COMPARATIVE GENOMIC ANALYSIS OF *PLASMODIUM FALCIPARUM* ISOLATES FROM ETHIOPIAN WEST ARSI, EAST AFRICA AND SOUTHEAST ASIA

DERIBA ABERA BEYENE

MASTER OF SCIENCE

(Molecular Biology and Bioinformatics)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2021

Comparative genomic analysis of *Plasmodium falciparum*isolates from Ethiopian West Arsi, East Africa and Southeast Asia

Deriba Abera Beyene

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Molecular Biology and Bioinformatics of the Jomo Kenyatta University of Agriculture and Technology

2021

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature  Date:10 - 3 – 2021

DeribaAberaBeyene

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

Signature…… Date…….

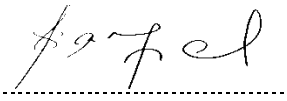
Dr. Joel L. Bargul, PhD

JKUAT, Kenya

Signature…… Date…….

Dr. Caleb Kibet, PhD

ICIPE, Kenya

Signature:  Date: 9-3-2021

Dr. LemuGolassa, PhD

ALIPB, Addis Ababa University, Ethiopia

DEDICATION

I would like to dedicate this work to my family for their unwavering support.

ACKNOWLEDGEMENTS

I would like to forward my gratitude acknowledge to the following institutions: DELGEME for financing my study and research project; Jomo Kenyatta University of Agriculture and Technology, where I'm registered for Master's degree;and Jimma University for providing infrastructural support, including internet connectivity. I extend my gratitude to my supervisors: 1) Dr. LemuGolassa for availing *P. falciparum* sequence data for my study, and for guidance, 2) Dr. Joel Bargul for his guidance during my research studies, as well as ensuring that I met all the University's requirements for the award of MSc degree, and 3) Dr. Caleb Kibet for his constructive advice and comments on my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ACT	Artemisinin-based Combination Therapy
AMA1	Apical Membrane Protein Gene 1
Bp	Base pairs
CDC	Center for Disease Control
CQ	Chloroquine
CQR	Chloroquine Resistance
CQS	Chloroquine Sensitive
CRT	Chloroquine Resistance Transporter Gene
BQSR	Base Quality Scoring Recalibration
DDT	Dichlorodiphenylchloroethane
DHFR	Dihydrofolate Reductase gene
DHPS	Dihydropteorate Synthase
DV	Digestive Vacuole
EBA175	Erythrocytes Binding Antigen Gene 175
EHH	Extended Haplotype Homozygosity
FST	Wright's F statistics
FMoH	Federal Ministry of Health
IHS	Integrated Haplotype Score
ITN	Insecticidal Treated Nets
Kbp	Kilo base pair
LD	Linkage Disequilibrium
LLIN	Long lasting Insecticidal treated Nets
MAF	Minor Allele Frequency

MBP	Mega base pairs
MDR1	Multidrug Resistance Transporter Gene1
MSP1	Merozoites Surface Protein Gene 1
PfCRT	*Plasmodium falciparum* Resistance Transporter Gene
PfMDR1	*Plasmodium falciparum* Multidrug Resistance Transporter Gene1
Phg1	P-glycoprotein related transporter
SNP	Single Nucleotide Polymorphism
SP	Sulfadoxine/Pyrimethamine drugs
VQSR	Variant Quality Scoring Recalibration

ABSTRACT

Local selective pressures from host immunityand drugs have led the *Plasmodium falciparum (P. falciparum)* parasite population to evolutionary adaptation. However, the evolutionary adaptations and genomic characteristics of *P. falciparum* isolates from the west Arsi, Ethiopia are poorly understood. After removing 9 samples with low quality SNPs, 25 *P. falciparum* isolates fromEthiopianwest Arsiwere studied to determine their population structures, genetic diversity, and signatures of selection in known drug resistance genesagainst isolates from Cambodia, DR Congo, Malawi and Thailand.About 18,517 high-quality single-nucleotide polymorphisms (SNPs) were identified in *P. falciparum* isolates from West Arsi, Ethiopia with overall average nucleotide diversity (π = 0.00022) across the genome. Most of the *P. falciparum* isolates (84%) from west Arsi of Ethiopia showed F_{WS} value >0.95 indicating single genotype infection dominance in most samples at the time of sample collection as expected in parasites in low area of transmissions. The Ethiopian *P. falciparum* population ofwest Arsiwas structurally differentiated from Southeast Asian and east African populations, suggesting independent evolution of the EthiopianWest Arsi parasite population. Furthermore, a total of 125 genes known to be under balancing selection were identified. These genesincluded*ama1, trap, eba175,* and *lsa3,* which were previously identified as human host immune targets and responsible for vaccine development. The directional selection signature analysis with integrated standardized haplotype score (IHS) did not detect any recent selection signatures in the *Pfdhps*,*Pfmdr*1,*Pfdhfr, Pfcrt,* and *PfK13* genesof *P. falciparum* isolates from west Arsi, Ethiopia. However, known antimalarial resistance-conferring SNPs analysis detected at least one SNP marker that was fixed in these genes, but not in *PfK13*and*Pfdhps*.The Ethiopian*P. falciparum* of West Arsiwas structurally diverged from both other East African and Southeast Asian*P. falciparum* populations. The Ethiopian*P. falciparum* population of West Arsicarries fixed chloroquine and Sulfadoxine/Pyrimethamine (SP) resistance markers despite the removal of these drugs from the National*P. falciparum* malaria treatment program of Ethiopia.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Malaria has been known to be a major public health problem worldwide (WHO, 2019). In 2018, malaria caused approximately 228 million cases and 405,000 deaths of which the majority of cases and deaths occur in sub-Saharan Africa (WHO, 2019). Most of malaria cases and deaths were reported from children below five years of age and pregnant women with *Plasmodium falciparum* considered as a major cause of malaria (WHO, 2017)**.**

Similar to othersub-Saharan African countries,Ethiopia suffered from the deadliest *P. falciparum* malaria that causesa significant morbidity and mortality(Lo *et al*., 2017)**.** More than 68% of the Ethiopian population that live on about 75% of the country's total area is at risk of getting malaria(Seyoum *et al*., 2017; Taffese *et al*., 2018). Both major human malaria-causing parasites,*P. falciparum* and *P. vivax* co-exist in Ethiopia (Lo *et al*., 2017), making malaria control and prevention more difficult than in other African countries.

Malaria transmission exhibits heterogeneity and seasonal patterns in Ethiopia, with the main transmission season from September to December, while the shorter transmission season covers from April to May in some parts of the country (FMoH, 2011).Malaria transmission occurs year-round in the southwestern lowland regions of Ethiopia(Taffese*et al*., 2018). Ethiopia has implemented massive malaria intervention strategies that include indoor residual spray (IRS), insecticide-treated bed nets (ITN) and antimalarial drug administration schemes throughout the country in the past few years (Seyoum *et al*., 2017). These massive deployments of intervention and control measures have led to the emergence of drug resistance to chloroquine (CQ) and anti-folate drugs (Golassa *et al*, 2014). The development of resistance to CQ and SP shifts the malaria treatment drug to artemether-lumefantrine (AL) currently being used as the first-line drug in the treatment of *P. falciparum* infection since 2004 (FMoH, 2011).

The development of drug resistance in *P. falciparum* hindersmalaria elimination progress, thus complicating malaria control and intervention. Similar to CQ and SP,

resistance to the current first-line drug, artemisinin combined therapy (ACT) was reported in Southeast Asia (Ashley *et al*., 2014). CQ was among the first antimalarial drugs used in the malaria-endemic areas, and the first resistance was reported from Thailand in the late 1950s (Mekonnen *et al*., 2014). Since then, more cases of drug resistance were reported in Southeast Asia, Africa in 1974, and eventually reached East Africa in the 1980s (Mekonnen *et al*., 2014). Ethiopia reported the first CQ resistance in 1985 from areas bordering Kenya, Somalia, and Sudan, whereas the central part of the country was free from resistant strains (Alene & Bennett, 2007). In 1998 SP replaced CQ and SP was in turn replaced by artemisinin combination therapy (ACT) called artemether-lumefantrine (AL) since 2004 (Heuchert et al., 2015; Mekonnen et al., 2014). However, CQ is still used as the first-line drug for *P. vivax* treatment in Ethiopia (Golassa *et al*., 2014) and continued selecting for chloroquine resistance mutation in *P. falciparum*.

Parasites change and select their new genetic variants owing to continued drug exposure, and therefore, they cause disease by overcoming challenges from host immunity and therapeutic interventions (Kidgell *et al*., 2006). For instance, previous studies have demonstrated that high pressure mounted on the parasites by host immunity and exposure to antimalarial drugs select adaptive strains, which in turn maintain malaria transmission (Oyebola *et al*., 2017). Consequently, many *P. falciparum* genes encoding immune and drug targets are under natural selection and show signatures of balancing or directional selection (Manske *et al*., 2012; Mobegi *etal*., 2014; Ocholla *et al*., 2014; Duffy *et al*., 2015; Oyebola *et al*., 2017; Amambua-Ngwa *et al*., 2018).

These selection signatures may vary due to differences in innate susceptibility of human or mosquito populations, variation in ecological transmission, degrees of acquired immunity in humans, or drug pressure(Mackinnon and Read, 2004). In low endemic areas,malaria parasites require prolonged maintenance of asexual parasite infection due to the reduced chances of transmission of the sexual stages. In contrast, parasites in high endemic regions are subjected to higher competition due to superinfection by different genotypes and face more robust acquired immune responses from their hosts (Mackinnon and Read, 2004).

For effective management of control and intervention strategies, it is important to determine genetic variation patterns due to parasite adaptation to host environments and drug exposure. Balancing selection brings the favored alleles of parasites to an intermediate equilibrium where they are maintained as genetic polymorphisms.At the same time, the directional forces cause parasite's genetic variants to increase in frequency and facilitate the occurrence of selective sweeps around affected loci (Biswas and Akey, 2006).

Population genomics of *P. falciparum* has been mostly studied in the West African parasite populations and showed signatures of balancing selection on multiple candidate vaccine antigens and strong directional selection around known drug resistance genes (Amambua-Ngwa *et al*., 2012; Mobegi *et al*., 2014). In contrast, there is little information about the genomic variations of *P. falciparum* populations in the horn of Africa, including Ethiopia, where *P. falciparum* and *P. vivax* malaria co-exist and are heterogeneously distributed. *P. falciparum* populations found in the horn of Africa, specifically in Ethiopia, were previously reported to be structurally divergent from the West, East, and central African *P. falciparum* populations (Amambua-Ngwa *et al*., 2019). These parasite populations share a chunk of genes with other sub-Saharan African *P. falciparum* populations across drug and immunity targeted genomic loci and facilitate the emergence and spread of drug-resistant strains (Amambua-Ngwa *et al*., 2019). This study called an in-depth analysis of the Ethiopian parasite genomes to deepen understanding of genome diversity and natural selection in Ethiopia's unique human populations with co-species transmission dynamics.

Understanding the population genetic diversity of the *P. falciparum* strains circulating in the specific regions is very important to monitor the effectiveness of control schemes and provide baseline information for making informed decisions (Jalei*et al*, 2018). Here, we analyzed the whole genome of 25 *P. falciparum* isolates obtained from central Ethiopia and identified genetic diversity, population structure, selection signatures, as well as the prevalence of drug resistance mutations in *Pfcrt, Pfmdr1, Pfdhps, Pfdhfr,* and *Pfkelch13* genes.

1.2 Statement of the Problem

Malaria continues to be one of the leading public health problems in sub-Saharan Africa, including Ethiopia. About two thirds of Ethiopian populations live in malaria-endemic regions and are at risk of contracting malaria, particularly from *P. falciparum.* Ethiopia has applied different malaria controlling strategies, including vector control (IRS and ITNs) and drug administration policies, to control malaria transmission and facilitate progress towards malaria elimination (Taffese *et al*., 2018). However, the development of drug-resistant *P. falciparum* and insecticide-resistant *Anopheles* mosquito makes malaria elimination difficult. Currently, artemisinin combined therapy is the common *P. falciparum* malaria treatment strategy in the malaria-endemic area. In Ethiopia, artemether-lumefantrine (AL) is the current effective anti-malarial drug for uncomplicated *P. falciparum* treatment (FMoH, 2011). However, ACT resistant *P. falciparum* was reported from Southeast Asia (Ashley *et al*., 2014) and recently in East Africa (Uwimana *et al*., 2020), promoting malaria transmission, threatening control and elimination strategies in sub-Saharan Africa. This is further complicated by reports of chloroquine and SP drug-resistant *P. falciparum* from Southeast Asia and later detected in East Africa (Lin, 2011). It is important to find alternative drugs or promising vaccines for this emerging drug-resistant *P. falciparum* to decrease the risks of malaria transmission and speed up disease elimination strategies in the near future.

To develop promising approaches and policy for malaria prevention and elimination, it is crucial to understand genomic variations of *P. falciparum* populations from specific geographical regions. Moreover, determining the evolutionary relationships, selection signatures, and population structure could reveal important region-specific parasite information to adjust malaria prevention, control, and elimination policies and programmes in malaria-endemic areas. For instance, Ocholla *et al.* 2014identified genomic variation in *P. falciparum* population and reported the return of chloroquine-sensitive *P. falciparum* malaria in Malawi after 15 years of chloroquine withdrawal; encouraging the idea that determining genomic diversity, selection signatures, and population structure are the key parameters to inform malaria treatment policies. This study aimed at determining molecular evolution and

signatures of selection in the genome of natural *P. falciparum* population in Ethiopia by comparing to other African and Southeast Asian isolates.

1.3 Justification

The drug-resistant *P. falciparum* was reported from almost all malaria-endemic regions (Schunk *et al*., 2006; Amambua-Ngwa, Park, *et al*., 2012; Manske *et al*., 2012) and limited the efforts of malaria control, intervention, and elimination. These reports changed the malaria drug administration policy from CQ to SP and SP to ACT over time. ACT is a common effective malaria drug used in sub-Saharan Africa, where *P. falciparum* malaria transmission is high, particularly in Ethiopia (FMoH, 2011). However, a genomic variation scan of *P. falciparum* population identified ACT drug-resistant parasite strains before 2014 in Southeast Asia (Ashley *et al*., 2014) and recently in East Africa (Uwimana *et al*., 2020), justifying the importance of surveillance of genomic variations across East Africa in general and Ethiopia in particular. The detection of*P. falciparum* genomic variation, determination of selection signatures, the understanding of population structure as well as estimation of allele frequency changes in each known drug target genes could help control the spread of drug-resistant parasites.

1.4 Research Questions

(i) Is there genetic diversity in the genome of natural *P. falciparum* isolates of west Arsi, Ethiopia?

(ii) Is the *P. falciparum* populationof west Arsi of Ethiopia evolutionarily diverged from East African and Southeast Asian *P. falciparum* populations?

(iii) Is there gene flow between *P. falciparum* population of west Arsi, East Africa, and Southeast Asia?

(iv) Are there selection signatures in the genome of *P. falciparum* population of west Arsi, Ethiopia?

1.5 Hypotheses

(i) There is low genetic diversity in the genomes of *P. falciparum* isolatesfrom west Arsi, Ethiopia.

(ii) *Plasmodium falciparum* population in west Arsi of Ethiopia is evolutionarily not diverged from other East African and Southeast Asian.

(iii) There are no selection signatures in the genome of the natural *P. falciparum*populationin west Arsi of Ethiopia.

1.6 Objectives

1.6.1 General Objective

To determine molecular evolution and signatures of selection in the genome of natural *P. falciparum* population in west Arsi,Ethiopia compared to other African and Southeast Asian parasite isolates.

1.6.2 Specific Objectives

(i) To determine genetic diversity within the genome of the *P. falciparum* population in west Arsi,Ethiopia.

(ii) To study evolutionary genetic relationships between Ethiopian(west Arsi), East African, and Southeast Asian *P. falciparum* populations.

(iii) To determine signatures of selection in the genome of the natural *P. falciparum* population of west Arsi, Ethiopia.

CHAPTER TWO

LITERATURE REVIEW

2.1 Malaria and Its Causative Agents, *Plasmodium* Parasites

Malaria is a preventable and curable disease caused by a protozoan parasite of the genus *Plasmodium* transmitted to humans from infected bite of female *Anopheles* mosquitoes during blood feeding on the human host.There are five known *Plasmodium* species that infect humans worldwide; they include *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae,* and *Plasmodium knowlesi*(< 1%) (Weekley and Smith, 2013). *Plasmodium falciparum* is the most prevalent malaria parasite globally, while *P. vivax* is the dominant parasite in the Americas (WHO, 2018).

2.2 Malaria Transmission

Malaria disease continues to be a critical public health problem worldwide. Approximately 228 million human malaria infection cases and 405,000 deaths occur each year (WHO, 2019). The majority of infection cases and death disproportionately affects vulnerable groups that include children below five years of age and pregnant women. Malaria threatens the economic and health development of countries in the subtropical and tropical regions of the world, particularly in sub-Saharan Africa where 90% of cases and 91% of deaths exist (WHO, 2017). Between 2010 and 2018, there was a significant decline in malaria mortality and morbidity due to intense malaria control and monitoring programs that include mass malaria drug administration, usage of long-lasting insecticide nets (LLINs), and indoor residual spray (WHO, 2019). However, malaria continues being a major cause of health problems in sub-Saharan Africa. It is heterogeneously distributed with a few regions progressing towards disease elimination, while transmission remains in others. This heterogeneity might be due to variations in environmental factors such as seasonality, vector distribution, rainfall, human, and host factors, and varying levels of interventions (WHO, 2018).

2.3 Malaria Infections in Ethiopia

Similar to other sub-Saharan African countries, Ethiopia has suffered from malaria, mostly from *P. falciparum* species. *P. falciparum* is responsible for 60% of cases. In comparison, *P. vivax* accounts for 40% of cases in the country (PMI, 2019). The malaria cases and the deaths have decreased over time in Ethiopia with spatial differences in disease transmission intensity (Taffese *et al*., 2018). Malaria is highly seasonal in many parts of the country, but it is constantly transmitted in some other areas. The main malaria season affecting most parts of Ethiopia occurs from September to December, while some regions experience minor malaria transmission from April to June. Since peak malaria transmission occurs during planting and harvesting season, most malaria burden affects working adults and olderchildren in the rural agricultural areas as well as migrants moving around the western corridor of the country that borders Sudan and South Sudan (PMI, 2019). This variability in disease transmission is as a result of many factors, including the co-existence of *P. falciparum* and *P. vivax* pathogens, antimalarial drug susceptibility levels, climate, the efficiency and distribution of mosquito vectors, and other environmental conditions (FMoH, 2011). The total malaria prevalence in the country was 0.5 percent by microscopy and1.2 percent by RDTs for areas below 2,000 meters and less than 0.1 percent in regions above 2,000 meters (PMI, 2019).

The variable winds, seasonal rains, and ambient temperatures experienced in Ethiopia create diverse micro-climates for proliferation of *Anopheles* mosquito that could rapidly accelerate focal malaria transmission (PMI, 2019). *Anopheles Arabiensis* is the main malaria vector in Ethiopia, while *An. funestus, An. pharoensis,* and *An. nili*are secondary vectors. *An. Pharoensis*showed a high level of insecticide resistance and are widely distributed in Ethiopia, while, *An. nili*is the key vector in the western part of the country in Gambela(PMI, 2019).

2.4 Life Cycle of Malaria

Malaria parasites display a complex life cycle involving human host and the mosquito vector. All human malaria-causing *Plasmodium* species*P. vivax,P.*

malariae, *P. ovale*,*P. falciparum*have a similar life cycle with minor variations (Wiser, 2000). The natural history of malaria begins when malaria-infected female mosquito bites and injects sporozoites into the human host. This cycle involves cyclical infection of female *Anopheles* mosquitoes and humans. Inside the human host, these parasites infect hepatocytes, multiply and grow in the liver cells (exoerythrocytic schizogony stage) and thereafter released to invade red blood cells (RBCs). Successive parasite progenies grows inside the RBCs(erythrocytic schizogony) and later releases merozoites, the daughter parasites, that invade the new RBCs to continue the life cycle (CDC, 2018).Merozoites undergo asexual replication to give rise to more merozoites and ring form trophozoites. These trophozoites can be enlarged by ingesting host RBCs cytoplasm and proteolysis hemoglobin into amino acids (Wiser, 2000).

On the other hand, the parasites changed to sexual stages known as micro & macrogametocytes that can be taken up by mosquito vector to initiate a new life cycle and induce gametogenesis (the production of gametes). Microgametes fertilize the macrogamete and lead to a zygote formation. This zygote is developed to ookinete that penetrates the epithelial cells of the gut of mosquito & developed to an oocyst. Then oocyst asexually replicates to produce sporozoites that migrate to the salivary glands of mosquito and complete the life cycle (Wiser, 2000; Simonetti, 1996).

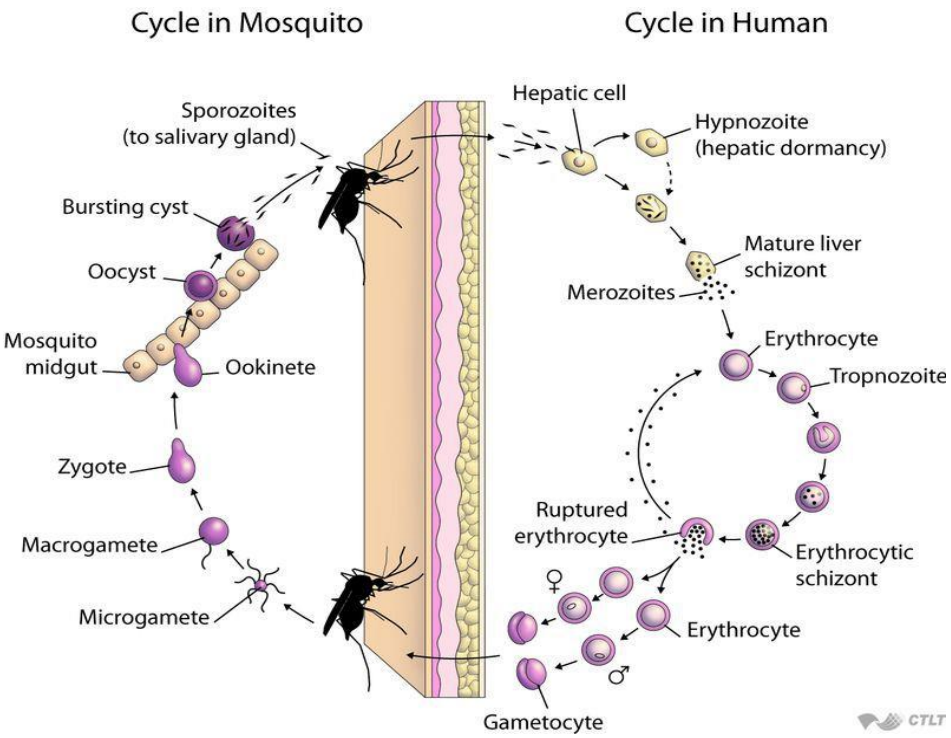


Figure 2.1 :Life Cycle of Malaria in Human Host and Mosquito Vector

Source: Malaria Site, 2018.

