FST." *Nature Reviews Genetics* 10(9): 639–50.

Hopkins, H et al. 2007. "COMPARISON OF HRP2AND PLDHBASED RAPID DIAGNOSTIC TESTS FOR MALARIA WITH LONGITUDINAL FOLLOWUP IN KAMPALA, UGANDA." *American Journal of Tropic Medical Hygiene* 76(6): 10921097.

Hume, Jennifer C C, Emily J Lyons, and Karen P Day. 2003. "Human Migration , Mosquitoes and the Evolution of Plasmodium Falciparum." 19(3): 144–49.

Karuri, Stella Wanjugu, and Robert W. Snow. 2016. "Forecasting Paediatric Malaria Admissions on the Kenya Coast Using Rainfall." *Global Health Action* 9(1).

Keenan, Kevin et al. 2013. "DiveRsity: An R Package for the Estimation and Exploration of Population Genetics Parameters and Their Associated Errors." *Methods in Ecology and Evolution* 4(8): 782–88.

Liu, Weimin et al. 2010. "Origin of the Human Malaria Parasite Plasmodium Falciparum in Gorillas." *Nature* 467(7314): 420–25. http://dx.doi.org/10.1038/nature09442.

Molina-Cruz, Alvaro et al. 2016. "Mosquito Vectors and the Globalization of Plasmodium Falciparum Malaria." *Annual Review of Genetics* 50: 447–65. http://www.annualreviews.org/doi/abs/10.1146/annurev-genet-120215-035211.

Neghina, Raul, Adriana Maria Neghina, Iosif Marincu, and Ioan Iacobiciu. 2010. "Malaria, a Journey in Time: In Search of the Lost Myths and Forgotten Stories." *The American journal of the medical sciences* 340(6): 492–98.

Ocan, Moses et al. 2016. "Prevalence of K13-Propeller Gene Polymorphisms among Plasmodium Falciparum Parasites Isolated from Adult Symptomatic Patients in Northern Uganda." *BMC Infectious Diseases* 16(1): 428. http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-016-1777-7.

Oyebola, Kolapo M. et al. 2017. "Pooled-DNA Sequencing Identifies Genomic Regions of Selection in Nigerian Isolates of Plasmodium Falciparum." *Parasites and Vectors* 10(1): 1–7.

Ramírez-Soriano, Anna et al. 2008. "Statistical Power Analysis of Neutrality Tests under Demographic Expansions, Contractions and Bottlenecks with Recombination." *Genetics* 179(1): 555–67.

Roper, Cally et al. 2004. "Intercontinental Spread of Pyrimethamine-Resistant Malaria." *Science* 305(5687): 1124. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15326348.

Sallares, Robert, Abigail Bouwman, and Cecilia Anderung. 2012. "The Spread of Malaria to Southern Europe in Antiquity: New Approaches to Old Problems." *Medical History* 48(03): 311–28.

Sifuna, Peter et al. 2014. "Health and Demographic Surveillance System Profile: The Kombewa Health and Demographic Surveillance System (Kombewa HDSS)." *International Journal of Epidemiology* 43(4): 1097–1104.

Snow, Robert W. 2015. "Global Malaria Eradication and the Importance of Plasmodium Falciparum Epidemiology in Africa." *BMC medicine* 13(1): 23.

Tajima, F. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123(3): 585–95.

Talisuna, A O, P Bloland, and U D'Alessandro. 2004. "History, Dynamics, and Public Health Importance of Malaria Parasite Resistance." *Clin.Microbiol.Rev.* 17(0893–8512 (Print)): 235–54.

doi: 10.1128/CMR.17.1.235.

Tatem, Aj J, Dj J Rogers, and S I Hay. 2006. “Global Transport Networks and Infectious Disease Spread.” *Advances in parasitology* 62(05): 293–343. http://www.sciencedirect.com/science/article/pii/S0065308X0562009X%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3145127&tool=pmcentrez&rendertype=abstract.

World Health Organization. 2015. World Health *World Malaria Report 2015*.