2.5 Malaria Diagnosis

Early disease diagnosis and treatment can reduce mortality and morbidity due to malaria, particularly the deadly type caused by the most severe *P. falciparum* species. There are different techniques to diagnose malaria, including clinical (symptoms and physical examination), microscopic, antigen detection, and molecular diagnostic techniques. Even though they all have varied sensitivity and specificity, microscopic diagnosis is considered as a gold standard for diagnosis of malaria (CDC, 2018), and antigen detection techniques (RDT) has been the most commonly used worldwide (WHO, 2018)

2.6 Malaria Prevention and Treatment

There are many strategies to control and prevent malaria transmission to decrease or prevent mortality and morbidity. However, the effectiveness and efficacy of each strategy depend on malaria transmission intensity and target population at risk of developing the disease.

This strategy includes treatment with antimalarial drugs, vector control (IRS, ITN), and vaccination (WHO, 2017).

The ACT is the backbone of global malaria control programs and becomes the main treatment drug combination for global malaria control over many years after CQ and SP drugresistance have been reported (Sibley, 2015).ACT achieves rapid removal asexual stages of the parasites, which cause blood-stage infection andhave an important role in reducing gametocytes responsible for transmission (Lin, 2011;Sibley, 2015). However,artemisinin-resistant *P. falciparum*,with diminished parasite response to artemisinin, was reported in the great Mekong region of Southeast Asia (Sibley, 2015). In this region, artemisinin has completely lost its potency (Sibley, 2015), and the drug-resistant parasites may spread to other countries in the world in a similar way that drug-resistant malaria against chloroquine and sulfadoxine/Pyrimethamine initially emerged in Southeast Asia and later spread to other countries like Ethiopia (Lin, 2011). Similarly, drug resistance against ACT was first reported in Southeast Asia and thereafter spread to East Africa (Uwimana *et al*., 2020). Emergence of drug resistant parasites and their spread to distant regions diminishes hope for malaria prevention, control, and elimination. Nonetheless, Ethiopia has an ambitious roadmap aimed at eliminating malaria in the next ten years, i.e. by 2030 (PMI, 2019).

Next-generation sequencing was used to generate genome sequence of *P. falciparum*,thus enabling identification of polymorphic site frequency distribution in genes with consistent selection. Currently, high throughput sequencing strategy is feasible for sequencing small genome size of 23 Mb in 14 chromosomes of *P. falciparum*. Despite high A+T content (70% in genes, 90% in intron and intergenic region), long tandem repeat gene families make malaria genomic analysis difficult. This high throughput sequencing is important in identifying genetic variation, single nucleotide polymorphisms (SNPs) and enable the estimation of allele frequencies (Gardner et al., 2002). For instance, high throughput sequences of *P. falciparum* isolates enabled the identification of various genes under potential strong directional and balancing selection from the antimalarial drug target or human host immunity targets (Oyebola*et al*., 2017). Examples of genome regions under natural selection are the *Pfcrt* gene found in chromosome 7 and the *Pfmdr1* in chromosome 5 (Ocholla

et al., 2014; Duffy *et al*., 2015) that are involved in chloroquine resistance. *P. falciparum* strains that developed resistance to CQ and two or more additional operational drugs are termed multidrug-resistant. Parasite with an increased copy number of SNPs in *Pfmdr1* has reduced responses to chloroquine, mefloquine, quinine, lumefantrine, and ACT combinations containing these drugs (Lin, 2011).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Area and Population

The study was conducted in West Arsi, Oromia (07’ 17’’ 34.2 S, 038’21’’ 46.3 W), located about 251 km from Addis Ababa, Ethiopia. This region with distinct seasons has an altitude of about 1500-2300 m above sea level with a human population of 176,671(Golassa *et al*., 2014). The inhabitants of this malaria region live with relatively high rates of poverty and are affected by malaria caused by *P. falciparum* and *P. vivax*. Malaria transmission was seasonal and unstable in this region (Golassa *et al*., 2014).

3.2 Sample Size Determination

DNA sequences of 34 *P. falciparum* samples collected from west Arsi, Ethiopiawere obtained from MalariaGEN*Plasmodium falciparum* Community releases 6.0. To compare the *P. falciparum* population in west Arsi, Ethiopia with other East African (Malawi and DR Congo) and Southeast Asian *P. falciparum* populations (Thailand and Cambodia), 50 *P. falciparum* clinical isolates from each country were downloaded from MalariaGEN*P. falciparum* Community release 5.0.

Figure 3.1:Ethiopian Map Showing the Sampling Site in west Arsi (blue pin).

The criteria used to download these *P. falciparum* isolates were as follows: 1) the DNA sequences should be from *P. falciparum* clinical isolates, 2) collection year of the isolates should be the same or close to the Ethiopian (west Arsi)*P. falciparum* isolates' collection year. The *shuf* Linux command line (https://linuxhint.com/bash_shuf_command/) was then used to randomly select 50 *P. falciparum* isolates from each country. The total sample size for this study was 234 computed by using the following formula by Pourhoseingholi *et al*., 2013n = Z^2 P (1-P)/d^2, where n is the sample size, P is the prevalence, d is effect size related to power and Z is the level of confidence at 95% confidence interval. The prevalence of *P. falciparum* in the study area is 0.03 (Meseret*et al*., 2016), Z at 95% confidence is 1.96, and the standard error is 0.05. This study was expected to detect single nucleotide polymorphisms in genomic loci with 99.88% precision (d=0.12) when 234 samples data were analyzed together.

In many cases, it is difficult to sequence large samples from one specific population and compare with other population due to time and resource limitation. To solve this problem, many researchers tried to determine the optimum sample size for

population genetic studies(Bemmels *etal*., 2017; Flesch *etal*., 2018; Hale *etal*., 2012; Trask *et al*., 2011). For instance, Bemmels *et al*. (2017) explained that a study with a sample size greater than 8 has little effect on accurate intra and inter-population genetic diversity estimates. Therefore, this study with at least 34 *P. falciparum* isolates from each population was feasible to determine population genomic variation within each country.

3.3 Sample Collection and Processing

Community health institution-based cross-sectional study was undertaken to collect *P.falciparum* positive samples from west Arsi, Ethiopia by Golassa*etal*., 2014. Briefly, Community health institution-based cross-sectional study was undertaken between July 2012 and December 2013 at West Arsi of Ethiopia.Venous blood (2 - 5mL) was collected from consented *P. falciparum* malaria positive patients following ethical and protocols approved by Addis Ababa University and AHRI-ALERT ethical review committees. Sequencing of *P. falciparum* from leukocyte-depleted infected whole blood was done as described in (Auburn *et al*., 2011) at the Welcome Sanger Institute as part of the MalariaGEN*Plasmodium falciparum* Community Project (www.malariagen.net/projects). Short sequence reads were generated on the Illumina HiSeq platform. Open-source sequences data from Southeast Asia were chosen randomly and sequences of clinical isolates from Cambodia and Thailand were randomly selected to be compared to west Arsi, Ethiopian isolates' sequences as similar species circulate in both Southeast Asia and Ethiopia.TheCongolese (DR Congo) and Malawian clinical isolates' sequences were the only freely accessible sequences data from east Africa during this analysis. All open-source sequences data wereaccessed via the Pf3K project (https://www.malariagen.net/data/pf3k-5).The sequence reads were aligned to Pf3D7 reference (version 3) by BWA (burrows-Wheeler-Alignment) tools version 0.7.17-r1188 (https://github.com/lh3/bwa.git). Single nucleotide polymorphism calling was done following the GATK best practice pipeline (Auwera et al., 2014).

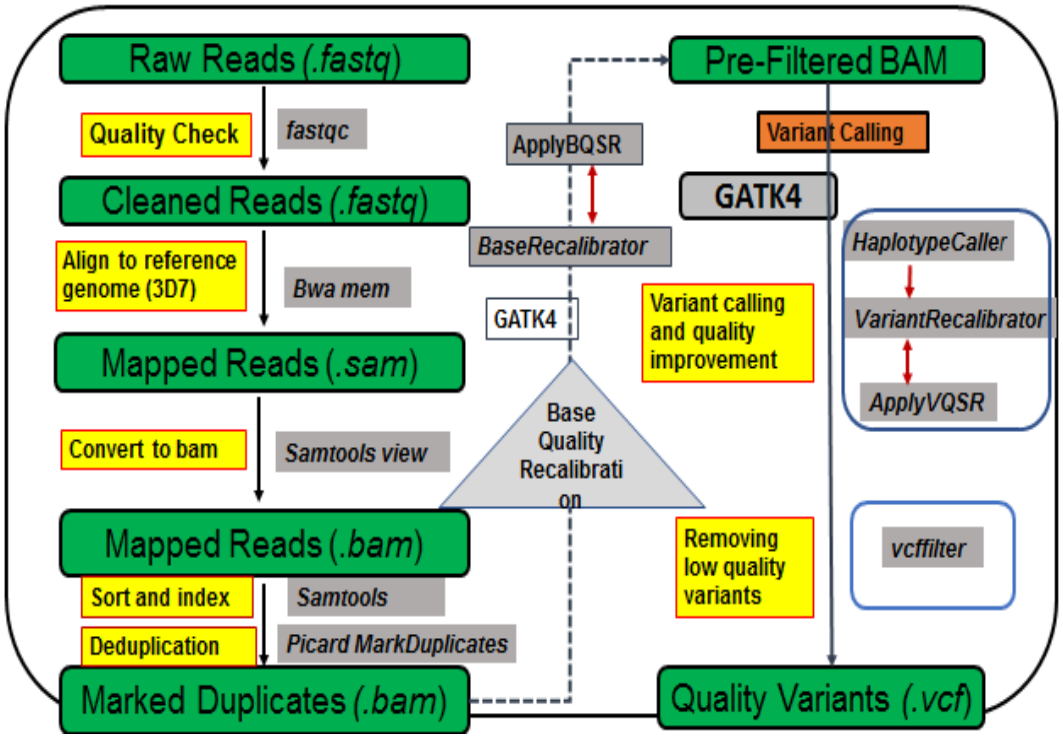


Figure 3.2:Data Processing Pipeline Summary.Grey boxes show different application tools used for data preprocessing and processing, and yellow boxes show the function of respective tools.Green boxes show sequence data in different data formats.

Each sample was genotyped for polymorphic coding SNPs across the genome, ensuring a minimum of 5x paired-end coverage across each variant per sample. Polymorphic sites within hypervariable, telomeric, and repetitive sequence regions were excluded. Vcftools(Danecek *et al.*, 2011) and bcftools (https://github.com/vcftools/vcftools.git) were used to filter the genomic data. Bi-allelic high-quality SNPs (MQ > 20, VQSLOD >= 3) in the core region loci with a minor allele frequency of at least 2% and individual sample and SNP-site missing less than 10% across the isolates was extracted and used for downstream analysis as used in (Amambua-Ngwa *et al.*, 2019).

3.4 Analysis of Population Genetic Diversity and Mixedness of Infection

Pairwise nucleotide diversity (π) (Nei and Li, 1979) was used to determine average pairwise nucleotide genetic diversity using a 5-kb sliding window (Ye *et al.*, 2019).F_{WS}(inbreeding coefficient within a population) metrics was used to calculate

within-host diversity (within-sample diversity) as described in Manske *et al*., 2012 to estimate genotype mixedness of infection in each isolate as used in West African *P. falciparum* genomics analysis (Mobegi et al., 2014). To derive F_{WS} (= 1- H_W/H_S), within isolate expected heterozygosity (H_W) was calculated from the relative allele frequencies for all genic SNPs and compared with the heterozygosity of the local population (H_S). F_{WS} value ranged from zero to one, where zero indicates a high diversity of infection, and one represents a single infection within the sample as compared to local population diversity. For this analysis, individual alleles with coverage of < 5 reads and positions with total coverage of < 20 reads were classified as missing data. Isolates with > 5% missing SNP data and SNPs with > 10% missing isolate data were discarded. Isolates with F_{WS} scores of > 0.95 were classified as isolates having a single predominant genotype due to low genome-wide diversity.

3.5Analysis of Population Structure and Admixture

Principal component analysis (PCA) was used to estimate population structure. This method simplified the complexity in high-dimensional data and reduced data by geometrically projecting them onto lower dimensions called principal components (PCs). PCA summarizes the data using a limited number of PCs. The first PC minimizes the total distance between the data and their projection onto the PC while maximizing the variance between the projected points. The subsequent PCs were similarly selected and uncorrelated with all previous PCs. PC1 has the greatest variance as compared to other PCs and followed by PC2 and so on. However, all the PCs are not typically useful because the majority of variance of the patterns of the data is covered by the first few PCs (Lever, Krzywinski, and Altman, 2017).

Further, the *glPCA* function in the R tool was used to calculate the first 10 principal components axis and retained the first three PCA axis to explain the genomic variation among *P. falciparum* populations. The data was thinned downed by pruning SNPs with pairwise linkage disequilibrium (LD) by r2 greater than 0.05 for determining PCA. To calculate an allele sharing matrix, we used the pruned SNP loci employed in *glPCA* function in custom R script. This function use between and within-group variance to determine population genetic structure. But targeting only

between groups alleles frequency variance is very important to understand the real relationship among populations from the different geographical regions.

Discriminant analysis of principal components (DAPC) (Jombart *etal*., 2010) was used to transform the data using PCA, and then perform a discriminant analysis on the retained principal components using the *adegenet*package for the R software version 3.6.2. This method identified and described population genetic clusters and their relationships among these clusters. The admixture was determined based on spatial modeling of allele sharing among geographical coordinates of sampling sites. *DAPC* determines ancestry proportions and membership probability modeled on genetic variation across space to determine admixtures as described in (Jombart et al., 2010).

It was important to select the Optimum number of PCS when we use the *DAPC* function. Retaining too many components can lead to over-fitting and instability in the membership probabilities or retaining too a few PCs can exclude usefulinformation from theanalysis. The resultant model may not be informative enough to discriminatebetween groups accurately. Eventually, the optimum PCs for DAPC analysis were chosen using *optim.a.score* function (Jombart *et al*., 2010).

3.6Population Allele Frequency Differentiation Analysis

Analyses of allele frequency distributions between-population F_{ST} values (Weir and Cockerham, 1984) were calculated using *vcftools* or *hierfstat* package from *adegenet* in R software, after excluding SNPs with greater than 10% missing data. For F_{ST} values analysis, missing data were observed and excluded as per the SNP basis, with the size of each population corrected to account for F_{ST} value difference due to population size variation. The F_{ST} value range from 0 to 1, while the zero Fst indicates no population differentiation, while F_{ST} value 1 shows complete population differentiation (Holsinger and Weir, 2009).

3.7Detection of Selection Signatures in *P. falciparum* Populations

Analyses of allele frequency distributions within-population Tajima's D indices (Tajima, 1989) were calculated using vcftools after removing missing data as per the SNP basis. Tajima D values were determined for each SNP, and the average value

was calculated for each gene using the *tapply* function in R software. Genes with at least five SNP and TajimaD values greater than one wereconsidered to be under positive balancing selection.

The standardized integrated haplotype score (IHS) analysis was done to identify positive directional selection signaturesusing SNP data with allele frequency > 5%. IHS was determined using the *rehh* package of R software with default parameters (Gautier & Vitalis, 2012) after imputing missing SNP data using *beagle* version 5.2 (Browning & Browning, 2007). For this IHS, we used |IHS| > 2.5 (top 1% of the expected distribution) (Voight *etal*.,2006) as cut off values to report genes under recent directional selection as reported for genome analysis of West African *P. falciparum*(Oyebola *et al*., 2017).

CHAPTER FOUR

RESULTS

4.1 Sequencing of *P. falciparum* Isolates and SNPs Allele Frequency Distributions

After removing nine samples with low quality sequences, high-quality sequence data obtained from 25 *P. falciparum* clinical isolates collected from West Arsi, Ethiopia enabled the identification of 672956 biallelic SNPs with less than 10% missing isolates and < 5% missing SNPs data in the individual isolate. Allele calls for all isolates were present for 95.95% SNPs. Sequences from the intergenic regions had lower read coverage than those in the coding regions, as expected, because of extreme A + T nucleotide contents in intergenic regions. Thus, 78.92% (531,120) of all SNPs called were located within the genes. Out of 5058 genes analyzed (var, rifin, stevor genes, and subtelomeric regions had been excluded), 3370 genes had at least one SNP (Figure 4.1A). To determine whether inferences from the analyses performed in this study were unique to the population sampled in Ethiopia (west Arsi) or present across East Africa and Southeast Asia (Figure 4.1A), we also analyzed previously sampled data from Malawi (Ocholla et al., 2014), DR Congo, Cambodia and Thailand (Ashley *et al*., 2014**).**

About 18517 SNPs were polymorphic marker in at least one sample in *P. falciparum* (n=25 samples) from west Arsi, Ethiopia of which 43.4 % (8037/18,517) were non-synonymous coding SNPs, 22.8% were synonymous coding SNPs, 26.6 %(4,932/18517) in intergenic regions, 3% (587/18,517) in intragenic regions and 3.5 %(656/18,517) SNPs in intron regions (Figure 4.1B) respectively. Similarly, *P. falciparum* populations from Cambodia (n=46), DR Congo (n=50), Malawi (n=50), and Thailand (n=49) had 32,854, 68,476, 79,250 and 30,427 biallelic polymorphic SNPs marker in at least one sample. Similar to the Ethiopian (west Arsi)*P. falciparum* population, *P. falciparum* from these countries had a high percentage of non-synonymous coding SNPs at the polymorphic marker and consistent with the previous study findings (Oyebola *et al*., 2017), showing all *P. falciparum* populations analyzed here were under different or the same selection pressure.

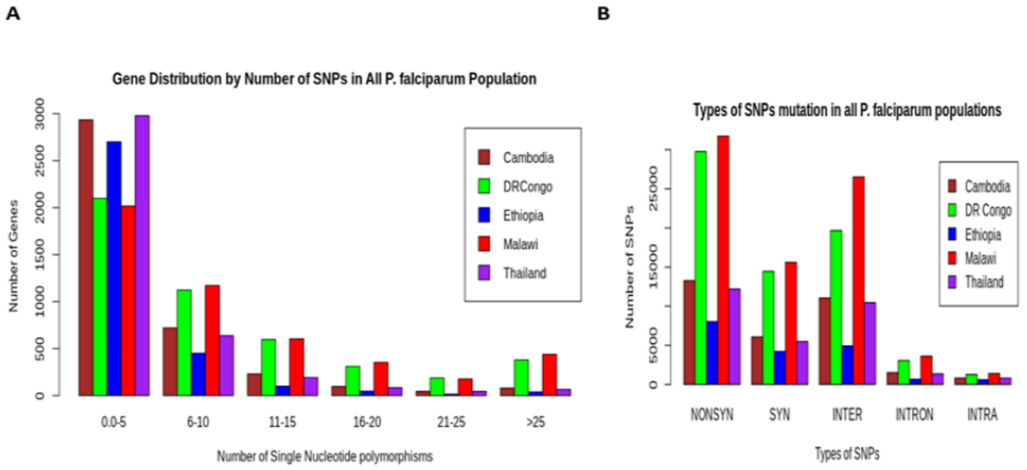


Figure 4.1:SNPs Distribution across Analyzed Genes and Their Respective Distribution of SNPs EffectsA) Distribution of genes (Cambodia; N=4134, DR Congo; N=4722, Ethiopia; N=3370, Malawi; N=4789, Thailand; N=4029)with at least one SNP across the P. falciparum populations whole genome. B) SNPs EffectsDistribution in each P. falciparum population. NONSYN: Non-synonymous, SYN: Synonymous, INTER: Intergenic, INTRON: Intronic and INTRA: other intragenic SNP, respectively. Color code represents P. falciparum population's country of origin

Single nucleotide polymorphisms with minor allele frequency (MAF) < 5% were common in all analyzed *P. falciparum* populations following the exclusion of monomorphic SNPs in each population. Further, SNPs with minor allele's frequency of < 5% occurred more frequently in samples from Malawi than Ethiopian (west Arsi) isolates (Figure 4.2).

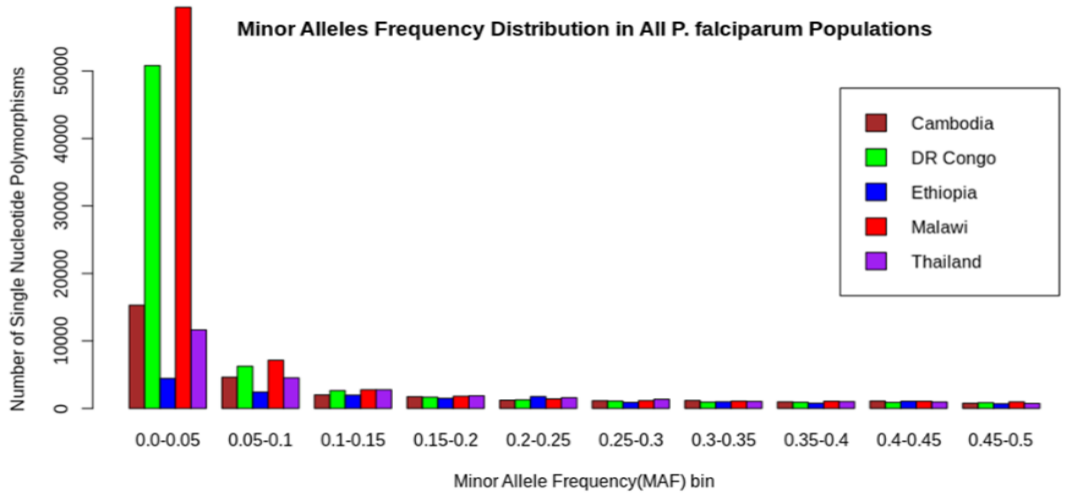


Figure 4.2: Minor Allele Frequency Distribution for Each SNPs Scored in a Population Sample of P. falciparum Clinical Isolates. SNPs Binned into 10 Equal Sizes of 0.05.There is an overabundance of low-frequency SNPs (MAF < 5%) in all parasite populations. Brown, green, blue, red, and purple color code stands for Cambodian, Congolese, Ethiopian, Malawian, and Thailand P. falciparum populations,respectively.

4.2 Genetic Diversity of Ethiopian *P. falciparum* Population

The overall π in the Ethiopian(west Arsi)*P. falciparum* population was 0.00022, which is relatively lower than the genetic diversity in Malawian *P. falciparum* (0.00025) (Ocholla *et al*., 2014). High variability of genetic diversity π values across the chromosomes was observed in Ethiopian (west Arsi)*P. falciparum* samples with a minimum value of 0.00015 in chromosome 12 and maximum value of 0.00045 in chromosome 4 (Figure 4.3), consistent with variable genetic diversity π values observed in different chromosomes of *P. falciparum* population (Volkman *et al.*, 2007).

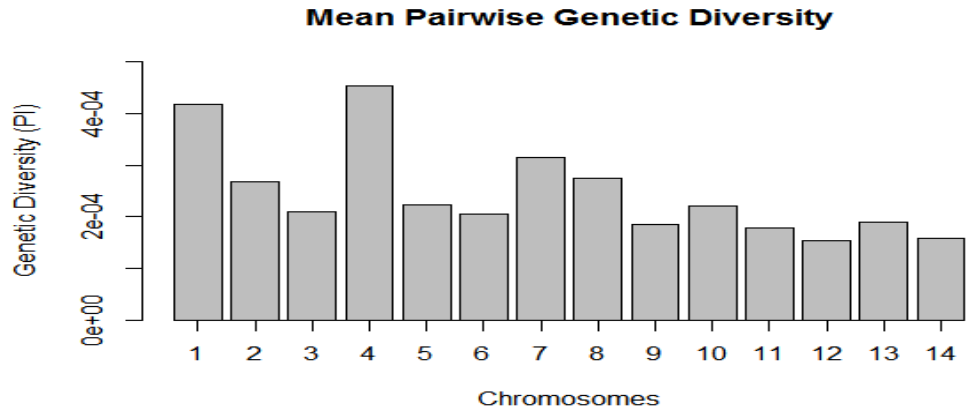


Figure 4.3: Mean Pairwise Nucleotide Diversity in the P. falciparum Populationof westArsi of Ethiopia.The x-axis represents the chromosome list from 1 to 14, and the y-axis shows the mean pairwise genetic diversity in the respective chromosome.

4.3 Genome Mixedness of *P. falciparum* Infections

Within samples, *P. falciparum* diversity was assessed through the FWS fixation index, which summarizes within individual infection diversity relative to the diversity in local *P. falciparum* populations, thus, informing the level of inbreeding within each local population (Manske *et al*., 2012).

FWS scores ranged from 0.837 to 0.997 (mean = 0.97, median = 0.99) for west Arsi, Ethiopian*P. falciparum* population, whereas the FWS values in Cambodia ranged from 0.702 to 0.999 (mean = 0.962, median = 0.995), from 0.483 to 0.998 (mean = 0.94, median = 0.994) in Thailand, from 0.321 to 0.998 (mean 0.94, median 0.994) in DR Congo and from 0.194 to 0.997 (mean = 0.747, median = 0.762) in Malawi (Figure 4.4).

The F_{WS} value of > 0.95 suggests that the individual sample predominantly contains a single genotype and could have other additional genotypes in lower proportions. In this study, F_{WS} values of > 0.95 were observed in 84%, 79.6%, 78%, 50%, and 36 % of samples from west Arsi of Ethiopia, Thailand, Cambodia, DR Congo, and Malawi, respectively.

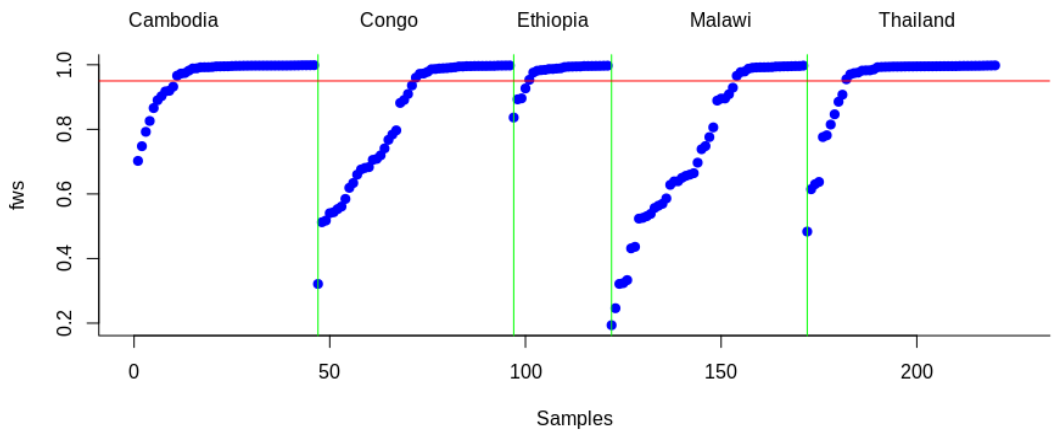


Figure 4.4: Within-Infection FWS Fixation Indices for Each Clinical Isolate Sampled from Cambodia, DR Congo, Ethiopia, Malawi, and Thailand P. falciparum Populations, Ordered by Increasing FWS Index Value within Samples of Each Population.Redline marks FWS = 0.95, above which an isolate is considered a single predominant genotype. Vertical green lines separate the P. falciparum population based on their country of origin. X-Axis label above the figure represents the analyzed P. falciparum populations' country of origin.

The mean F_{WS} scores of the Ethiopian(west Arsi)*P. falciparum* population were not significantly different from those of Cambodia (Welch two Sample t-test, P = 0.42) and Thailand (Welch two-sample t-test, p = 0.083) at 95% confidence intervals. However, it was significantly different from the mean F_{WS} scores of the *P. falciparum* population of DR Congo (Welch two-sample t-test, p = $5.603e^{-06}$) and Malawi (Welch two-sample t-test, p = $3.242e^{-08}$) at 95% confidence intervals.

4.4 Population Structure and Admixtures

PCA, admixture, and FST analysis were performed to investigate major geographical divisions of population structures. The outcome of PCA analysis clearly distinguished four major groups of the parasite isolates that were consistent with their respective geographical origins from which they were sampled (Figure 4.5A-C). Similarly, the findings from the admixture analysis were consistent with the outcome of PCA. The isolates from the three regions were distinguished. This admixture analysis showed that four major components were differentiated according to the optimized cluster value of K = 5. Multiple parasite subpopulations were detected in

parasites from Malawi and DR Congo (Figure 4.6). There was no gene flow between the isolates from Ethiopia, East Africa, and Southeast Asia isolates.

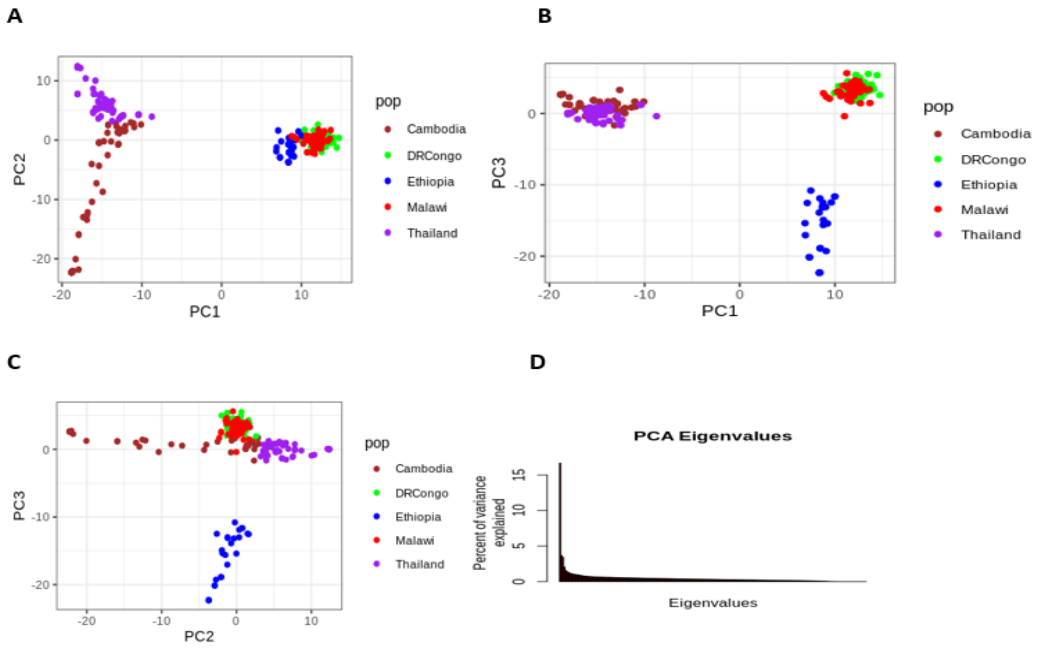


Figure 4.5: Principal Component Analysis.A) Principal component axis1 and axis2 identified Southeast Asian from African P. falciparum population B) Principal component axis1 and axis3 identified Ethiopian (west Arsi), south East Asian and African P. falciparum populations C) Principal Component axis2 and axis3 identified Ethiopian (west Arsi) P. falciparum population from Southeast Asian and African P. falciparum population D) Percent of variance explained by each principal component axis. The first three Principal components axis explained more each Principal component axis. The first three principal component axis explained more variance found in the data. PC1 explained 16.7%, PC2 explained 3.7% and PC3 explained 3.4% of the variance in the analyzed data. Pop stands for the country of origin for each population.

The clustering of Ethiopian(west Arsi) parasite isolates was consistent with the fixation index (F_{ST}) values with or without correcting for sample size. The F_{ST} values of Ethiopian (west Arsi) isolates versus those from the two other east African regions (DR Congo and Malawi) ranged from 0.08 to 0.09 and 0.18 in both Asian regions (Thailand and Cambodia) respectively (Table 4.1), which suggested population differentiation between Ethiopia(west Arsi) and East Africa as well as East Africa and Southeast Asia regions.

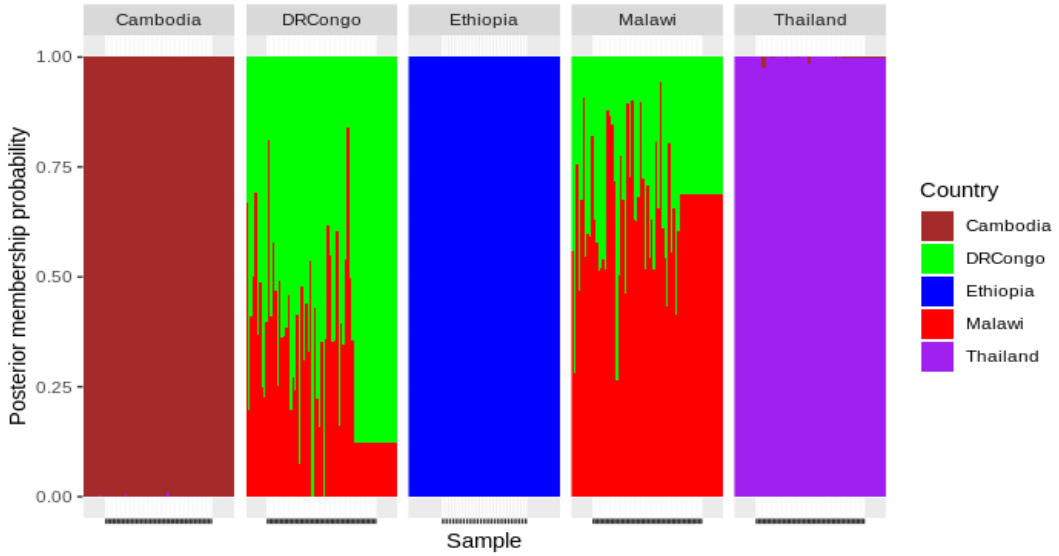


Figure 4.6: Analysis of an Admixture.X-axis represents the assigned samples to country listed above the figure after modeling allele sharing probability, while Y-axis represents the membership probability of the samples.Color coded country listed at the right of the figure indicates the origin of P. falciparum samples. Analysis of admixture identified four major components based on an optimized cluster value of K = 5, and multiple subpopulations of P. falciparum were found in DR Congo and Malawi. There was a bidirectional gene flow between Malawian and Congolese P. falciparum populations.

4.5 Signatures of Selection in the Ethiopian P. falciparum Isolates

We found that the average Tajima's D value of the Ethiopian (west Arsi) isolates was 0.18 across the entire genome (One sample t-test, $p < 2x10^{-16}$). About 1450 genes had at least one SNP with TajimaD value > 1, out of which 125 genes had at least five SNPs with Tajima D values > 1 (Appendix II), suggesting that the genes were most likely under balancing selection. These genes include *ama1, eba175, msp1, trap, dblmsp*, and *clag2*, previously reported for being under balancing selection (Amambua-Ngwa et al., 2012; Ye et al, 2019).

The standardized integrated haplotype homozygosity score (IHS) was employed to investigate genome-wide evidence for recent positive directional selection due to drug pressure, the impact of the immune response, or other mechanisms. Using |IHS| score of > 2.5 (top 1% of the expected distribution) as a threshold hits, we identified 36 genes with at least one SNP that could be under significant positive selection and out of these, and 15 genes had at least two SNPs (Table 4.2 and Figure 4.7).

Table 4.1: Pairwise *Plasmodium falciparum* Population Divergence Measured by FST.Ethiopian*P. falciparum* population of west Arsi was highly diverged from both Southeast Asian and East African *P. falciparum* populations.

Country	Cambodia	DRCongo	Ethiopia	Malawi	Thailand
Cambodia	0	0.16	0.18	0.17	0.06
DR Congo	0.16	0	0.08	0.02	0.16
Ethiopia	0.18	0.08	0	0.09	0.18
Malawi	0.17	0.02	0.09	0	0.17
Thailand	0.06	0.16	0.18	0.17	0

Thirteen (13) out of above 15 genes under positive directional selection showed both positive balancing and directional selections (Table 4.2 and Table 4.3) and these genes include vaccine candidate gene *SURF4.2* on chromosome 4 and cytoadherence linked asexual protein 8 (*CLAG8*) on chromosome 8 (Iriko *et al*., 2008).Our study did not detect the selection signals in drug-resistance that includes genes such as *Pfcrt, Pfmdr1, Pfdhfr,* and*Pfdhps* possibly because the IHS may not be suitable for detecting positive selection for those SNPs that have reached or are near to fixation in the local *P. falciparum* population (Voight *et al*., 2006).

4.6 Prevalence of Drug Desistance-Conferring Mutations in P. *falciparum*

Inter-population drug resistance-conferring allele frequency difference was observed among P. falciparum populations from Ethiopia (west Arsi), Cambodia, DR Congo, Malawi, and Thailand. This finding was consistent with the previous reports (Ocholla et al., 2014; Verity et al., 2020), showing the evidence that different drug resistance-conferring mutation distribution in various geographical regions. For instance, chloroquine resistance-conferring alleles (Pfcrt-K76T, Pfcrt-A220S, Pfcrt- Q271E, and Pfcrt-N326S) were fixed in Ethiopia (west Arsi), Cambodia, and Thailand. However, about 66 to 72% of P. falciparum from DR Congo contained chloroquine resistance alleles, whereas no chloroquine resistance markers were absent in Malawi samples (Table 4.4).

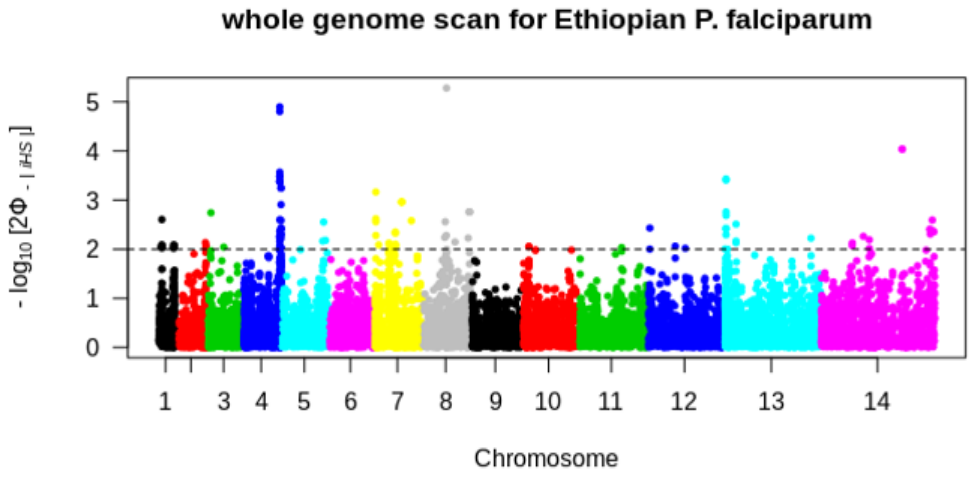


Figure 4.7:Recent Directional Selection in P. falciparum Population of west Arsiof Ethiopia. The dashed line indicates genome-wide integrated haplotype scores at a significance threshold of P< 0.01.

A fixed allele's frequency at *Pfcrt*-K76T mutation was previously reported in Ethiopian(west Arsi)*P. falciparum* (Golassa *et al*., 2014). Similarly, drug resistance mutations in *pfmdr1* (*pfmdr1-N86Y* and *pfmdr1-Y184F*) were also variable among populations. For instance, the Ethiopian(west Arsi) parasite population showed 14% *Pfmdr1-N86Y* and 100% *Pfmdr1-Y184F* gene mutations, whereas *Pfmdr1-N86Y*was detected in 48% of DR Congo isolates and 3% of Malawi's. In addition, the *pfmdr1*-Y184F drug resistance marker was detected in 58% of the *P. falciparum* population in Cambodia, 32% in DR Congo, 35% in Malawi, and 6% in Thailand's parasite isolates.

The major pyrimethamine resistance-conferring alleles such as *Pfdhfr-N51I* and *Pfdhfr-C59R* were also identified in all parasite populations with fixed or near fixation in frequency. *Pfdhfr-S108N* was fixed in other *P. falciparum* populations, except in Ethiopia's (west Arsi).

Table 4.2: Genes Under a Recent Positive Directional Selection in *P. falciparum* of west Arsi of Ethiopia, Identified Using the Integrated Haplotype Score at a Significance Threshold of P< 0.01.SNPs and ID stand for single nucleotide polymorphisms and gene identification numbers, respectively.

Chromosome	Number of SNPs	Genes Name/ ID	Product description
1	9	PF3D7_0104100	conserved *Plasmodium* membrane protein, unknown function
1	3	PF3D7_0113600	surface-associated interspersed protein 1.2 (SURFIN 1.2)
4	2	PF3D7_0424300	erythrocyte binding antigen-165
4	14	SURF4.2	surface-associated interspersed protein 4.2 (SURFIN 4.2)
4	5	PF3D7_0425200	*Plasmodium* exported protein (hyp15), unknown function
4	4	PF3D7_0425250	*Plasmodium* exported protein (PHIST), unknown function
7	5	PF3D7_0713900	conserved *Plasmodium* protein, unknown function
7	3	CRMP2	cysteine repeat modular protein 2
8	3	CLAG8	cytoadherence linked asexual protein 8
10	2	PF3D7_1004800	ADP/ATP carrier protein, putative

12	2	PF3D7_1201400	*Plasmodium* exported protein, unknown function
13	7	PF3D7_1301800	surface-associated interspersed protein 13.1 (SURFIN 13.1)
13	3	PF3D7_1308400	conserved *Plasmodium* protein, unknown function
14	2	PF3D7_1434500	dynein-related AAA-type ATPase, putative
14	2	PF3D7_1477500	*Plasmodium* exported protein (PHISTb), unknown function

We also found drug resistance-conferring alleles in *Pfdhps* (*Pfdhps-S436A, Pfdhps-G437A*, *Pfdhps-K540N*, and *Pfdhps-A581G*) for the parent drug, sulfadoxine.

The non-synonymous mutation was observed in the *Pfkelch13* gene, which is a known gene in the artemisinin drug response. African population-specific *Pfk13*-K189T mutation was observed in west Arsi ofEthiopia (in 20% of the samples), DR Congo (17%), and Malawi (3%), but not observed in Southeast Asian isolates.This mutation was previously identified in African *P. falciparum* populations (Ocholla *et al*., 2014; Verity *et al.*, 2020). As previously reported (Ashley *et al*., 2014), the known artemisinin resistance-conferring mutation *PfK13*-C580Y was identified in Cambodia (36% of the samples) and Thailand (26%) (Table 4.4).

Table 4.3: Genes Under Both Positive Balancing and Directional Selections in *P. falciparum* Population from west Arsi of Ethiopia.ID stands for a gene identification number.

Chromosome	Gene Name/ ID	Product Description
1	PF3D7_0104100	conserved *Plasmodium* membrane protein, unknown function
1	PF3D7_0113600	surface-associated interspersed protein 1.2 (SURFIN 1.2)
4	PF3D7_0424300	erythrocyte binding antigen-165, pseudogene
4	SURF4.2	surface-associated interspersed protein 4.2 (SURFIN 4.2)
4	PF3D7_0425200	*Plasmodium* exported protein (hyp15), unknown function
4	PF3D7_0425250	*Plasmodium* exported protein (PHIST), unknown function
7	PF3D7_0713900	conserved Plasmodium protein, unknown function
8	CLAG8	cytoadherence linked asexual protein 8

10	PF3D7_1004800	ADP/ATP carrier protein, putative
12	PF3D7_1201400	*Plasmodium* exported protein, unknown function
13	PF3D7_1301800	surface-associated interspersed protein 13.1 (SURFIN 13.1)
13	PF3D7_1308400	conserved Plasmodium protein, unknown function
14	PF3D7_1434500	dynein-related AAA-type ATPase, putative

Table 4.4: Alleles Frequency of SNPs Responsible for Drug Resistance in 5 Populations of *P. falciparum*

Genes	**Chr**	**Position**	**mutation site**	**Ethiopia**	**Cambodia**	**DRCongo**	**Malawi**	**Thailand**
DHFR	4	748577	I164L	0	0.5	0	0	0.84
DHFR	4	748410	S108N	0	1	1	1	1
DHFR	4	748262	C59R	0.86	1	0.86	0.99	1
DHFR	4	748239	N51I	1	0.95	1	1	0.95
MDR1	5	961625	D1246Y	0	0	0.17	0	0
MDR1	5	958145	N86Y	0.14	0	0.48	0.03	0
MDR1	5	961566	F1226Y	0	0.04	0	0	0.59
MDR1	5	958440	Y184F	1	0.58	0.32	0.35	0.06
CRT	7	405600	I356T	0	0.52	0.27	0	1
CRT	7	405362	N326S	0.98	0.51	0	0	1
CRT	7	405838	R371I	0	0.8	0.71	0	1
CRT	7	404407	A220S	1	1	0.663	0	1
CRT	7	403625	K76T	1	1	0.66	0	1
CRT	7	404836	Q271E	1	1	0.7143	0	1
DHPS	8	549685	G437A	0.08	0.13	0.08	0.01	0
DHPS	8	549995	K540N	0	0.4	0	0	0.03
DHPS	8	549681	S436A	0	0.2	0.11	0.02	0.17
DHPS	8	550117	A581G	0.02	0.4	0.03	0.02	0.82
K13	13	1726432	K189T	0.2	0	0.17	0.13	0
K13	13	1725259	C580Y	0	0.36	0	0	0.26

CHAPTER FIVE

DISCUSSION

This study aimed to analyze the whole genome sequence of *P. falciparum* from diverse geographical origins to identify valuable genes and genetic variation in the *P. falciparum* parasite populations across the malaria-endemic regions in the world. This process enabled the identification of *P. falciparum* genome sequence variation, population structure, admixture, selection signature, and adaptation, undertaken in response to variable environmental stimuli in the different *P. falciparum* malaria endemic geographical regions (Manske *et al*., 2012). These malaria endemic regions include Africa (Amambua-Ngwa *et al*., 2019; Verity *et al*., 2020), where *P. falciparum* transmission intensity is high and causes significant loss of life, especially in children under the age of five that occurs each year (WHO, 2017). In addition, these efforts also identified different mutation and selection signatures in drug target genes, including *Pfcrt, Pfdhps, Pfdhfr, Pfmdr1* (Duffy *et al*., 2015; Ravenhall *et al*., 2016; Samad et al., 2015), and *Pfkelch-13*(Ashley et al., 2014; Sibley, 2015) and immune target genes such as *ama1, eba175, trap, msp1, dblmsp,* and others (Amambua-Ngwa, Park, *et al*., 2012; Ye *et al*., 2019).

The whole-genome analysis also enabled the identification of *P. falciparum* population origin, evolutionary relatedness of parasites in different geographical locations, allele sharing probability, and gene flow intensity among *P. falciparum* populations, which is crucial in the containment and prevention of *P. falciparum* malaria diseases.

Many studies have endeavored to understand the prevalence of drug resistance mutations in Ethiopian *P. falciparum* isolates (Golassa *et al*., 2014; Hailemeskel et al., 2013; Heuchert et al., 2015; Jalei, Chaijaroenkul, & Na-bangchang, 2018; Lo et al., 2017; Mekonnen et al., 2014).None of them used whole-genome sequencing and characterized these deadly malaria parasites and compared it to other East African and Southeast Asian *P. falciparum* populations to understand the evolutionary relationship and allele sharing probabilities.Understanding the evolutionary

relationship and allele sharing probabilities of population may play a key role developingspecific malaria treatment and prevention methods. Here, we sequenced 34 *P. falciparum* clinical isolates from the west Arsi region of Ethiopia intended to determine genetic variation, population structure, selection signatures, and its genetic relatedness to other East African (DR Congo and Malawi) and Southeast Asian (Cambodia and Thailand) *P. falciparum* populations.

5.1 Genetic Diversity in Ethiopian *P. falciparum* Population

We found that overall the pairwise genetic diversity (π) in the Ethiopian (west Arsi)*P. falciparum* population was 0.00022, which was relatively lower than genetic diversity (π) in the Malawian *P. falciparum* population, probably due to the difference in parasites transmission intensity between the two countries. *Plasmodium falciparum* malaria transmission occurs throughout the year in Malawi (Ocholla et al., 2014). In contrast, malaria transmission is seasonal and low (*P. falciparum* prevalence = 0.03) in the Ethiopian region(west Arsi) where the *P. falciparum* samples were collected (Wolde and Bikila, 2016). We also observed pairwise genetic diversity (π) difference in chromosomes of Ethiopian *P. falciparum* parasites with the maximum and minimum value in chromosome 4 and 12 respectively (Figure 4.3).Genetic diversity variation across chromosomes could be due to different recombination activities acting on *P. falciparum* chromosomes, as observed in the previous studies (Ocholla *et al*., 2014; Volkman *et al.*, 2007).

5.2 The Multiplicity of Infections (Genome Mixedness) in *P. falciparum* Population

F_{WS} (parasite inbreeding coefficient), which is inversely related to the multiplicity of infection or genome mixedness, was used to estimate the relative number of the genotypically distinct parasite within individual samples.We observed a significant difference in F_{WS} values between regions. High values of F_{WS}> 0.95 were much common in Ethiopia, Thailand, and Cambodia (84%, 79.6%, and 78% of samples) than DR Congo and Malawi (50% and 36 % of samples) (Figure 4.4 and Appendix I). The high level of F_{WS}> 0.95 observed in Ethiopia(west Arsi), Thailand, and Cambodia could be due to the low *P. falciparum* malaria transmission intensity

within the countries. High transmission intensity can also lead to high F_{WS}> 0.95 if *P. falciparum* circulates in the geographically isolated community, which limits the chance of outcrossing with other genetically distinct *P. falciparum* parasites as observed in the previous study (Manske *et al*., 2012). In Ethiopia(west Arsi), the high level of F_{WS}> 0.95 was most likely due to the low transmission of *P. falciparum* in the study area (Wolde et al., 2016). In contrast, the low F_{WS}> 0.95 in DR Congo and Malawi was probably due to the high *P. falciparum* malaria transmission intensity in both countries (Ocholla *et al*., 2014; Verity *et al*., 2020). These findings were consistent with the study finding from West Africa (Duffy *et al*., 2015) that a low percentage of samples with a level of F_{WS}> 0.95 in the *P. falciparum* population of high transmission region.

5.3 Population Structure Analysis

Population structure within and between continents was evident from principal components (Figure 4.5), admixture (Figure 4.6), and F_{ST} analysis (Table 4.1). All of these methods indicated a higher degree of population structure in Southeast Asia as well as in East Africa. Cambodia's and Thailand's samples form clear separate clusters. Similarly, Ethiopian samples and other East African samples showed separate clusters, while samples from DR Congo and Malawi were intermixed. These population structure variations observed here are probably due to malaria transmission differences within and between the continents. Consistent with previous evidence that parasite population structure was high in regions of patchy or low malaria transmission (Manske *et al*., 2012) and low in the area of high malaria transmission (Amambua-Ngwa *et al*., 2019).

5.4 Selection Signatures in Ethiopian *P. falciparum* Population

Allele frequency spectrum analysis identified many genes with Tajima D value greater than one in Ethiopian *P. falciparum*. These genes include known vaccine candidates such as *ama1, trap, msp1, eba175,* and *clag2* (Appendix II) which were previously identified in the different populations from different malaria-endemic regions (Amambua-Ngwa, Park, et al., 2012; Amambua-Ngwa, Tetteh, et al., 2012; Ye et al., 2019), indicating balancing selection were operating in a similar subset of

genes. We also identified genes under both balancing and directional selections (complex selection) which includes *SURFIN 4.2, CLAG8, and PHIST* families and estimated to be targets of immunity as previously observed (Amambua-Ngwa, Park, et al., 2012) indicating immune target genes are under complex selections.

Haplotype analysis with integrated haplotype score identified 15 genes being under recent positive directional selections that include *SURFIN* and *PHIST* families that require further study. This analysis failed to detect selection signatures in known drug target genes such as *Pfcrt, Pfmdr1, Pfdhfr, Pfdhps,* and *Pfkelch-13*. The reason could be that the standardized integrated haplotype score (IHS) may not be suitable for detecting positive selection for those SNPs that have reached or near fixation in the local *P. falciparum* population. However, analyzing drug resistance-conferring mutations prevalence in each known drug target of *P. falciparum* population showed that drug resistance-conferring SNPS in *Pfcrt, Pfmdr1,* and *Pfdhfr* were already fixed or near to fixation in Ethiopian (west Arsi)*P. falciparum* (Table 4.4*)*. This result was consistent with the previous study finding (Golassa et al., 2014) that chloroquine resistance-conferring SNP was already fixed in west Arsi, Ethiopia. After comparing Ethiopian (west Arsi)*P. falciparum* drug resistance SNPs (*Pfcrt*-K76T, *Pfcrt*-A220S, *Pfcrt*- Q271E, and *Pfcrt*-N326S), we noted chloroquine resistance SNPs were fixed in both Cambodia and Thailand, cover from 66-70% isolates in DR Congo and not found in Malawi.

Similarly, Pyrimethamine resistance SNPs in *Pfdhfr*(*Pfdhfr*-N51I and *Pfdhfr*-C59R) were almost fixed in all populations. But, drug resistance SNPs in *Pfdhps* genes (*Pfdhps*-S436A, *Pfdhps*-G437A, *Pfdhps*-K540N, and *Pfdhps*-A581G) were account about less than 18% in all population while SNPs in *Pfmdr1* was fixed in Ethiopia(west Arsi) and less than 60% in the other populations. The high prevalence of drug resistance mutation in *Pfcrt, Pfmdr1,* and *Pfdhfr*of Ethiopian (west Arsi)*P. falciparum* informed that chloroquine and SP are not yet to be effective in *P. falciparum* malaria treatment. Chloroquine has been used to treat *P. vivax* malaria and resistance SNP markers in *Pfdhfr*against the pyrimethamine were already fixed to evolve through the Ethiopian (west Arsi)*P. falciparum* population.

We also found *PfK13* confirmed drug resistance-conferring *PfK13*-C580Y SNP in about 36% and 26% of Cambodia and Thailand samples respectively, but not in all African samples. Similarly, we identified African specific *PfK13*-K189T mutation in Ethiopia(west Arsi) (20% of samples), DR Congo (17%), and Malawi (13% samples), and its role in artemisinin resistance is unknown. These findings were consistent with the previous study findings (Verity et al., 2020). However, (Ashley et al., 2014) reported that mutation in *PfK13* at the position of amino acid less than 441is not related to artemisinin resistance.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

The whole-genome analysis of the Ethiopian (west Arsi)*P. falciparum* population revealed the low overall genetic diversity ($\pi = 0.00022$) across the 14 chromosomes of 25 clinical isolates. A high percentage of samples (84%) had FWS > 0.95 (low multiplicity of infection), which was consistent with low malaria transmission of the study area. Further analysis ofthe population structure, admixture, and population allele frequency differentiation (F_{ST}) showed that the *P. falciparum* population from west Arsi of Ethiopia was separated from other East Africa and Southeast Asian populations. No gene flow between *P. falciparum*from west Arsi of Ethiopia and other population indicatethat *P. falciparum*population in west Arsi of Ethiopia is independently evolving and circulating in the area. We also identified positive balancing selection in 125 genes of parasites in west Arsi, including known vaccine target *ama1, eba175, msp1*and *trap,* and directional selection in 15 genes, of which 13 genes were under complex selection. However, we did not identify any selection in *Pfcrt*, *Pfmdr1, Pfdhfr, Pfdhps,* and *Pfkelch13*. However, drug resistance-conferring allele frequency analysis showed that at least one allele was fixed in *Pfcrt*, *Pfmdr1, Pfdhfr,*suggesting that the Ethiopian *P. falciparum* is not yet sensitive to chloroquine and SP drugs despite the withdrawal of these drugs before 2004. African specific *PfK13-K189T* mutation was also observed in Ethiopian *P. falciparum,* and its role in artemisinin resistance was unknown.

6.2 Recommendation

Our study showed that Chloroquine and Pyrimethamine resistance were fixed in Ethiopia. We recommend the continuation of *P. falciparum* malaria treatment with artemisinin combined therapy with careful molecular surveillance. Further molecular studies involving deeper sampling of Ethiopian parasite populations are essential to understand the genetic diversity, gene flow, and temporal evolution of

drug resistance loci within Ethiopia. Our findings can be used to support national malaria control decision making for optimal impact in further reducing malaria transmission in Ethiopia.

REFERENCES

Alene, G.D. and Bennett, S. (2007). Chloroquine resistance of *Plasmodium falciparum* malaria in Ethiopia and Eritrea. *Tropical Medicine & International Health, 1*(6), 810–815. https://doi.org/10.1111/j.1365-3156.1996.tb00115.x

Amambua-Ngwa, A., Amenga-Etego, L., Kamau, E., Amato, R., Ghansah, A., Golassa, L., … Djimde, A.A. (2019). Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. *Science, 365*(6455),813–816. https://doi.org/10.1126/science.aav5427

Amambua-Ngwa, A., Jeffries, D., Amato, R., Worwui, A., Karim, M., Ceesay, S., … D'Alessandro, U.(2018). Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from the Gambia. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-28017-5

Amambua-Ngwa, A., Park, D.J., Volkman, S.K., Barnes, K.G., Bei, A.K., Lukens, A.K., … Schaffner, S.F.(2012). SNP genotyping identifies new signatures of selection in a deep sample of west African *Plasmodium falciparum* malaria parasites. *Molecular Biology and Evolution*, *29*(11), 3249–3253. https://doi.org/10.1093/molbev/mss151

Amambua-Ngwa, A., Tetteh, K.K.A., Manske, M., Gomez-Escobar, N., Stewart, L.B., Deerhake, M.E., … Conway, D.J. (2012). Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in

malaria parasites. *PLoS Genetics*, *8*(11), 1–14. https://doi.org/10.1371/journal.pgen.1002992

Ashley, E., Dhorda, M., Fairhurst,R.M., Amaratunga, C., Lim, P., Suon, S., … Dondorp,A.M., (2014). Spread of Artemisinin Resistance in *Plasmodium falciparum* Malaria. *The New Engl and Journal of Medicine*, *371*(5), 411–423. https://doi.org/10.1056/NEJMoa1314981

Auburn, S., Campino, S., Clark, T.G., Djimde, A.A., Zongo, I., Pinches, R., … Kwiatkowski, D.P. (2011). An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole-genome high-throughput sequencing. *PLoS ONE*, *6*(7), 4–11. https://doi.org/10.1371/journal.pone.0022213

Auwera, G.A. Van Der, C.M.O., Hartl, C., Poplin, R., Levy-moonshine, A., Jordan, T., … Depristo, M.A. (2014). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, *11*(1110). https://doi.org/10.1002/0471250953.bi1110s43

Bemmels, J.B., Dick, C.W., Nazareno, A.G. and Lohmann, I. A. G. (2017). Minimum sample sizes for population genomics : an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, 17, 1136–1147. https://doi.org/10.1111/1755-0998.12654

Biswas, S. and Akey, J.M. (2006). Genomic insights into positive selection. *Genetics*, *22*(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Browning, S.R. and Browning, B.L. (2007). Rapid and accurate haplotype phasing

and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

CDC. (2018a). Malaria Diagnosis.from

https://www.cdc.gov/malaria/diagnosis_treatment/diagnosis.html

CDC. (2018b). Malaria lifecycle. Retrieved from

https://www.cdc.gov/malaria/about/biology/index.html

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Duffy, C.W., Assefa, S.A., Abugri, J., Amoako, N., Owusu-Agyei, S., Anyorigiya, T., … Awandare, G.A. (2015). Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. *BMC Genomics*, *16*(1), 1–11. https://doi.org/10.1186/s12864-015-1746-3

Flesch, E.P., Garrott, R.A., Rotella, J.J., Thomson, J.M., & Graves, T.A. (2018). Evaluating sample size to estimate genetic management metrics in the genomics era. *Molecular Ecology Resources*, 1–15. https://doi.org/10.1111/1755-0998.12898

FMoH. (2011). National five-year strategic plan for malaria prevention & control in Ethiopia. Retrieved from

https://extranet.who.int/countryplanningcycles/sites/default/files/planning_cycle

_repository/ethiopia/five_year_malaria_control_sp_2006_2010.pdf

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., … Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, *419*(6906), 498–511. https://doi.org/10.1038/nature01097

Gautier, M., and Vitalis, R. (2012). Rehh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure, *28*(8), 1176–1177. https://doi.org/10.1093/bioinformatics/bts115

Golassa, L., Enweji, N., Erko, B., Aseffa, A., and Swedberg, G. (2014). High prevalence of pfcrt-CVIET haplotype in isolates from asymptomatic and symptomatic patients in south-central Oromia, Ethiopia. *Malaria Journal*, *13*(120), 1–8. https://doi.org/10.1186/1475-2875-13-120

Hailemeskel, E., Kassa, M., Taddesse, G., Mohammed, H., Woyessa, A., Tasew, G., … Petros, B. (2013). Prevalence of sulfadoxine-pyrimethamine resistance-associated mutations in dhfr and dhps genes of *Plasmodium falciparum* three years after SP withdrawal in Bahir Dar, Northwest Ethiopia. *Acta Tropica*, *128*(3), 636–641. https://doi.org/10.1016/j.actatropica.2013.09.010

Hale, M.L., Burg, T.M., and Steeves, T.E. (2012). Sampling for Microsatellite-Based Population Genetic Studies : 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. *PLoS ONE*, *7*(9), 1–10. https://doi.org/10.1371/journal.pone.0045170

Heuchert, A., Abduselam, N., Zeynudin, A., Eshetu, T., Löscher, T., Wieser, A., … Berens-Riha, N. (2015). Molecular markers of anti-malarial drug resistance in southwest Ethiopia over time: Regional surveillance from 2006 to 2013. *Malaria Journal*, *14*(1), 1–7. https://doi.org/10.1186/s12936-015-0723-2

Holsinger, K.E., and Weir, B.S. (2015). Genetics in geographically structured populations: defining, estimating, and interpreting FST. *Nature Genetics*, *10*(9), 639–650. https://doi.org/10.1038/nrg2611.Genetics

Iriko, H., Kaneko, O., Otsuki, H., Tsuboi, T., Su, X.Z, Tanabe, K., and Torii, M. (2008). Diversity and evolution of the rhoph1/clag multigene family of *Plasmodium falciparum.Molecular and Biochemical Parasitology*, *158*(1), 11–21. https://doi.org/10.1016/j.molbiopara.2007.11.004

Jalei, A.A., Chaijaroenkul, W., and Na-bangchang, K. (2018).*Plasmodium falciparum* drug resistance gene status in the Horn of Africa : A systematic review. *African Journal of Pharmacy and Pharmacology*, *12*(25), 361–373. https://doi.org/10.5897/AJPP2018.4942

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components : a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(94), 1–15.

Kidgell, C., Volkman, S.K., Daily, J., Borevitz, J.O., Plouffe, D., Zhou, Y., … Winzeler, E.A. (2006). Erratum: A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathogens*, *2*(6), 0562–0577. https://doi.org/10.1371/journal.ppat.0020057

Lever, J., Krzywinski, M., and Altman, N. (2017). Principal component analysis. *Nature*, *14*(7), 1–2. https://doi.org/10.1038/nmeth.4346

Lin, J.T. (2011). Drug-Resistant Malaria : The Era of ACT. *Current Infectious Diseases Reports*, *12(*3), 165–173. https://doi.org/10.1007/s11908-010-0099-y

Lo, E., Hemming-Schroeder, E., Yewhalaw, D., Nguyen, J., Kebede, E., Zemene, E., … Yan, G. (2017). Transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum* in Ethiopia and the prevalence of antimalarial resistant genotypes. *PLoS Neglected Tropical Diseases*, *11*(7), 1–25. https://doi.org/10.1371/journal.pntd.0005806

Mackinnon, M.J., and Read, A.F. (2004). Virulence in malaria: An evolutionary viewpoint. *Philosophical Transactionsof the Royal. Society B, 359*(1446), 965–986. https://doi.org/10.1098/rstb.2003.1414

Malaria Site. (2018). Malaria lifecycle. Retrieved April 20, 2019, from https://www.malariasite.com/life-cycle/

Manske, M., Miotto, O., Campino, S., Auburn, S., Zongo, I., Ouedraogo, J., … Amenga-etego, L. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*, *487(*7407), 1–19. https://doi.org/10.1038/nature11174

Wiser,M.F. (2000). *Plasmodium* Life Cycle. Retrieved April 20, 2019, from https://www.tulane.edu/~wiser/malaria/mal_lc.PDF

Mekonnen, S.K., Aseffa, A., Berhe, N., Teklehaymanot, T., Clouse, R.M., Gebru, T., … Velavan, T.P. (2014). Return of chloroquine-sensitive *Plasmodium falciparum*

parasites and the emergence of chloroquine-resistant *Plasmodium vivax* in Ethiopia. *Malaria Journal*, *13*(1), 1–9. https://doi.org/10.1186/1475-2875-13-244

Mobegi, V.A., Duffy, C.W., Amambua-Ngwa, A., Loua, K.M., Laman, E., Nwakanma, D.C., … Conway, D.J. (2014). Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Molecular Biology and Evolution*, *31*(6), 1490–1499. https://doi.org/10.1093/molbev/msu106

Nei, M. and Li, W.H. (1979). A mathematical model for studying genetic variation in terms of restriction endonucleases. *Genetics*, *76*(10), 5269–5273.

Ocholla, H., Preston, M.D., Mipando, M., Jensen, A.T.R., Campino, S., Macinnis, B., … Montgomery, J. (2014). Whole-genome scans provide evidence of adaptive evolution in Malawian *Plasmodium falciparum* isolates. *Journal of Infectious Diseases*, *210*(12), 1991–2000. https://doi.org/10.1093/infdis/jiu349

Oyebola, K.M., Idowu, E.T., Olukosi, Y.A., Awolola, T.S. and Amambua-Ngwa, A. (2017). Pooled-DNA sequencing identifies genomic regions of selection in Nigerian isolates of *Plasmodium falciparum*. *Parasites and Vectors*, *10*(1), 1–7. https://doi.org/10.1186/s13071-017-2260-z

Payne, D. (1987). Spread of chloroquine resistance in *Plasmodium falciparum*. *Parasitology Today*, 3(8), 241–246. https://doi.org/10.1016/0169-4758(87)90147-5

PMI. (2019). President's malaria initiative Ethiopia*:* Malaria Operational Plan.

Pourhoseingholi, M.A., Vahedi, M. and Rahimzadeh, M. (2013). Sample size calculation in medical studies. *Gastroenterol Hepatol Bed Bench*, *6*(1), 14–17.

Ravenhall, M., Benavente, E.D., Mipando, M., Jensen, A.T.R., Sutherland, C.J., Roper, C., … Clark, T.G. (2016). Characterizing the impact of sustained sulfadoxine/pyrimethamine use upon the *Plasmodium falciparum* population in Malawi. *Malaria Journal*, 1–11. https://doi.org/10.1186/s12936-016-1634-6

Samad, H., Coll, F., Preston, M.D., Ocholla, H., Fairhurst, R.M. and Clark, T.G. (2015). Imputation-Based Population Genetics Analysis of *Plasmodium falciparum* Malaria Parasites. *PLoS Genetics*, *11*(4), 1–21. https://doi.org/10.1371/journal.pgen.1005131

Schunk, M., Kumma, W.P., Miranda, I.B., Osman, M.E., Roewer, S., Alano, A., … Mockenhaupt, F.P. (2006). High prevalence of drug-resistance mutations in *Plasmodium falciparum* and *Plasmodium vivax* in southern Ethiopia. *Malaria Journal*, *5*, 1–5. https://doi.org/10.1186/1475-2875-5-54

Seyoum, D., Speybroeck, N., Duchateau, L., Brandt, P. and Rosas-Aguirre, A. (2017). Long-lasting insecticide net ownership, access, and use in southwest Ethiopia: A community-based cross-sectional study. *International Journal of Environmental Research and Public Health*, *14*(11), 1–16. https://doi.org/10.3390/ijerph14111312

Sibley, C.H. (2015). Observing in real-time the evolution of artemisinin resistance in *Plasmodium falciparum*. *BMC Medicine*, *13*(1), 5–8. https://doi.org/10.1186/s12916-015-0316-5

Simonetti, A.B. (1996). The biology of malarial parasite in the mosquito - A review. *Memorias Do Instituto Oswaldo Cruz*, *91*(5), 519–541. https://doi.org/10.1590/S0074-02761996000500001

Taffese, H.S., Hemming-Schroeder, E., Koepfli, C., Tesfaye, G., Lee, M.C., Kazura, J., … Zhou, G.F. (2018). Malaria epidemiology and interventions in Ethiopia from 2001 to 2016. *Infectious Diseases of Poverty*, *7*(1), 1–9. https://doi.org/10.1186/s40249-018-0487-3

Tajima, F. (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/PMC1203831

Trask, J.A.S., Malhi, R.S., Kanthaswamy, S., Johnson, J., Garnica, W.T., Malladi, V.S. and Smith, D.G. (2011). The effect of the SNP discovery method and sample size on the estimation of population genetic data for Chinese and Indian *rhesus macaques* (*Macaca mulatta*). *Springer*, 52, 129–138. https://doi.org/10.1007/s10329-010-0232-4

Uwimana, A., Legrand, E., Stokes, B.H., Ndikumana, J.M., Warsame, M., Umulisa, N., … Menard, D. (2020). Emergence and clonal expansion of in vitro *kelch13 R561H* mutant parasites in Rwanda. *Nature Medicine*, 1–14. https://doi.org/10.1038/s41591-020-1005-2

Verity, R., Aydemir, O., Brazeau, N.F., Watson, O.J., Hathaway, N.J., Mwandagalirwa, M.K., … Juliano, J.J. (2020). The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the Democratic Republic of Congo. *Nature Communications*, *11*(2107), 1–10.

https://doi.org/10.1038/s41467-020-15779-8

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in thehuman genome. *Plos Biology*, *4*(3), 0446–0458. https://doi.org/10.1371/journal.pbio.0040072

Volkman, S.K., Sabeti, P.C., Decaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., … Wirth, D.F. (2007). A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genetics*, *39*(1), 113–119. https://doi.org/10.1038/ng1930

Weekley, C. and Smith, D.S. (2013). Malaria : the clinical basics Learning Overview (Clinical).

Weir, B.S. and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

WHO. (2010). World Malaria Report. *World Malaria Report*. https://doi.org/ISBN 978 92 4 1564403

WHO. (2017). World Malaria Report 2017, retrieved from http://www.who.int/malaria/publications/world-malaria-report-2016/report/en/

WHO. (2018). World Malaria Report 2018, retrieved from https://www.who.int/docs/default-source/gho-documents/world-health-statistic-reports/6-june-18108-world-health-statistics-2018.pdf

WHO. (2019). World Malaria Report 2019, retrievedfrom https://www.who.int/publications/i/item/9789241565721

Wolde, M.F., Berhanu, M.Y. and Bikila, L.E. (2016). Analysis of malaria prevalence

in Arsi Negelle health. *International Scholars Journals*, *3*(2), 90–95.

Ye, R., Tian, Y., Huang, Y., Zhang, Y., Wang, J., Sun, X., … Pan, W. (2019). Genome-wide analysis of genetic diversity in *Plasmodium falciparum* isolates from China–Myanmar border. *Genetics*, *10*(1065), 1–8. https://doi.org/10.3389/fgene.2019.01065

APPENDICES

Appendix I:Distribution of Samples' FWS Values in P. falciparum Populations Analyzed in this Study

Sample_id	FWS	Country
PH0877-C	0.998690919	Cambodia
PH0878-C	0.998507556	Cambodia
PH0871-C	0.998437472	Cambodia
PH0874-Cx	0.998413499	Cambodia
PH0899-C	0.997997614	Cambodia
PH0912-C	0.997857453	Cambodia
PH0933-C	0.997678974	Cambodia
PH0987-C	0.997517973	Cambodia
PH0676-C	0.997493392	Cambodia
PH0984-C	0.9974781	Cambodia
PH0904-C	0.997467041	Cambodia
PH0915-C	0.997424293	Cambodia
PH0866-C	0.997396315	Cambodia
PH0991-C	0.997378275	Cambodia
PH0928-C	0.997338708	Cambodia
PH0979-C	0.997333917	Cambodia
PH0880-C	0.997252615	Cambodia
PH0994-C	0.997094249	Cambodia
PH0921-C	0.996979338	Cambodia
PH0925-C	0.996699048	Cambodia
PH0990-C	0.996591379	Cambodia
PH0983-C	0.996146208	Cambodia
PH0999-C	0.995344397	Cambodia
PH0982-C	0.995251636	Cambodia
PH0924-C	0.994970852	Cambodia
PH0902-C	0.994845337	Cambodia
PH0911-Cx	0.993187684	Cambodia
PH0939-C	0.992480642	Cambodia
PH1003-C	0.992416573	Cambodia
PH0910-C	0.992192175	Cambodia
PH0937-C	0.98904881	Cambodia
PH0920-Cx	0.988537676	Cambodia
PH0931-C	0.981846208	Cambodia
PH0932-C	0.974487547	Cambodia

PH0901-C	0.972807072	Cambodia
PH0976-C	0.966201517	Cambodia
PH0905-C	0.932159067	Cambodia
PH0935-C	0.9194328	Cambodia
PH0903-C	0.917528283	Cambodia
PH0876-C	0.902756451	Cambodia
PH1007-C	0.890177489	Cambodia
PH0918-Cx	0.86594622	Cambodia
PH0917-Cx	0.825958692	Cambodia
PH0940-Cx	0.792392294	Cambodia
PH0941-C	0.747816169	Cambodia
PH0938-C	0.702386267	Cambodia
QG0249-C	0.997692542	DRCongo
QG0216-C	0.997577666	DRCongo
QG0205-C	0.99741637	DRCongo
QG0210-C	0.997328232	DRCongo
QG0257-C	0.99660103	DRCongo
QG0256-C	0.996551137	DRCongo
QG0258-C	0.996349511	DRCongo
QG0269-C	0.996287138	DRCongo
QG0223-C	0.996283167	DRCongo
QG0232-C	0.99615296	DRCongo
QG0225-C	0.995753299	DRCongo
QG0201-C	0.995726821	DRCongo
QG0291-C	0.994737603	DRCongo
QG0247-C	0.99471141	DRCongo
QG0202-C	0.992266101	DRCongo
QG0252-C	0.991138906	DRCongo
QG0286-C	0.989867132	DRCongo
QG0245-C	0.989166849	DRCongo
QG0239-C	0.988489213	DRCongo
QG0281-C	0.986934239	DRCongo
QG0255-C	0.986224999	DRCongo
QG0219-C	0.978083795	DRCongo
QG0211-C	0.973025032	DRCongo
QG0236-C	0.972046424	DRCongo
QG0204-C	0.960144235	DRCongo
QG0270-C	0.9358572	DRCongo
QG0263-C	0.909575787	DRCongo
QG0188-C	0.8911024	DRCongo
QG0215-C	0.881827704	DRCongo

QG0194-C	0.796905074	DRCongo
QG0207-C	0.78407086	DRCongo
QG0260-C	0.767546537	DRCongo
QG0280-C	0.740994475	DRCongo
QG0240-C	0.719600979	DRCongo
QG0253-C	0.708655312	DRCongo
QG0203-C	0.70554302	DRCongo
QG0294-C	0.683044572	DRCongo
QG0244-C	0.680622122	DRCongo
QG0231-C	0.675880387	DRCongo
QG0279-C	0.659683153	DRCongo
QG0213-C	0.633873479	DRCongo
QG0230-C	0.618988035	DRCongo
QG0235-C	0.584641838	DRCongo
QG0183-C	0.561061434	DRCongo
QG0189-C	0.5532151	DRCongo
QG0197-C	0.543402641	DRCongo
QG0191-C	0.540405897	DRCongo
QG0217-C	0.517619618	DRCongo
QG0193-C	0.512434014	DRCongo
QG0284-C	0.321381961	DRCongo
QS0159-C	0.997136885	Ethiopia
QS0144-C	0.996894132	Ethiopia
QS0104-C	0.99658765	Ethiopia
QS0132-C	0.996440061	Ethiopia
QS0056-C	0.996260953	Ethiopia
QS0133-C	0.996108538	Ethiopia
QS0110-C	0.996066757	Ethiopia
QS0109-C	0.994809978	Ethiopia
QS0168-C	0.994700867	Ethiopia
QS0169-C	0.994347938	Ethiopia
QS0087-C	0.994030515	Ethiopia
QS0170-C	0.992927704	Ethiopia
QS0166-C	0.989183394	Ethiopia
QS0162-C	0.988265921	Ethiopia
QS0157-C	0.986959585	Ethiopia
QS0155-C	0.986861821	Ethiopia
QS0163-C	0.984240975	Ethiopia
QS0154-C	0.983354429	Ethiopia
QS0135-C	0.981551903	Ethiopia
QS0129-C	0.975800195	Ethiopia

QS0116-C	0.953468797	Ethiopia
QS0127-C	0.926686018	Ethiopia
QS0126-C	0.896055262	Ethiopia
QS0156-C	0.892934956	Ethiopia
QS0128-C	0.836655462	Ethiopia
PT0215-C	0.997241538	Malawi
PT0270-C	0.99612365	Malawi
PT0146-C	0.996025603	Malawi
PT0234-Cx	0.995686924	Malawi
PT0080-C	0.995190162	Malawi
PT0034-C	0.995083926	Malawi
PT0129-C	0.994550549	Malawi
PT0086-C	0.994402865	Malawi
PT0197-C	0.992789437	Malawi
PT0009-CW	0.992515695	Malawi
PT0218-C	0.992230944	Malawi
PT0133-C	0.992120534	Malawi
PT0018-CW	0.991633553	Malawi
PT0145-C	0.990702691	Malawi
PT0137-C	0.987986943	Malawi
PT0109-C	0.979194491	Malawi
PT0173-C	0.977232294	Malawi
PT0195-C	0.965850816	Malawi
PT0073-C	0.928994174	Malawi
PT0239-Cx	0.908744295	Malawi
PT0090-C	0.89561901	Malawi
PT0091-C	0.895346962	Malawi
PT0196-C	0.889123117	Malawi
PT0237-C	0.806107062	Malawi
PT0271-C	0.77600774	Malawi
PT0117-C	0.748263464	Malawi
PT0124-C	0.738545663	Malawi
PT0078-Cx	0.696855267	Malawi
PT0043-C	0.664292794	Malawi
PT0260-C	0.659907505	Malawi
PT0055-C	0.656130062	Malawi
PT0085-C	0.649960585	Malawi
PT0228-C	0.639394426	Malawi
PT0209-C	0.63888405	Malawi
PT0136-C	0.628147225	Malawi
PT0103-C	0.586214079	Malawi

PT0266-C	0.569529891	Malawi
PT0001-CW	0.563728211	Malawi
PT0139-C	0.556603933	Malawi
PT0252-C	0.538318633	Malawi
PT0125-C	0.531098361	Malawi
PT0212-C	0.526038854	Malawi
PT0089-C	0.523655338	Malawi
PT0150-C	0.4363166	Malawi
PT0250-C	0.431551954	Malawi
PT0140-C	0.333279395	Malawi
PT0112-C	0.323973609	Malawi
PT0113-C	0.321524824	Malawi
PT0051-C	0.246233567	Malawi
PT0231-C	0.193640899	Malawi
PD0561-C	0.997940764	Thailand
PD0528-C	0.997578863	Thailand
PD0516-C	0.997454349	Thailand
PD0532-C	0.997120614	Thailand
PD0563-C	0.996820479	Thailand
PD0632-C	0.996755504	Thailand
PD0529-C	0.996614585	Thailand
PD0571-C	0.99651691	Thailand
PD0564-C	0.996505575	Thailand
PD0481-C	0.996242335	Thailand
PD0637-C	0.995867127	Thailand
PD0520-C	0.9957822	Thailand
PD0553-C	0.995775987	Thailand
PD0521-C	0.995729764	Thailand
PD0497-C	0.995598799	Thailand
PD0570-C	0.995454445	Thailand
PD0496-C	0.995291792	Thailand
PD0573-C	0.995283216	Thailand
PD0525-C	0.995178234	Thailand
PD0549-C	0.995087879	Thailand
PD0527-C	0.995046883	Thailand
PD0556-C	0.994846562	Thailand
PD0494-C	0.994842244	Thailand
PD0565-C	0.994702537	Thailand
PD0539-C	0.994415521	Thailand
PD0547-C	0.994309371	Thailand
PD0546-C	0.994308946	Thailand

PD0487-C	0.99383041	Thailand
PD0530-C	0.993578749	Thailand
PD0535-C	0.993173999	Thailand
PD0523-C	0.992455448	Thailand
PD0636-C	0.985587657	Thailand
PD0574-C	0.982754014	Thailand
PD0543-C	0.98258244	Thailand
PD0495-C	0.981884525	Thailand
PD0559-C	0.975710947	Thailand
PD0541-C	0.974143818	Thailand
PD0569-C	0.970297154	Thailand
PD0545-C	0.955092474	Thailand
PD0538-C	0.90732396	Thailand
PD0531-C	0.885504395	Thailand
PD0482-C	0.846604655	Thailand
PD0534-C	0.81486168	Thailand
PD0536-C	0.781611419	Thailand
PD0537-C	0.776151622	Thailand
PD0485-C	0.637166916	Thailand
PD0484-C	0.630129606	Thailand
PD0566-C	0.614597494	Thailand
PD0517-C	0.483456278	Thailand

Appendix II: Genes with Number of SNPs ≥5 and Average Tajima D ≥ 1: Defined as Genes Under Balancing Selection

Name/Identification number of gene	Chromosome	Number of SNP	TAJIMAD Value	Product Description
PF3D7_0425300	4	5	1.723878	Plasmodium exported protein (PHISTa), unknown function, pseudogene
FP2A	11	5	1.69833	cysteine proteinase falcipain 2a
PF3D7_0702000	7	6	1.6948417	Plasmodium exported protein (hyp12), unknown function
PF3D7_0619600	6	5	6	1.67197
LSA3	2	5	1.675734	liver stage antigen 3
PF3D7_0103500	1			conserved Plasmodium protein, unknown function
PF3D7_1303800	13	6	1.661325	conserved Plasmodium protein, unknown function
AP2-G	12	5	1.655106	AP2 domain transcription factor AP2-G
ATPase2	12	5	1.65117	aminophospholipid-transporting P-ATPase
PF3D7_0321800	3	5	1.632504	WD repeat-containing protein, putative
PF3D7_0711500	7	5	1.629558	regulator of chromosome condensation, putative
SURF14.1	14	13	1.6138031	surface-associated interspersed protein 14.1 (SURFIN 14.1)
RCC1	5	5	1.61187	guanidine nucleotide exchange factor
PF3D7_0221000	2	6	1.611385	Plasmodium exported protein, unknown function
ACPS	4	7	1.6082271	holo-[acyl-carrier-protein] synthase, putative
PF3D7_0826100	8	7	1.60401	HECT-like E3 ubiquitin ligase, putative
TRAP	13	8	1.6033988	thrombospondin-related anonymous protein
PF3D7_1140900	11	8	1.59787	conserved Plasmodium protein, unknown function
MSP7	13	6	1.5925483	merozoite surface protein 7
PF3D7_0414100	4	11	1.5865936	conserved Plasmodium membrane protein, unknown function
GEXP08	7	7	1.5840629	alpha/beta hydrolase, putative
PF3D7_0516300	5	7	1.58302	tRNA pseudouridine synthase, putative
PF3D7_0630300	6	6	1.5819067	DNA polymerase epsilon catalytic subunit A, putative
PF3D7_0522400	5	9	1.5799967	conserved Plasmodium protein, unknown function

CTRP	3	6	1.57454	circumsporozoite- and TRAP-related protein
PF3D7_1113600	11	5	1.57257	conserved Plasmodium protein, unknown function
PF3D7_0508900	5	6	1.56471	conserved Plasmodium protein, unknown function
CLAG8	8	19	1.5601663	cytoadherence linked asexual protein 8
DBLMSP	10	14	1.5551429	duffy binding-like merozoite surface protein
PF3D7_1475800	14	13	1.5499746	conserved Plasmodium protein, unknown function
PF3D7_1359900	13	5	1.547026	conserved Plasmodium membrane protein, unknown function
G377	12	5	1.543096	osmiophilic body protein G377
SURF8.2	8	72	1.5420215	surface-associated interspersed protein 8.2 (SURFIN 8.2)
PF3D7_0619800	6	8	1.53831	conserved Plasmodium membrane protein, unknown function
AMA1	11	12	1.535235	apical membrane antigen 1
CLAG3.1	3	5	1.534172	cytoadherence linked asexual protein 3.1
PF3D7_0104100	1	26	1.5290008	conserved Plasmodium membrane protein, unknown function
MSP1	9	43	1.524154	merozoite surface protein 1
PF3D7_1301800	13	49	1.5128173	surface-associated interspersed protein 13.1 (SURFIN 13.1), pseudogene
SMC4	5	5	1.50969	structural maintenance of chromosomes protein 4, putative
PF3D7_0820900	8	5	1.508708	conserved Plasmodium protein, unknown function
PF3D7_1408700	14	8	1.5082175	conserved protein, unknown function
CLAG3.2	3	7	1.5078371	cytoadherence linked asexual protein 3.2
PF3D7_1035100	10	7	1.5064629	probable protein, unknown function
PF3D7_0113600	1	27	1.5057607	surface-associated interspersed protein 1.2 (SURFIN 1.2), pseudogene
PF3D7_0914100	9	5	1.50085	conserved Plasmodium protein, unknown function
PF3D7_0418600	4	5	1.499868	regulator of chromosome condensation, putative
EBA175	7	12	1.4988025	erythrocyte binding antigen-175
PF3D7_1004600	10	10	1.498394	conserved Plasmodium membrane protein, unknown function
PRPF6	11	5	1.496918	pre-mRNA-processing factor 6, putative
AARP1	12	7	1.4931286	asparagine and aspartate rich protein 1
PF3D7_0614100	6	5	1.492988	conserved Plasmodium protein, unknown function

PF3D7_0609600	6	5	1.491024	conserved Plasmodium protein, unknown function
PF3D7_0113800	1	67	1.490863	DBL containing protein, unknown function
PF3D7_0422200	4	10	1.488567	erythrocyte membrane-associated antigen
PF3D7_1035000	10	9	1.48611	U2 snRNA/tRNA pseudouridine synthase, putative
PF3D7_1308400	13	5	1.48513	conserved Plasmodium protein, unknown function
RH1	4	6	1.484475	reticulocyte binding protein homologue 1
PF3D7_0919900	9	6	1.4803783	regulator of chromosome condensation-PP1-interacting protein
ABCB4	3	5	1.479234	ABC transporter B family member 4, putative
PF3D7_0221200	2	7	1.47721	Plasmodium exported protein (hyp15), unknown function
PF3D7_1239800	12	10	1.474321	conserved Plasmodium protein, unknown function
DDX60	9	5	1.471374	ATP-dependent RNA helicase DDX60, putative
SURF1.1	1	16	1.4710669	surface-associated interspersed protein 1.1 (SURFIN 1.1)
PF3D7_0809600	8	12	1.469735	peptidase family C50, putative
PF3D7_0630600	6	5	1.469408	conserved protein, unknown function
PF3D7_0425250	4	5	1.46253	Plasmodium exported protein (PHIST), unknown function, pseudogene
UT	7	15	1.4599107	E3 ubiquitin-protein ligase
CLAG2	2	11	1.45753	cytoadherence linked asexual protein 2
LISP2	4	5	1.44681	liver specific protein 2, putative
PF3D7_0421700	4	18	1.4451733	conserved Plasmodium protein, unknown function
PF3D7_0701900	7	10	1.443246	Plasmodium exported protein, unknown function
PF3D7_0424300	4	13	1.4426538	erythrocyte binding antigen-165, pseudogene
PF3D7_0415200	4	8	1.4412838	conserved Plasmodium protein, unknown function
ACC	14	5	1.43895	acetyl-CoA carboxylase
PF3D7_0425000	4	7	1.4355814	Plasmodium exported protein, unknown function, pseudogene
PF3D7_0419400	4	5	1.43404	conserved Plasmodium protein, unknown function
PF3D7_1434500	14	7	1.4327757	dynein-related AAA-type ATPase, putative
PF3D7_1429800	14	6	1.4320717	coatomer subunit beta, putative

PF3D7_0221300	2	8	1.4309863	Plasmodium exported protein, unknown function, pseudogene
PF3D7_1352700	13	5	1.43011	intron-binding protein aquarius, putative
SURF8.1	8	6	1.42716	surface-associated interspersed protein 8.1 (SURFIN 8.1)
PF3D7_0420000	4	14	1.4211343	zinc finger protein, putative
Cap380	3	6	1.4181533	oocyst capsule protein Cap380
pfa55-14	8	9	1.4162433	asparagine-rich antigen Pfa55-14
SURF4.2	4	33	1.4145235	surface-associated interspersed protein 4.2 (SURFIN 4.2)
RPB1	3	5	1.414388	DNA-directed RNA polymerase II subunit RPB1
PF3D7_0526600	5	8	1.4105813	conserved Plasmodium protein, unknown function
PF3D7_0221400	2	6	1.410555	Stevor
CRMP4	14	5	1.408492	cysteine repeat modular protein 4
PF3D7_0425100	4	12	1.4077008	Plasmodium exported protein (hyp6), unknown function
PF3D7_0220900	2	5	1.404562	Plasmodium exported protein, unknown function, pseudogene
PF3D7_0212400	2	7	1.3962829	conserved Plasmodium membrane protein, unknown function
ATPase1	5	10	1.393263	cation-transporting ATPase 1
PF3D7_0402200	4	22	1.3909873	surface-associated interspersed protein 4.1 (SURFIN 4.1), pseudogene
PF3D7_1475900	14	10	1.388352	KELT protein
PF3D7_1467600	14	5	1.38393	conserved Plasmodium protein, unknown function
SMC3	4	6	1.3796733	structural maintenance of chromosomes protein 3, putative
PF3D7_1339700	13	5	1.377052	conserved Plasmodium protein, unknown function
SET10	12	6	1.3706667	histone-lysine N-methyltransferase, H3 lysine-4 specific
PF3D7_0713900	7	9	1.3698478	conserved Plasmodium protein, unknown function
PF3D7_0425200	4	6	1.3690983	Plasmodium exported protein (hyp15), unknown function
SERA2	2	6	1.36821	serine repeat antigen 2
PF3D7_0307900	3	6	1.36739	conserved Plasmodium protein, unknown function
PF3D7_1201400	12	6	1.3624783	Plasmodium exported protein, unknown function
PF3D7_0504800	5	7	1.35628	conserved Plasmodium protein, unknown function
PF3D7_0808000	8	5	1.35249	conserved Plasmodium protein, unknown function
PF3D7_1225000	12	5	1.35249	conserved Plasmodium protein, unknown function
PF3D7_1004800	10	10	1.345123	ADP/ATP carrier protein, putative

CCT7	3	5	1.34434	T-complex protein 1 subunit eta
PF3D7_0808100	8	6	1.3428283	AP-3 complex subunit delta, putative
PF3D7_0713500	7	13	1.3417585	conserved Plasmodium protein, unknown function
PF3D7_0419900	4	11	1.3400391	phosphatidylinositol 4-kinase, putative
PF3D7_0513200	5	7	1.3317171	conserved Plasmodium protein, unknown function
PF3D7_0511300	5	6	1.32891	MORN repeat protein, putative
PF3D7_1349500	13	5	1.326946	conserved Plasmodium protein, unknown function
NFS	7	5	1.321048	cysteine desulfurase, putative
ApiAP2	4	5	1.3062579	AP2 domain transcription factor, putative
PF3D7_1228800	12	5	1.29747	WD repeat-containing protein, putative
PF3D7_0418000	4	10	1.296487	conserved Plasmodium protein, unknown function
PF3D7_0703900	7	10	1.287644	conserved Plasmodium membrane protein, unknown function
PF3D7_0904300	9	8	1.2463125	conserved protein, unknown function
PF3D7_0720400	7	5	1.238518	apoptosis-inducing factor, putative
BRR2	4	7	1.2243429	pre-mRNA-splicing helicase BRR2, putative
PF3D7_0308100	3	5	1.147146	zinc finger protein, putative

Appendix III: summary of sequence read and accession number of ENA per isolate for P. falciparum samples from west Arsi of Ethiopia

Sample	ENA accession Number	avg read length	bases of 1X coverage	bases of 5X coverage	mean coverage	%callable	Country
QS0056-C	ERR1035493	99	23028000	22471869	44.32	88.45	Ethiopia
QS0087-C	ERR1035520	99	22955070	21678711	24.03	73.76	Ethiopia
QS0104-C	ERR1035536	99	23044435	22700913	102.73	89	Ethiopia
QS0109-C	ERR1045266	99	23190557	23097616	255.46	90.33	Ethiopia
QS0110-C	ERR1045267	99	23127881	22980833	248.64	90.01	Ethiopia
QS0116-C	ERR1045271	99	23179531	22976729	151.38	89.98	Ethiopia
QS0126-C	ERR1045280	99	23224596	23118000	241.66	90.45	Ethiopia
QS0127-C	ERR1045281	99	22878355	20783836	16.46	80.49	Ethiopia
QS0128-C	ERR1045282	99	23037470	21637188	18.23	84.07	Ethiopia
QS0129-C	ERR1045283	99	23207516	22995958	104.71	90.25	Ethiopia
QS0132-C	ERR1045286	99	23179105	23063444	256.29	90.26	Ethiopia
QS0133-C	ERR1045287	99	23179617	23075708	243.61	90.29	Ethiopia

QS0135-C	ERR1045288	99	23147463	22815003	64.55	89.83	Ethiopia
QS0144-C	ERR1045295	99	23047268	22751022	106	89.18	Ethiopia
QS0154-C	ERR1106575	99	23132274	22597080	96.65	88.06	Ethiopia
QS0155-C	ERR1106606	99	23034693	22092716	47.49	85.31	Ethiopia
QS0156-C	ERR1106576	99	23099463	22610985	76.89	88.41	Ethiopia
QS0157-C	ERR1106577	99	23123132	22810518	85.32	89.13	Ethiopia
QS0159-C	ERR1106579	99	23148013	22866562	101.35	89.65	Ethiopia
QS0162-C	ERR1106581	99	22883708	21476485	107.77	82.99	Ethiopia
QS0163-C	ERR1106582	99	22964709	22175144	58.35	86.8	Ethiopia
QS0166-C	ERR1106584	99	21607223	15832040	8.99	58.99	Ethiopia
QS0168-C	ERR1106586	99	23005125	22363066	77.52	87.42	Ethiopia
QS0169-C	ERR1106587	99	23053276	22442425	87.88	87.76	Ethiopia
QS0170-C	ERR1106590	99	23126842	22737150	108.78	88.81	Ethiopia

Appendix IV: Sequence Read Summary of Samplesand Accession Number of ENA per Isolate Analyzed in this Study

Sample	ENA accession Number	avg read length	bases of 1X coverage	bases of 5X coverage	mean coverage	%callable	Country
PH0676-C	ERS175810	99	23188045	23090600	126.24	90.36	Cambodia
PH0866-C	ERS174566	99	23123389	22964223	91.2	90.2	Cambodia
PH0871-C	ERS174475	99	23141858	22948934	58.66	90.2	Cambodia
PH0874-Cx	ERS336335	99	22585609	18816036	15.08	73.48	Cambodia
PH0876-C	ERS174477	99	23231072	23004095	53.79	90.28	Cambodia
PH0877-C	ERS174478	99	23150890	22937730	55.59	89.95	Cambodia
PH0878-C	ERS174479	99	23152145	22926569	57.49	89.91	Cambodia
PH0880-C	ERS174481	99	23188878	22993121	65.35	90.18	Cambodia
PH0899-C	ERS174512	99	23139265	22839417	46.84	89.67	Cambodia
PH0901-C	ERS174484	99	23103229	22128316	24.13	86.51	Cambodia
PH0902-C	ERS174485	99	23058574	22194525	22.82	86.82	Cambodia
PH0903-C	ERS174486	99	23212656	22899897	39.52	89.86	Cambodia
PH0904-C	ERS174487	99	22995608	21433210	15.33	83.58	Cambodia
PH0905-C	ERS174488	99	23237074	22970872	39.26	89.89	Cambodia
PH0910-C	ERS174491	99	22967429	20676714	14.42	79.49	Cambodi

							a
PH0911-Cx	ERS336382	99	22182789	16731499	12.41	63.69	Cambodia
PH0912-C	ERS174492	99	23108999	22679700	33.69	89.16	Cambodia
PH0915-C	ERS174579	99	23173166	23019959	83.73	90.19	Cambodia
PH0917-Cx	ERS336365	99	23243746	22953166	151.05	89.52	Cambodia
PH0918-Cx	ERS336359	99	22327255	17503432	19.29	67.24	Cambodia
PH0920-Cx	ERS336377	99	22600261	19611687	27.72	76.7	Cambodia
PH0921-C	ERS193646	99	22858651	18466147	8.52	69.47	Cambodia
PH0924-C	ERS193651	99	23102176	22696063	25.48	89.02	Cambodia
PH0925-C	ERS193656	99	23191017	23067084	89.87	90.33	Cambodia
PH0928-C	ERS193671	99	23110707	22944542	69.03	89.97	Cambodia
PH0931-C	ERS193622	99	23226410	23054144	95.21	90.2	Cambodia
PH0932-C	ERS193627	99	23143961	22697092	29.06	89.14	Cambodia
PH0933-C	ERS193632	99	23152925	23001530	75.82	90.14	Cambodia
PH0935-C	ERS193637	98	23275803	23198419	109.68	90.55	Cambodia
PH0937-C	ERS193647	99	23117224	22412201	20.28	87.61	Cambodia
PH0938-C	ERS193652	99	23168166	21289309	14.2	81.76	Cambodia
PH0939-C	ERS193657	99	23058163	22172300	19.35	87.23	Cambodia
PH0940-Cx	ERS336371	99	21980576	15760546	12.35	59.81	Cambodia
PH0941-C	ERS193662	99	23182672	21780111	14.81	84.43	Cambodia
PH0976-C	ERS224907	99	23076545	22213345	17.55	87.35	Cambodia
PH0979-C	ERS224912	99	23191510	23092190	143.95	90.38	Cambodia
PH0982-C	ERS224868	99	23117878	22525921	24.85	88.3	Cambodia
PH0983-C	ERS224873	99	23101705	22710964	27.36	89.39	Cambodia
PH0984-C	ERS224878	99	23158659	23008280	105.58	90.29	Cambodia
PH0987-C	ERS199597	98	23157767	22961073	40.51	89.81	Cambodia
PH0990-C	ERS199607	99	23161893	22994314	62.2	90	Cambodia
PH0991-C	ERS199612	98	23100247	22638619	22.89	88.65	Cambodia
PH0994-C	ERS199617	98	23184525	23020331	51.26	90.09	Cambodi

							a
PH0999-C	ERS199632	99	23104455	22321583	18.82	87.15	Cambodia
PH1003-C	ERS224889	98	23169790	22894410	33.63	89.78	Cambodia
PH1007-C	ERS224875	99	22981494	19634213	9.07	74.31	Cambodia
QG0183-C	ERS347575	99	22656354	20622309	104.16	79.42	CongoDR
QG0188-C	ERS347615	98	23125681	21903585	40.56	84.53	CongoDR
QG0189-C	ERS347623	99	23276484	23034688	101.89	90.02	CongoDR
QG0191-C	ERS347544	99	23204200	22287121	81.18	86.44	CongoDR
QG0193-C	ERS347560	99	23273748	23045502	118.95	89.85	CongoDR
QG0194-C	ERS347568	99	23061016	21355021	29.7	82.71	CongoDR
QG0197-C	ERS347592	99	23143086	22122834	65.22	85.92	CongoDR
QG0201-C	ERS347624	99	23162059	22796110	114.31	88.96	CongoDR
QG0202-C	ERS347632	99	22722223	20540156	28.63	79.61	CongoDR
QG0203-C	ERS347545	98	23165274	22166376	67.03	85.69	CongoDR
QG0204-C	ERS347553	98	22998231	21360644	26.61	82.58	CongoDR
QG0205-C	ERS347561	99	23164911	22759752	104.39	88.79	CongoDR
QG0207-C	ERS347577	99	23250744	23022389	114.2	89.94	CongoDR
QG0210-C	ERS347601	99	22994684	21687466	47.23	83.96	CongoDR
QG0211-C	ERS347609	99	23161998	22459258	58.46	87.51	CongoDR
QG0213-C	ERS347625	99	23199591	22642423	79.98	88.2	CongoDR
QG0215-C	ERS347546	99	23134634	22145037	37.25	86.04	CongoDR
QG0216-C	ERS347554	99	23163105	22762354	99.28	88.75	CongoDR
QG0217-C	ERS347562	99	23256643	22865792	72.64	89.21	CongoDR
QG0219-C	ERS347578	99	21474574	15140060	10.18	57.28	CongoDR
QG0223-C	ERS347610	99	19100656	11504828	7.17	43.84	CongoDR
QG0225-C	ERS347626	99	22204104	17840659	19.16	66.82	CongoDR
QG0230-C	ERS347571	99	20976636	13150816	8.7	48.58	CongoDR
QG0231-C	ERS347579	99	23016305	21275953	36.33	82.04	CongoDR
QG0232-C	ERS347587	99	21222221	15965547	13.02	61.28	CongoD

							R
QG0235-C	ERS347611	99	23033202	21075374	29.2	81.02	CongoDR
QG0236-C	ERS347619	99	23156896	22244950	44.88	86.47	CongoDR
QG0239-C	ERS347548	99	22882885	21376244	47.23	82.85	CongoDR
QG0240-C	ERS347556	99	22777435	19646656	18.17	75.11	CongoDR
QG0244-C	ERS347588	99	22948417	21073358	22.53	81.63	CongoDR
QG0245-C	ERS347596	98	23186787	22780894	90.47	88.8	CongoDR
QG0247-C	ERS347612	99	22976556	21757790	37.77	84.34	CongoDR
QG0249-C	ERS347628	98	22990291	21858283	101.19	84.62	CongoDR
QG0252-C	ERS347557	99	22303493	19448507	30.98	74.79	CongoDR
QG0253-C	ERS347565	99	23188024	22693761	78.06	88.57	CongoDR
QG0255-C	ERS347581	99	22846569	20908971	34.92	80.59	CongoDR
QG0256-C	ERS347589	99	21776912	17160167	15.64	65.97	CongoDR
QG0257-C	ERS347597	99	23078095	22178499	78.38	86.15	CongoDR
QG0258-C	ERS347605	98	23034532	21591596	43.94	82.91	CongoDR
QG0260-C	ERS347621	98	22770690	19860080	33.71	75.62	CongoDR
QG0263-C	ERS347638	98	22826628	20464008	44.47	78.03	CongoDR
QG0269-C	ERS347680	98	23164240	22993306	108.55	90.12	CongoDR
QG0270-C	ERS347687	99	23199992	22639241	106.21	88.28	CongoDR
QG0279-C	ERS347667	99	23259953	22871215	74.48	89.33	CongoDR
QG0280-C	ERS347674	99	23230433	22865100	84.98	89.09	CongoDR
QG0281-C	ERS347681	97	23129667	22175410	34.47	84.93	CongoDR
QG0284-C	ERS347700	99	23280003	22757397	58.17	88.62	CongoDR
QG0286-C	ERS347712	99	22335005	18918520	18.7	72.5	CongoDR
QG0291-C	ERS347668	99	22012226	16907038	13.98	63.94	CongoDR
QG0294-C	ERS347689	99	22114129	16431569	13.09	60.96	CongoDR
PT0001-CW	ERS032647	74	23231526	23021767	64.27	89.76	Malawi
PT0009-CW	ERS032655	74	23125908	22868806	63.05	89.22	Malawi
PT0018-	ERS032662	74	23050199	22224782	31.53	86.74	Malawi

CW							
PT0034-C	ERS164677	99	23073481	22327994	49.07	86.8	Malawi
PT0043-C	ERS157475	99	23285638	23184510	83.78	90.73	Malawi
PT0051-C	ERS040107	74	23323324	23304436	145.17	90.66	Malawi
PT0055-C	ERS053948	98	23262160	23172684	55.31	90.35	Malawi
PT0073-C	ERS055903	98	23247434	23123720	60.72	90.56	Malawi
PT0078-Cx	ERS053888	98	23258697	23190922	119.74	90.59	Malawi
PT0080-C	ERS053940	98	23169363	22911749	41.12	89.29	Malawi
PT0085-C	ERS053890	99	23277253	23155152	70	90.67	Malawi
PT0086-C	ERS055908	98	23189333	23071700	51.88	90.23	Malawi
PT0089-C	ERS053954	99	23310571	23165833	74.5	90.32	Malawi
PT0090-C	ERS053866	98	23268294	23177938	94.71	90.78	Malawi
PT0091-C	ERS053939	98	23240960	23093396	51.33	90.04	Malawi
PT0103-C	ERS053956	98	23300169	23239598	108.2	90.69	Malawi
PT0109-C	ERS055909	98	23260476	23163649	126.29	90.5	Malawi
PT0112-C	ERS055910	98	23316893	23292619	94.76	91.18	Malawi
PT0113-C	ERS053946	98	23320051	23240733	47.8	90.7	Malawi
PT0117-C	ERS168650	99	23293280	23228553	123.01	90.86	Malawi
PT0124-C	ERS168610	99	23296965	23247959	201.86	91.01	Malawi
PT0125-C	ERS168596	99	23285720	23206255	105.52	90.65	Malawi
PT0129-C	ERS168597	99	23182294	23077313	124.34	90.15	Malawi
PT0133-C	ERS168614	99	23251776	23148946	217.93	90.56	Malawi
PT0136-C	ERS168636	98	23281240	23224037	142.93	90.84	Malawi
PT0137-C	ERS168646	99	23245944	23133165	180.1	90.43	Malawi
PT0139-C	ERS168616	99	23283735	23242641	196	90.88	Malawi
PT0140-C	ERS175813	99	23312003	23277859	104.8	91.24	Malawi
PT0145-C	ERS168620	98	23230543	23162294	135.38	90.49	Malawi
PT0146-C	ERS168598	99	23186961	23082558	128.35	90.21	Malawi
PT0150-C	ERS168644	99	23310211	23264249	118.09	91.07	Malawi
PT0173-C	ERS168629	99	23182872	23053000	159.03	90.29	Malawi
PT0195-C	ERS188101	99	23281919	23185355	149.99	90.52	Malawi
PT0196-C	ERS188142	99	22956543	18885446	8.52	71.5	Malawi
PT0197-C	ERS188108	98	23193235	23079283	85.98	90.27	Malawi
PT0209-C	ERS188102	99	23262180	23202951	102.8	90.83	Malawi
PT0212-C	ERS188123	99	23315382	23270961	81.76	91.09	Malawi
PT0215-C	ERS188144	99	23222984	23153388	201.22	90.44	Malawi
PT0218-C	ERS188082	98	23263511	23169095	123.77	90.49	Malawi
PT0228-C	ERS188145	98	23269414	23210128	91.66	90.73	Malawi
PT0231-C	ERS188083	98	23327363	23313595	118.69	91.34	Malawi
PT0234-Cx	ERS188097	99	23210804	23098318	126.88	90.34	Malawi
PT0237-C	ERS193638	98	23246510	23150659	81.03	90.37	Malawi

PT0239-Cx	ERS188118	99	23257604	23161098	146.63	90.46	Malawi
PT0250-C	ERS188091	98	23309137	23274203	99.05	90.93	Malawi
PT0252-C	ERS188098	98	23285553	23187416	52.11	90.53	Malawi
PT0260-C	ERS193663	99	23287543	23197284	76.36	90.77	Malawi
PT0266-C	ERS188071	99	23306798	23264207	105.6	91.04	Malawi
PT0270-C	ERS193673	98	23193994	23054017	59.46	90.1	Malawi
PT0271-C	ERS193678	98	23283511	23155115	62.96	90.47	Malawi
PD0481-C	ERS142884	99	23196022	23081764	136.98	90.22	Thailand
PD0482-C	ERS142825	99	23241301	23170798	158.95	90.62	Thailand
PD0484-C	ERS142827	99	23260269	23171441	115.24	90.59	Thailand
PD0485-C	ERS142824	99	23285583	23246330	209.91	90.89	Thailand
PD0487-C	ERS142818	99	23191659	23046863	108.52	90.13	Thailand
PD0494-C	ERS142833	99	23193432	23108802	182.18	90.34	Thailand
PD0495-C	ERS142854	99	23135951	22742720	42.57	89.18	Thailand
PD0496-C	ERS142834	99	23182722	23076349	126.36	90.29	Thailand
PD0497-C	ERS142869	99	23176708	23068188	97.03	90.31	Thailand
PD0516-C	ERS174522	99	23214859	23118842	103.17	90.51	Thailand
PD0517-C	ERS174523	98	23249417	23166754	65.14	90.62	Thailand
PD0520-C	ERS174636	99	23131220	22881192	58.59	89.72	Thailand
PD0521-C	ERS174525	99	23143714	22972965	73.15	90.16	Thailand
PD0523-C	ERS174637	99	23211810	23045537	126.76	90.18	Thailand
PD0525-C	ERS174639	98	23153483	23031855	87.2	90.15	Thailand
PD0527-C	ERS174527	99	23139091	22953974	64.09	90.15	Thailand
PD0528-C	ERS174641	99	23135778	22993305	116.69	90.19	Thailand
PD0529-C	ERS174642	99	23191690	23076830	97.7	90.27	Thailand
PD0530-C	ERS174643	98	23171069	23058447	95.14	90.12	Thailand
PD0531-C	ERS174528	99	23192739	23018278	58.41	90.25	Thailand
PD0532-C	ERS174644	99	23172758	23037160	87.66	90.1	Thailand
PD0534-C	ERS174530	99	23262869	23170162	114.7	90.65	Thailand
PD0535-C	ERS174645	99	23177437	23005570	85.12	90.05	Thailand
PD0536-C	ERS174646	99	23257024	23162818	84.71	90.49	Thailand
PD0537-C	ERS174531	99	23282320	23179994	104.25	90.67	Thailand
PD0538-C	ERS174532	99	23229566	23097668	67.44	90.45	Thailand
PD0539-C	ERS174647	98	23179304	23056406	102.25	90.1	Thailand
PD0541-C	ERS174649	98	23220342	23033661	71.67	90.14	Thailand
PD0543-C	ERS174650	98	23197563	23051813	83.57	90.08	Thailand
PD0545-C	ERS174652	99	23139841	22841949	38.25	89.62	Thailand
PD0546-C	ERS174653	99	23095871	22666520	29.59	88.91	Thailand
PD0547-C	ERS174654	98	23148272	22990233	91.44	89.97	Thailand
PD0549-C	ERS174533	99	23171262	23045306	121.33	90.28	Thailand
PD0553-C	ERS174536	99	23157323	22977548	67.55	90.13	Thailand

PD0556-C	ERS174538	99	23146828	22900755	49.61	89.77	Thailand
PD0559-C	ERS174539	99	23173615	22851571	46.77	89.74	Thailand
PD0561-C	ERS174659	98	23150186	22956252	69.44	89.84	Thailand
PD0563-C	ERS174661	98	23165932	23034075	97.4	90.16	Thailand
PD0564-C	ERS174662	98	23170399	23033936	79.07	90.15	Thailand
PD0565-C	ERS174663	98	23133531	22923970	54.49	89.72	Thailand
PD0566-C	ERS174664	98	23254716	23139969	59.05	90.32	Thailand
PD0569-C	ERS174666	98	23193723	23032886	78.67	90.11	Thailand
PD0570-C	ERS174667	99	23148167	23019860	81.6	90.15	Thailand
PD0571-C	ERS174668	99	23150941	23016095	98.58	90.11	Thailand
PD0573-C	ERS174669	99	23165596	23014419	85.51	90.1	Thailand
PD0574-C	ERS174541	99	23226577	23086894	194.85	90.29	Thailand
PD0632-C	ERS347472	99	23134394	22651332	69.72	88.38	Thailand
PD0636-C	ERS347504	99	22705713	19879383	22.62	76.56	Thailand
PD0637-C	ERS347512	99	22902102	20698320	24.12	79.82	Thailand

Appendix V: Publication

Abera *et al. Malar J* (2021) 20:129
https://doi.org/10.1186/s12936-021-03660-y

Malaria Journal

RESEARCH Open Access

Genomic analysis reveals independent evolution of *Plasmodium falciparum* populations in Ethiopia

Deriba Abera[1,2], Caleb K. Kibet[3], Teshome Degefa[2], Lucas Amenga-Etego[4], Joel L. Bargul[1,3] and Lemu Golassa[5*]

Abstract

Background: *Plasmodium falciparum* parasite populations in Ethiopia have been experiencing local selective pressures from drugs and immunity, leading to evolutionary adaptation. However, there was a paucity of data on genomic characterization and evolutionary adaptations of *P. falciparum* isolates from the central area of Ethiopia.

Methods: Whole-genome analysis of 25 *P. falciparum* isolates from central Ethiopia, specifically from West Arsi, were studied to determine their genetic diversity, population structures, and signatures of selection in known drug resistance alleles against global isolates from Cambodia, Thailand, DR Congo, and Malawi.

Results: A total of 18,517 high-quality single-nucleotide polymorphisms (SNPs) were identified in Ethiopian *P. falciparum* isolates. About 84% of the Ethiopian *P. falciparum* isolates had a F_{WS} value > 0.95 showing a dominant single genotype infection in most isolates at the time of collection with little potential for out-crossing as expected in areas with low transmission intensity. Within-host diversity of Ethiopian infections was significantly different from East African ($p < 0.001$), but not Southeast Asian infections ($P > 0.05$). A significant population structure has been observed by PCA and population differentiation between Ethiopian parasites and East African (Fst ~ 10%) and Southeast Asian populations (Fst ~ 18%), suggesting limited gene flow and the independent evolution of the Ethiopian parasite population. Moreover, a total of 125 genes under balancing selection was found that include *ama1*, *trap*, *eba175*, and *Isa3*, previously identified as targets of human host immunity. Recent directional selection analysis using integrated standardized haplotype score (IHS) did not detect any selection signatures in the *Pfcrt*, *Pfdhfr*, *Pfdhps*, *Pfmdr*1, and *PfK13* genes. However, known drug resistance-conferring mutations analysis showed that at least one SNP marker was fixed in these genes, but not in *Pfdhps* and *PfK13*.

Conclusion: *Plasmodium falciparum* populations in the central region of Ethiopia was structurally diverged from both Southeast Asian and other East African populations. Malaria infections in Ethiopia had low within-host diversity, and parasites carry fixed chloroquine resistance markers despite the withdrawal of this drug for the treatment of *P. falciparum*.

Keywords: *Plasmodium falciparum*, Ethiopia, Population structure, Drug resistance, Admixture, Positive selection

Background

Plasmodium falciparum **malaria remains one of the major public health problems worldwide accounting for 228 million cases in 2018 compared to 231 million in 2017, while the number of deaths due to malaria**

*Correspondence: lgolassa@gmail.com
[5] Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia
Full list of author information is available at the end of the article

decreased by just 2.5%, from 415,000 to 405,000 during the same period [1]. Sub-Saharan Africa (sSA) still accounts for 94% of global death. In Ethiopia, more than 75% of the total area is malarious, and *P. falciparum* and *Plasmodium vivax* co-exist [2] making malaria control more complicated than in other African countries.

Across malaria-endemic regions, large-scale deployment of anti-malarial drugs has led to the emergence of drug resistance to chloroquine (CQ) and sulfadoxine/pyrimethamine (SP) antifolate drugs [3–5]. Like many other countries, Ethiopia has switched from CQ to SP in 1998 and from SP to artemether–lumefantrine (AL) in 2004 [6] for the treatment of uncomplicated *P. falciparum* malaria in response to the development of parasite resistance. However, CQ remained the first-line drug for *P. vivax* treatment in the country [7], leading to a continued selection of CQ-resistance markers in *P. falciparum* as the result of indirect pressure from CQ and the presence of mixed infections of *P. falciparum* and *P. vivax*. Similar to CQ and SP, *P. falciparum* developed resistance to AL first at the Thai-Cambodian border [8] and recently in East Africa [9]. The rapid development of resistance in *P. falciparum* to series of the first-line anti-malarial hinders malaria prevention, control, and elimination efforts.

Anti-malarial drugs are known to pose tremendous selective pressure on *P. falciparum* leading to the worldwide spread of resistant parasites [10]. It was well noted that *P. falciparum* resistance to the two conventional anti-malarial drugs, CQ and SP, has resulted in increased malaria morbidity and mortality across endemic settings. Apart from increased morbidity and mortality, selective sweeps of drug resistance mutations have reduced levels of polymorphism in *P. falciparum* as these resistant and sensitive strains continue to recombine in mosquitoes [11, 12], with perhaps a reduced diversity around the selected loci. However, the greatly reduced level of diversity across the entire *P. falciparum* genome most likely resulted from a severe population bottleneck as has been observed in gorilla-to-human cross-species transmission [13].

Based on the analysis made on 12 strains collected from different countries in Africa and Asia, the average diversity of *P. falciparum* at fourfold degenerate sites was estimated to be 8×10^{-4} per site [13]. However, published mutation rates for *P. falciparum* were in the range of $1–10 \times 10^{-9}$ mutations per site per replication cycle [14, 15]. Depending on *P. falciparum* life cycle and assuming varying lengths of time that the parasites spend either in the vector or in the mammalian host, the parasites are likely to undergo at least 200 replication cycles per year suggesting that the observed level of genetic diversity in *P. falciparum* could have readily accumulated within the past 10,000 years [10].

Plasmodium falciparum parasite change and select its new genetic variants owing to drug exposure to cause disease and overcome challenges from host immunity and therapeutic interventions [16]. Indeed, high pressure from immunity and drugs are known to select adaptive parasite strains that maintain transmission [17] and, therefore, many *P. falciparum* genes encoding immune and drug targets are under natural selection and show signatures of balancing or directional selection [4, 17–21]. This selection may vary due to differences in innate susceptibility of human populations, variations in ecological transmission, resulting in varying degrees of acquired immunity and/or drug pressure. Malaria parasites from low and high endemic regions have a distinct opportunity for transmission and host acquired immune responses [22]. For effective management of malaria control and intervention strategies, it is important to determine genetic variation patterns due to parasite adaptation to host environments and drug interventions. Balancing selection brings the favoured alleles of parasites to an intermediate equilibrium where they are maintained as genetic polymorphisms, while the directional selection forces cause the parasite's genetic variants to increase in frequency and facilitate the occurrence of selective sweeps around the affected loci [23].

Plasmodium falciparum population genomics has been highly studied in West African populations and showed signatures of balancing selection on multiple candidate vaccine antigens and strong directional selection around known drug resistance genes [19, 24]. In contrast, there is little information about the genomic variations of *P. falciparum* populations in the horn of Africa, including Ethiopia, where *P. falciparum* and *P. vivax* malaria coexist and are heterogeneously distributed. A recent study reported *P. falciparum* populations in the horn of Africa, specifically in Ethiopia, are unique and structurally diverged from other West, East, and central African *P. falciparum* populations [25]. These parasite populations share a chunk of genes with other sub-Saharan African *P. falciparum* populations across drug and immune targets and facilitate the spread of drug-resistant strains [25]. These studies call for in-depth analysis of Ethiopian parasite genomes to deepen understanding of genome diversity and natural selection in Ethiopia's unique human populations with co-species transmission dynamics.

Understanding the population genetic diversity of *P. falciparum* strains circulating in the specific region of central Ethiopia is very important to monitor the effectiveness of control schemes and provide baseline information for making informed decisions by the national malaria control programme [26]. This study aimed to characterize the *P. falciparum* populations in West Arsi,

1.

in central Ethiopia, using whole-genome analysis of data generated by Illumina next-generation sequencing.

Methods

Study area and population

The study was conducted in West Arsi, Oromia (07′ 17″ 34.2 S, 038′ 21″ 46.3 W) located about 251 km from Addis Ababa, Ethiopia. This region with distinct wet and dry seasons has an altitude of about 1500–2300 m above sea level with the human population of 176,671. The inhabitants of this malarious region have high levels of poverty worsened by the malaria diseases caused predominantly by *P. falciparum* and *P. vivax* with a seasonal and unstable pattern of transmission [7].

Sample collection and processing

Venous blood (2–5 mL) was collected from July 2012 to December 2013 from consented *P. falciparum* malaria patients following standard procedures. Sequencing of 34 *P. falciparum* samples from leukocyte-depleted infected whole blood was done as described by Auburn et al. [27], at the Welcome Sanger Institute as part of the MalariaGEN *P. falciparum* Community Project (www.malariagen.net/projects). Freely available *P. falciparum* sequence data were accessed via the Pf3K project (https://www.malariagen.net/data/pf3k-5) for Southeast Asian and East African samples. Among East Africa, samples are available only in DR Congo and Malawi from East Africa and randomly took Cambodia and Thailand. Sample collection site with the number of samples greater or equal to Ethiopian samples with closer/similar *P. falciparum* samples collection year to Ethiopian sample were randomly selected. Then, 50 samples were randomly selected from each country.

Short sequence reads were generated on the IlluminaHiSeq platform and aligned to Pf3D7 reference (version 3) by burrows-wheeler-aligners (BWA). SNP calling was done following a customized genome analysis tool kit (GATK) pipeline. Each sample was genotyped for polymorphic coding SNPs across the genome, ensuring a minimum of 5× paired-end coverage across each variant per sample. Polymorphic sites within hyper-variable, telomeric, and repetitive sequence regions were excluded. Biallelic high-quality SNPs with mapping quality (MQ) > 20 and Variant Quality Score (VQSLOD) ≥ 3 in the core region loci with a minor allele frequency of at least 2% and individual sample with less than 10% missing data and SNP-site missing less than 10% across the isolate was extracted and used for downstream analysis. After quality filtering, 46, 50, 25, 50, 49 samples of Cambodia, DR Congo, Ethiopia, Malawi, and Thailand were left, respectively.

Analysis of population genetic diversity and within-host infection diversity

The genome-wide F_{WS} (inbreeding coefficient within a population) metric was used to calculate within-host diversity as described by Manske et al. [21]. To derive F_{WS} $(=1-H_W/H_S)$, within isolate, expected heterozygosity (H_W) was calculated from the relative allele frequencies for all genic SNPs, averaged across the genome and compared with the heterozygosity of local population (H_S). F_{WS} value ranged from zero to one, where zero indicates high diversity of infection, and one represents a single infection within the sample as compared to local population diversity. For this analysis, individual alleles with coverage of less than < 5 reads and positions with total coverage of < 20 reads were classified as undermined (missing). Isolates with greater than > 10% missing SNP data and SNPs with > 10% missing isolate data were discarded. Isolates with F_{WS} scores of > 0.95 were classified as single predominant genotype infections.

Population structure and admixture analysis

Principal component analysis (PCA) was used to estimate population structure using the *glPCA* function in the open-source R statistical software version 3.6.2. The first 10 principal components axis (PCs) were calculated and the first three PCs which explained the majority of the variation in the data were retained. The data was thinned downed by pruning SNPs with pairwise linkage disequilibrium (LD) by r^2 greater than 0.05 for determining the PCs. The pruned SNP loci employed in the *glPCA* function was used to calculate an allele sharing matrix in custom Rscripts. This function use variance between and within groups to determine population genetic structure. A discriminant analysis of principal components (DAPC) [28] was used to transform the PCA data, and perform discriminant analysis on the retained principal components using the *adegenet* package in the *R* software version 3.6.2. Population admixture was determined based on spatial modelling of allele sharing among geographical coordinates of sampling sites. DAPC determines ancestry proportions and membership probability modelled on genetic variation across space to determine admixtures as described by Jombart et al. [28].

Allele frequency and differentiation analysis

Analyses of allele frequency distributions between-population F_{ST} values [29] were calculated using Vcftools or *hierfstat* package from *adegenet* in Rafter excluding SNPs with greater than 10% missing data. For F_{ST} analysis, missing data were excluded on the SNP basis with the

size of each population corrected to account for F_{ST} value difference due to population size variation.

Detection of signatures of natural selection

Within-population Tajima's D index [30] was calculated using Vcf tools. Tajima D values were determined for each SNP and the average value for each gene was calculated. Genes with at least five SNP and positive Tajima D values > 1 were considered as genes under balancing selection.

The standardized integrated haplotype score (IHS) analysis was used to identify positive directional selection signatures by using phased SNP data with allele frequency > 5%. IHS was determined using the *rehh* package in R software with default parameters [31] after imputing missing SNP data using Beagle version 5.2. The |IHS| > 2.5 (top 1% of the expected distribution) was used as cut off value [32] to report genes under recent directional selection as reported for genome analysis of West African *P. falciparum* [17].

Results

Sequencing of *P. falciparum* and analysis of allele frequency

High-quality sequence data obtained from 25 *P. falciparum* clinical isolates (Additional file 1) collected from the West Arsi of Ethiopia enabled the identification of 672,956 biallelic SNPs with less than 10% missing SNPs data and < 10% sample missing data in the individual isolate. All isolates had 95.95% (645,715/672,956) SNPs call. Sequences from the intergenic regions had lower read coverage compared to those sequences in the coding regions, and as a result, 78.92% (531,120/672,956) of all SNPs called were located within genes. Of 5058 genes analysed, 3370 genes had at least one SNP (Table 1, Additional file 2: Fig. S1A). About 18,517 SNPs were polymorphic marker in at least one sample in Ethiopian (n = 25) samples of which 43.4% (8037/18,517) were non-synonymous coding SNPs, 22.8% (4222/18,517) synonymous coding SNPs, 26.6% (4932/18,517) in intergenic regions, 3.6% (666/18,517) other intragenic regions and 3.5% (656/18,517) SNPs in intron region. Similarly, *P. falciparum* populations from Cambodia (n = 46), DR Congo (n = 50), Malawi (n = 50), and Thailand (n = 49) had 32,854, 68,476, 79,250 and 30,427 biallelic polymorphic SNPs marker in at least one sample, respectively (Table 1, Additional file 2: Fig. S1B). The proportion of non-synonymous coding to synonymous coding and the intragenic to intergenic SNPs were ~ 2 or above in all populations (Table 1).

In general, all populations had a high percentage of non-synonymous coding SNPs at polymorphic marker consistent with previous findings [17]. SNPs with minor allele frequency (MAF) < 5% were common in all analysed *P. falciparum* populations following the exclusion of monomorphic SNPs in each population. Further, SNPs with minor allele frequency of < 5% occurred more frequently in samples from Malawi than in Ethiopian isolates (Additional file 2: Fig. S2).

Genomic diversity of *P. falciparum* infections

F_{WS} scores ranged from 0.837 to 0.997 (mean = 0.97, median = 0.99) for Ethiopian *P. falciparum* infections whereas the F_{WS} values in Cambodia ranged from 0.702 to 0.999 (mean = 0.962, median = 0.995), from 0.483 to 0.998 (mean = 0.94, median = 0.994) in Thailand, from 0.321 to 0.998 (mean = 0.94, median = 0.994) in DR Congo and from 0.194 to 0.997 (mean = 0.747, median = 0.762) in Malawi (Fig. 1; Additional file 3).

The F_{WS} value of > 0.95 suggests that the individual samples predominantly contained a single genotype and could have other additional genotypes in lower proportions. In this study, F_{WS} values of > 0.95 were observed in 84%, 79.6%, 78%, 50%, and 36% of samples from Ethiopia, Thailand, Cambodia, DR Congo, and Malawi, respectively.

The mean F_{WS} scores of the Ethiopian *P. falciparum* population were not significantly different from Cambodia's (Welch two Sample t-test, P = 0.42) and Thailand's (Welch two-sample t-test, p = 0.083) at 95% confidence intervals. However, mean F_{WS} was

Table 1 Distribution of polymorphic SNP marker effects and their relative proportion in each *P. falciparum* population

Country of origin	Non-synonymous coding	Synonymous coding	Intergenic	Intron	Other intragenic	Proportion of non-synonymous to synonymous coding	Proportion of intragenic to intergenic	Total
SNP in Cambodia	13,271	6063	11,059	1494	806	2.19	1.96	32,693
SNP in DR Congo	29,762	14,450	19,685	3034	1255	2.06	2.46	68,186
SNP in Ethiopia	8037	4226	4932	656	666	1.90	2.75	18,517
SNP in Malawi	31,756	15,623	26,538	3608	1384	2.03	1.97	78,909
SNP in Thailand	12,194	5489	10,466	1342	792	2.22	1.89	30,283

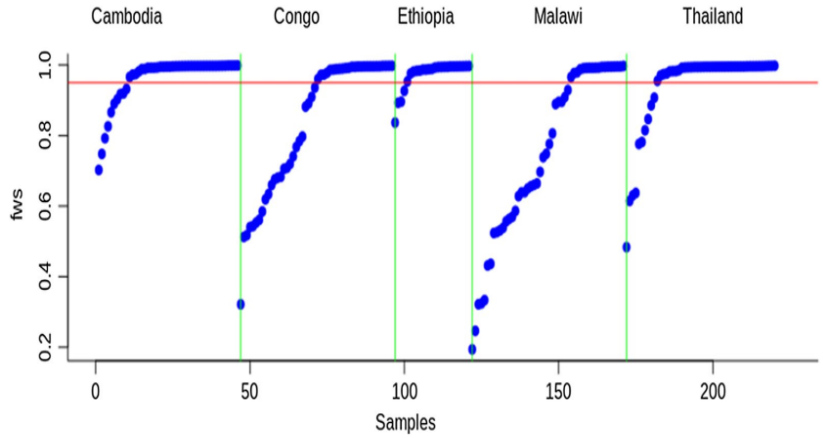


Fig. 1 Ranked distribution of F_{WS} values by population. Redline marks $F_{WS} = 0.95$, above which an infection is considered clonal. A vertical green line separates the *P. falciparum* population

significantly higher in Ethiopia compared to DR Congo (Welch two-sample t-test, $p = 5.603e^{-06}$) and Malawi (Welch two-sample t-test, $p = 3.242e^{-08}$) at 95% confidence intervals.

Population structure and admixtures

Analysis using PCA revealed the presence of four clear major population groups of isolates, which were coincident with their geographical origins (Fig. 2a–c). Similarly, the findings from admixture analysis were consistent with the PCA clustering. The isolates from

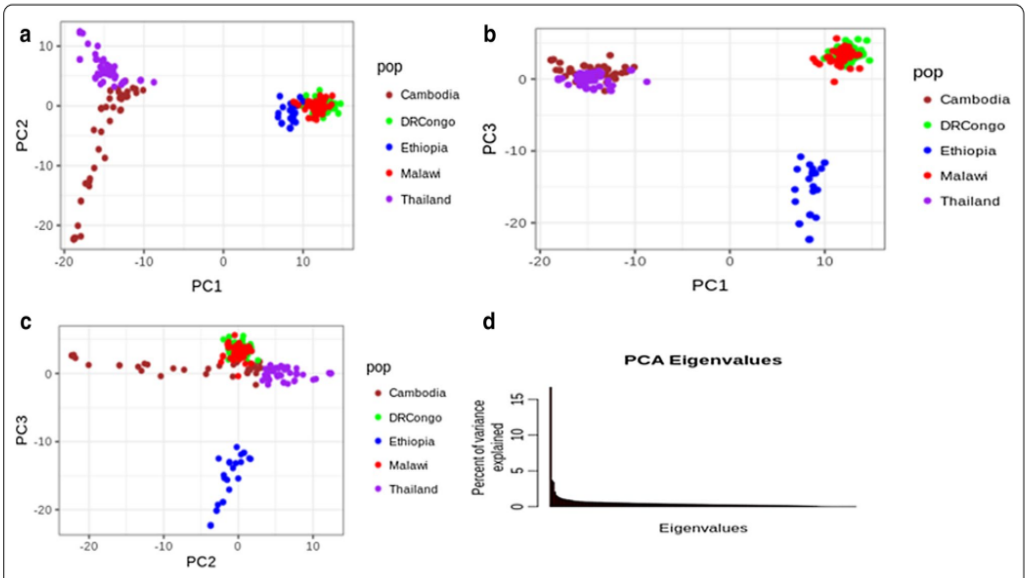


Fig. 2 Principal component analysis. **a** Principal component axis1 and axis2 identified Southeast Asian from African *P. falciparum* population. **b** Principal component axis1 and axis3 identified Ethiopian, Southeast Asian and other East African *P. falciparum* populations. **c** Principal Component axis2 and axis3 identified Ethiopian *P. falciparum* population from Southeast Asian and African *P. falciparum* population. **d** Percent of variance explained by each principal component axis. PC1 explained 16.7%, PC2 explained 3.7% and PC3 explained 3.4% of the variance in the data. Pop stands for the *P. falciparum* country of origin

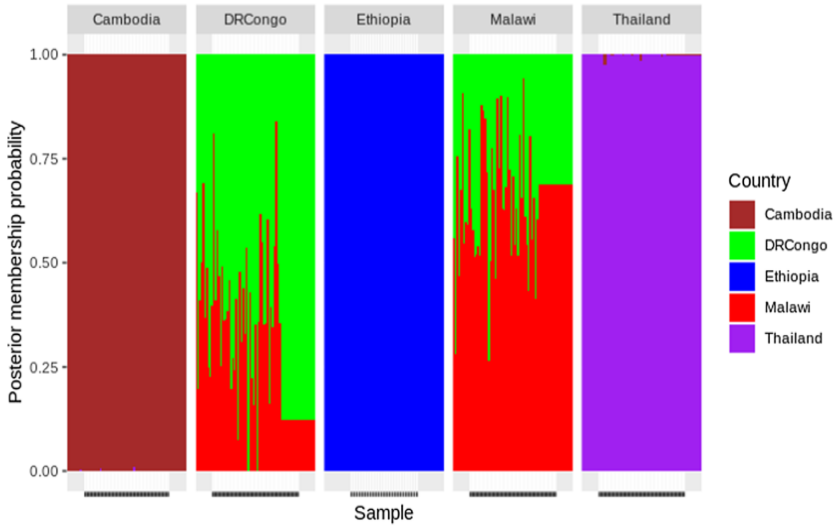


Fig. 3 An admixture analysis: X-axis represent the assigned samples to country listed above the figure after modeling allele sharing probability, while Y-axis represent the membership probability of samples. Colour-coded country listed at the right of figure indicated the origin of *P. falciparum* sample collection. Analysis of admixture identified four major components based on an optimized cluster value of K = 5, and multiple subpopulations of *P. falciparum* were found in DR Congo and Malawi. There was a bidirectional gene flow between Malawian and Congolese *P. falciparum* populations

Table 2 Pairwise population divergence (measured by FST) among *P. falciparum* populations

Country	Cambodia	DR Congo	Ethiopia	Malawi	Thailand
Cambodia	0	0.16	0.18	0.17	0.06
DR Congo	0.16	0	0.08	0.02	0.16
Ethiopia	0.18	0.08	0	0.09	0.18
Malawi	0.17	0.02	0.09	0	0.17
Thailand	0.06	0.16	0.18	0.17	0

Ethiopian *P. falciparum* highly diverged from both Southeast Asian and East African *P. falciparum* populations

the three regions were distinguished. This admixture analysis showed that four major components could be differentiated with a cluster value of K = 5. Multiple parasite subpopulations were observed in Malawi and DR Congo parasite populations suggestive of gene flow between these two populations (Fig. 3). There was no detectable gene flow between the isolates from Ethiopia and East African or Southeast Asia.

The clustering of Ethiopian *P. falciparum* isolates was consistent with the fixation index (F_{ST}) values with or without correcting for sample size. The F_{ST} values of Ethiopian isolates versus those from the two other East African regions (DR Congo and Malawi) ranged from 0.08 to 0.09, while the F_{ST} value of Ethiopian *P. falciparum* versus the two southeast Asian regions (Thailand and Cambodia) was 0.18 (Table 2).

Signatures of selection in the *P. falciparum* isolates

The Ethiopian isolates had the average Tajima's D value of 0.18 across the entire genome (one sample t-test, $p < 2 \times 10^{-16}$). 1,450 genes had at least one SNP with Tajima D value > 1 of which 125 genes had at least five SNPs with Tajima D values > 1 of which 125 genes had at least five SNPs with Tajima D values > 1 (Additional file 4). These genes include *apical membrane antigen-1 (ama1), erythrocyte binding antigen-175 (eba175), merozoites surface protein-1 (msp1), thrombospondin-related anonymous protein (trap), duffy binding like merozoites surface protein (dblmsp)*, and *cytoadherence linked asexual protein 2 (clag2)*, that were previously reported for the balancing selection [24, 33].

The standardized integrated haplotype homozygosity score (IHS) was applied to investigate genome-wide evidence for recent positive directional selection due to drug pressure, immune impact, or other mechanisms. Using |IHS| score of > 2.5 (top 1% of the expected distribution) as a threshold for hits, 36 genes with at least one SNP that could be under significant positive selection were identified, and out of these, 15 genes had at least two SNPs (Table 3).

Table 3 Genes with at least two SNPs that had a recent positive directional selection in *P. falciparum* of Ethiopia, identified using the integrated haplotype score at a significance threshold of P < 0.01

Chromosomes	Number of SNPs	Genes name or ID	Product description
1	9	PF3D7_0104100	Conserved *Plasmodium* membrane protein, unknown function
1	3	PF3D7_0113600	Surface-associated interspersed protein 1.2 (SURFIN 1.2)
4	2	PF3D7_0424300	Erythrocyte binding antigen-165
4	14	SURF4.2	Surface-associated interspersed protein 4.2 (SURFIN 4.2)
4	5	PF3D7_0425200	*Plasmodium* exported protein (hyp15), unknown function
4	4	PF3D7_0425250	*Plasmodium* exported protein (PHIST), unknown function
7	5	PF3D7_0713900	Conserved *Plasmodium* protein, unknown function
7	3	CRMP2	Cysteine repeat modular protein 2
8	3	CLAG8	Cytoadherence linked asexual protein 8
10	2	PF3D7_1004800	ADP/ATP carrier protein, putative
12	2	PF3D7_1201400	*Plasmodium* exported protein, unknown function
13	7	PF3D7_1301800	Surface-associated interspersed protein 13.1 (SURFIN 13.1)
13	3	PF3D7_1308400	Conserved *Plasmodium* protein, unknown function
14	2	PF3D7_1434500	Dynein-related AAA-type ATPase, putative
14	2	PF3D7_1477500	*Plasmodium* exported protein (PHISTb), unknown function

SNPs and ID stand for single nucleotide polymorphisms and gene identification numbers, respectively

Table 4 Genes under both recent positive directional selection and positive balancing selections in Ethiopian *P. falciparum* populations

Chromosomes	Gene name/ID	Product description
1	PF3D7_0104100	Conserved *Plasmodium* membrane protein, unknown function
1	PF3D7_0113600	Surface-associated interspersed protein 1.2 (SURFIN 1.2)
4	PF3D7_0424300	Erythrocyte binding antigen-165, pseudogene
4	SURF4.2	Surface-associated interspersed protein 4.2 (SURFIN 4.2)
4	PF3D7_0425200	*Plasmodium* exported protein (hyp15), unknown function
4	PF3D7_0425250	*Plasmodium* exported protein (PHIST), unknown function
7	PF3D7_0713900	Conserved Plasmodium protein, unknown function
8	CLAG8	Cytoadherence linked asexual protein 8
10	PF3D7_1004800	ADP/ATP carrier protein, putative
12	PF3D7_1201400	*Plasmodium* exported protein, unknown function
13	PF3D7_1301800	Surface-associated interspersed protein 13.1 (SURFIN 13.1)
13	PF3D7_1308400	Conserved Plasmodium protein, unknown function
14	PF3D7_1434500	Dynein-related AAA-type ATPase, putative

ID stands for a gene identification number

Thirteen (13) out of the above 15 genes under positive directional selection showed both positive balancing and directional selections (Table 4) and these genes include the vaccine candidate gene *SURF4.2* on chromosome 4 and *CLAG8* (cytoadherence-linked asexual protein 8) on chromosome 8 [34]. Interestingly, attempts to detect selection signals in drug resistance genes, such as *Pfcrt*, *Pfmdr1*, *Pfdhfr*, and *Pfdhps* were unsuccessful. The reason could be that IHS may not be suitable for detecting positive selection for those SNPs that have reached or are near fixation in the local *P. falciparum* population [32].

Prevalence of mutations conferring anti-malarial drug resistance in *P. falciparum*

Table 5 shows inter-population differences in the prevalence of drug resistance genes observed among the *P. falciparum* global datasets analysed. In tandem with previous studies [20, 35] that suggest temporal differences in the geographical distribution of anti-malarial drug resistance mutations, in this study CQ-resistance alleles (*Pfcrt-K76T*, *Pfcrt-A220S*, and *Pfcrt-Q271E*) were fixed in Ethiopia, Cambodia, and Thailand, regions where malaria transmission rates are comparably low. In contrast, the

Table 5 Drug resistance-conferring allele's frequency across the 5 *P. falciparum* populations

Genes	Chromosome	Position	Mutation site	Ethiopia	Cambodia	DR Congo	Malawi	Thailand
DHFR	4	748,577	*I164L*	0	0.5	0	0	0.84
DHFR	4	748,410	*S108N*	0	1	1	1	1
DHFR	4	748,262	*C59R*	0.86	1	0.86	0.99	1
DHFR	4	748,239	*N51I*	1	0.95	1	1	0.95
MDR1	5	961,625	*D1246Y*	0	0	0.17	0	0
MDR1	5	958,145	*N86Y*	0.14	0	0.48	0.03	0
MDR1	5	961,566	*F1226Y*	0	0.04	0	0	0.59
MDR1	5	958,440	*Y184F*	1	0.58	0.32	0.35	0.06
CRT	7	405,600	*I356T*	0	0.52	0.27	0	1
CRT	7	405,362	*N326S*	0.98	0.51	0	0	1
CRT	7	405,838	*R371I*	0	0.8	0.71	0	1
CRT	7	404,407	*A220S*	1	1	0.663	0	1
CRT	7	403,625	*K76T*	1	1	0.66	0	1
CRT	7	404,836	*Q271E*	1	1	0.7143	0	1
DHPS	8	549,685	*G437A*	0.08	0.13	0.08	0.01	0
DHPS	8	549,995	*K540N*	0	0.4	0	0	0.03
DHPS	8	549,681	*S436A*	0	0.2	0.11	0.02	0.17
DHPS	8	550,117	*A581G*	0.02	0.4	0.03	0.02	0.82
K13	13	1,726,432	*K189T*	0.2	0	0.17	0.13	0
K13	13	1,725,259	*C580Y*	0	0.36	0	0	0.26

prevalence of these same alleles was 0% in Malawi and ranged from 66 to 72% in DR Congo.

Similarly, drug resistance mutations in *Pfmdr1* (*Pfmdr1-N86Y* and *Pfmdr1-Y184F*) were also variable among populations. For instance, the Ethiopian parasite population showed the presence of 14% *Pfmdr1-N86Y* and 100% *Pfmdr1-Y184F* gene mutations, whereas *Pfmdr1-N86Y*was detected in 48% of DR Congo isolates and in 3% of Malawi's. Also, the *Pfmdr1*-Y184F drug resistance marker was detected in 58% of the *P. falciparum* population in Cambodia, 32% in DR Congo, 35% in Malawi, and 6% in Thailand's parasite isolates.

Sulfadoxine/pyrimethamine drug resistance mutations were also present in *Pfdhfr* and *Pfdhps* genes in all analysed *P. falciparum* populations. The major pyrimethamine resistance-conferring alleles, such as *Pfdhfr-N51I* and *Pfdhfr-C59R*, were also identified in all parasite populations with fixed or near fixation in frequency. *Pfdhfr-S108N* was fixed in other *P. falciparum* populations, except in Ethiopia. The variable prevalence of drug resistance-conferring alleles were also observed in *Pfdhps* (*Pfdhps-S436A*, *Pfdhps-G437A*, *Pfdhps-K540N*, and *Pfdhps-A581G*), for the parent drug sulfadoxine resistance.

In terms of artemisinin resistance, the African population-specific *Pfk13-K189T* mutation was observed in Ethiopia (in 20% of the samples), DR Congo (17%), and Malawi (13%). This mutation was previously identified in African *P. falciparum* populations [20, 35]. As previously reported [8], the validated and most characterized artemisinin resistance-conferring mutation *PfK13-C580Y* was identified in Cambodia (36% of the samples) as well as in Thailand (26%), but not in Africa.

Discussion

The transmission dynamic coupled with the unique history, ecology, and demography of Ethiopia raises interest in the genetics of its parasite population. High-resolution whole-genome SNP data was used to analyse *P. falciparum* parasite genetic diversity in the central region of Ethiopia and compared with similar parasite data from mainland Africa (DR Congo and Malawi) and Southeast Asian parasites, from Cambodia and Thailand. In this analysis, similar MAF across all five parasite populations with over-representation of low frequency (< 5%) variants was observed as previously reported [19, 21]. Interestingly, mean F_{WS} values were significantly higher in the Ethiopian parasite isolates as compared to the other African populations, but not the Southeast Asian parasite populations. F_{WS} is a genome-wide metric that averages heterozygosity across the genome in comparison with heterozygosity within the local parasite population [21]. Hence, it is a measure of within-host diversity of infections that helps to gauge the potential for inbreeding (or outcrossing). The higher F_{WS} values (> 0.95) in Ethiopia (*P. falciparum* prevalence of 0.02) [36] and East Asian

infections is underscored by the low malaria transmission rates in these settings which supports a higher inbreeding and clonal propagation of infections (Fig. 1; Additional file 3). Unlike the other East African countries (DR Congo and Malawi) where transmission intensities are higher [20, 35], and there was a good distribution of F_{WS} values with the majority of infections being polyclonal with high potential for outcrossing (Fig. 1). These findings are supported by similar studies that link lower F_{WS} values to in west African where transmission is high [18]. However, of note, is the possibility for high F_{WS} values to occur in areas of high transmission intensity if *P. falciparum* circulates in a geographically isolated community which limits the chance of outcrossing with other genetically distinct *P. falciparum* parasites as observed in the previous study [21].

An analysis of parasite population structure within and between continents revealed a higher degree of population structure between Ethiopian parasites and other East African parasites and between Southeast Asia and East Africa. However, neither PCA (Fig. 2) nor admixture analysis (Fig. 3) could resolve parasite populations in DR Congo and Malawi. These observations are corroborated by several studies that report regional and inter-continental level structure in global *P. falciparum* parasite populations [21]. However, the separation of Ethiopian parasites from the two East African populations is worth noting. Notwithstanding the increased human mobility between Addis Ababa and the rest of Africa, particularly East Africa, there remain important barriers to gene flow between parasite populations in central Ethiopia and the rest of the sub-region. Indeed, one possible factors that severely limit gene flow between Ethiopia and its neighbours is the local malaria transmission intensity as a function of poor vectoral capacity determined by the ecological landscape (highlands).

Against the backdrop of this unique eco-epidemiology of *P. falciparum* malaria in Ethiopia, Tajima D and IHS was used to explore the mechanisms of natural selection in the country. However, identification of many antigenic genes under balancing selection with Tajima D value greater than one were observed in Ethiopia. These genes included known vaccine candidates, such as *ama1*, *trap*, *msp1*, *eba175*, and *clag2* (Additional file 4), which were previously identified in different populations that vary in transmission intensity [24, 33, 37], to be under balancing selection. Besides, 15 genes under positive directional selection by IHS were identified, which includes *SURFIN* and *PHIST* families previously suggested to be targets of immunity [24]. It can be hypothesized that the low seasonal transmission in Ethiopia maintains significant immune selection pressure on the infection reservoir than drug pressure due to clinical malaria. Therefore, the candidate vaccine antigen loci under balancing selection may be largely due to immune modulation and not positive adaptive selection influenced by drug pressure. This is supported by our failure to detect selection signatures in known drug target genes such as *Pfcrt*, *Pfmdr1*, *Pfdhfr*, *Pfdhps*, and *Pfkelch-13*. The ability of IHS to detect selection in these drug resistance genes in Ethiopia may be because the frequency of polymorphisms in these loci are either fixed or near fixation in the Ethiopian population (Table 5). These findings are supported by a previous study in Ethiopia which showed that the CQ-resistant haplotype (CVIET) was fixed [7]. The continued use of CQ in Ethiopia for the treatment of *P. vivax* malaria may account for the high prevalence of CQ resistant markers. Also, *Pfmdr 1* mutations have been demonstrated to mediate AL resistance. Therefore, the high prevalence of *Pfmdr 1* mutations may signal poor efficacy of AL as the first treatment for *P. falciparum* malaria in Ethiopia. Variable prevalences of CQ-resistant polymorphisms were observed only in DR Congo and not in Malawi, evidence that supports the complete reversal of CQ susceptibility in Malawi as reported by Ochola et al. [20].

Undoubtedly, artemisinin resistance has taken root in Southeast Asia. Despite 36% and 26% prevalence of *PfKelch13-C580Y* mutation in Cambodia and Thailand samples, respectively, no validated *PfKelch13* mutation was found in the African samples. However, an uncharacterized *Pfkelch13* mutation (*PfK13-K189T)* found at prevalence > 10% in all the African datasets and previously reported in other studies [35], may be important, but its role in artemisinin resistance is unknown. One study [8] reported that mutation in *Pfkelch13* at amino acid positions less than 441 may not play any role in mediating artemisinin resistance. A validated *Pfkelch13-R561H* mutation for artemisinin resistance was recently reported in other East African *P. falciparum* populations [9].

Conclusion

Overall, this study reveals the presence of a comparably low genetic diversity of *P. falciparum* parasites in Ethiopia. The majority of infections were of low complexity, demonstrated significant population structure with Ethiopian parasites diverged from parasite populations within the sub-region. Based on the analysis made it is suggested the presence of limited gene flow between parasite populations in the East African sub-region and Ethiopia. More importantly, apparent balancing selection in antigenic loci known to be targets of immunity and adaptive positive selection in *SURFIN* and *PHIST* gene families that are potential vaccine antigens. Though selection analysis did not pick up any adaptive mutations in known drug-resistant genes, CQ-resistance *Pfcrt-K76T* genotype

seems fixed in Ethiopia like the wild-type genotype (*K*) in Malawi. In this analysis no *PfKelch13* validated mutations were reported in Ethiopia, DR Congo, and Malawi except a *PfK13-K189T* African specific uncharacterized mutation. Further molecular studies involving deeper sampling of Ethiopian parasite populations are essential to understand the genetic diversity, gene flow, and temporal evolution of drug resistance loci within Ethiopia. Furthermore, such findings can be used to support national malaria control decision-making for optimal impact in further reducing malaria transmission in Ethiopia.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12936-021-03660-y.

Additional file 1. Summary of sequence reads and ENA accession number per isolate included in this study.

Additional file 2: Figure S1. Distribution of the number of SNPs across all analyzed genes and their respective SNPs type distribution. **Figure S2.** Minor allele frequency distribution by population. SNPs are binned into 10 equal sizes of 0.05. In all parasite populations, there is an overabundance of low-frequency SNPs (MAF < 5%).

Additional file 3. F_{WS} values of samples in *P. falciparum* populations analyzed in this study.

Additional file 4. Genes that had ≥ 5 SNPs and a Tajima D of ≥ 1 that are defined as genes under positive balancing selection.

Abbreviations

ACT: Artemisinin-based combination therapy; AL: Artemether–lumefantrine; ART: Artemisinin; CQ: Chloroquine; FWS: Within infection diversity fixation index; FST: Population differentiation fixation index; HW: Heterozygosity within infection; HS: Heterozygosity within a local population; IHS: Standardized integrated haplotype score; IRS: Indoor Residual Spraying; ITN: Insecticide-treated nets; *Pfcrt*: *Plasmodium falciparum chloroquine*; *Pfdhfr*: *Plasmodium falciparum dihydrofolatereductase*; *Pfdhps*: *Plasmodium falciparum pteroate synthase*; *PfK13*: *Plasmodium falciparum kelch-13*; SNP: Single nucleotide polymorphisms; SP: Sulfadoxine/pyrimethamine.

Acknowledgements

Plasmodium falciparum genome sequencing was done at the Wellcome Sanger Institute as part of the MalariaGEN *P. falciparum* Community Project (www.malariagen.net/projects). We thank the MalariaGEN *P. falciparum* Community Project and Pf3K Project for allowing access to sequence data. JLB is supported by DELTAS Africa Initiative grant # DEL-15-011 to THRiVE-2. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust grant # 107742/Z/15/Z and the UK government. The views expressed herein do not necessarily reflect the official opinion of the donors.

Authors' contributions

DA participated in study design, data analysis, interpretation, and writing manuscript, CK, TD, LAE, and JLB reviewed the manuscript, and LG participated in data collection, DNA extraction, genome sequencing, and manuscript reviewing. All authors read and approved the final manuscript.

Funding

This project was funded by the Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME). DELGEME is a training program sponsored by the Welcome Trust Developing Excellence in Leadership, Training and Science Africa (DELTAS Africa) initiative in partnership with the Department of International Development (DFID) and the Alliance for Accelerating Excellence in Science in Africa (AESA).

Availability of data and materials

Datasets generated and/or analysed during the study are available through the MalariaGEN Pf3K Project. The *P. falciparum* genome sequences used in this study are available in the ENA and SRA databases (see Additional file 1 for accession numbers).

Ethics approval and consent to participate

The use of human subjects and scientific merit for this study was approved by the institutional Ethical Review Board (IRB) and the scientific committee of Addis Ababa University and AHRI-ALERT (Armauer Hansen Research Institute and the Africa Leprosy Rehabilitation and Training Hospital). Written informed consent was obtained from all adult subjects and the parent or legal guardians of minors.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Department of Biochemistry, College of Health Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya. [2] School of Medical Laboratory Sciences, Faculty of Health Sciences, Jimma University, Jimma, Ethiopia. [3] International Centre of Insect Physiology and Ecology, Nairobi, Kenya. [4] West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), University of Ghana, Accra, Ghana. [5] Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia.

Received: 9 September 2020 Accepted: 20 February 2021
Published online: 04 March 2021

References

1. WHO. World malaria report 2019. Geneva: World Health Organization; 2019.
2. Lo E, Hemming-Schroeder E, Yewhalaw D, Nguyen J, Kebede E, Zemene E, *et al*. Transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum* in Ethiopia and prevalence of antimalarial resistant genotypes. PLoS Negl Trop Dis. 2017;11:e0005806.
3. Abdifatah AJ, Wanna C, Kesara N-B. *Plasmodium falciparum* drug resistance gene status in the Horn of Africa: a systematic review. Afr J Pharm Pharmacol. 2018;12:361–73.
4. Amambua-Ngwa A, Jeffries D, Amato R, Worwui A, Karim M, Ceesay S, *et al*. Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from The Gambia. Sci Rep. 2018;8:9687.
5. Mekonnen SK, Aseffa A, Berhe N, Teklehaymanot T, Clouse RM, Gebru T, *et al*. Return of chloroquine-sensitive *Plasmodium falciparum* parasites and emergence of chloroquine-resistant *Plasmodium vivax* in Ethiopia. Malar J. 2014;13:244.
6. National five-year strategic plan for malaria prevention & control in Ethiopia. https://extranet.who.int/countryplanningcycles/sites/default/files/planning_cycle_repository/ethiopia/five_year_malaria_control_sp_2006_2010.pdf. Accessed 20 Dec 2019.
7. Golassa L, Enweji N, Erko B, Aseffa A, Swedberg G. High prevalence of *Pfcrt-CVIET* haplotype in isolates from asymptomatic and symptomatic patients in South-Central Oromia, Ethiopia. Malar J. 2014;13:120.
8. Ashley E, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, *et al*. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. N Engl J Med. 2014;371:411–23.
9. Uwimana A, Legrand E, Stokes BH, Ndikumana JM, Warsame M, Umulisa N, *et al*. Emergence and clonal expansion of in vitro *Pfkelch13 R561H* mutant parasites in Rwanda. Nat Med. 2020;26:1602–8.
10. Loy DE, Liu W, Li Y, Learn GH, Plenderleith LJ, Sundararaman SA, *et al*. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. Int J Parasitol. 2017;47:87–97.

11. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, *et al*. A selective sweep driven by pyrimethamine treatment in Southeast Asian malaria parasites. Mol Biol Evol. 2003;20:1526–36.
12. Volkman SK, Sabeti PC, Decaprio D, Neafsey DE, Schaffner SF, Milner DA, *et al*. A genome-wide map of diversity in *Plasmodium falciparum*. Nat Genet. 2007;39:113–9.
13. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, *et al*. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. Nat Commun. 2016;7:11078.
14. Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, *et al*. Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. PLoS Genet. 2013;9:e1003293.
15. Paget-Mcnicol S, Saul A. Mutation rates in the *dihydrofolate reductase* gene of *Plasmodium falciparum*. Parasitology. 2001;122:497–505.
16. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, Zhou Y, *et al*. Erratum: A systematic map of genetic variation in *Plasmodium falciparum*. PLoS Pathog. 2006;2:e57.
17. Oyebola KM, Idowu ET, Olukosi YA, Awolola TS, Amambua-Ngwa A. Pooled-DNA sequencing identifies genomic regions of selection in Nigerian isolates of *Plasmodium falciparum*. Parasites Vectors. 2017;10:320.
18. Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, *et al*. Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. BMC Genom. 2015;16:527.
19. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, *et al*. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. Mol Biol Evol. 2014;31:1490–9.
20. Ocholla H, Preston MD, Mipando M, Jensen ATR, Campino S, Macinnis B, *et al*. Whole-genome scans provide evidence of adaptive evolution in Malawian *Plasmodium falciparum* isolates. J Infect Dis. 2014;210:1991–2000.
21. Manske M, Miotto O, Campino S, Auburn S, Zongo I, Ouedraogo J, *et al*. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. Nature. 2012;487:375–9.
22. Mackinnon MJ, Read AF. Virulence in malaria: an evolutionary viewpoint. Philos Trans R Soc B Biol Sci. 2004;359:965–86.
23. Biswas S, Akey JM. Genomic insights into positive selection. Trends Genet. 2006;22:437–46.
24. Amambua-Ngwa A, Park DJ, Volkman SK, Barnes KG, Bei AK, Lukens AK, *et al*. SNP genotyping identifies new signatures of selection in a deep sample of west African *Plasmodium falciparum* malaria parasites. Mol Biol Evol. 2012;29:3249–53.
25. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, *et al*. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. Science. 2019;365:813–6.
26. Jalei AA, Chaijaroenkul W, Na-Bangchang K. *Plasmodium falciparum* drug resistance gene status in the Horn of Africa : a systematic review. Afr J Pharm Pharmacol. 2018;12:361–73.
27. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, *et al*. An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole-genome high-throughput sequencing. PLoS ONE. 2011;6:e22213.
28. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11:94.
29. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.
30. Tajima F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
31. Gautier M, Vitalis R. Rehh 2.0: a reimplementation of the R package *rehh* to detect positive selection from haplotype structure. Mol Ecol Resour. 2017;17:78–90.
32. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLos Biol. 2006;4:e72.
33. Ye R, Tian Y, Huang Y, Zhang Y, Wang J, Sun X, *et al*. Genome-wide analysis of genetic diversity in *Plasmodium falciparum* isolates from China-Myanmar border. Genetics. 2019;10:1065.
34. Iriko H, Kaneko O, Otsuki H, Tsuboi T, Su XZ, Tanabe K, *et al*. Diversity and evolution of the rhoph1/clag multigene family of *Plasmodium falciparum*. Mol Biochem Parasitol. 2008;158:11–21.
35. Verity R, Aydemir O, Brazeau NF, Watson OJ, Hathaway NJ, Mwandagalirwa MK, *et al*. The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the Democratic Republic of Congo. Nat Commun. 2020;11:2107.
36. https://www.internationalscholarsjournals.org/print.php?article=analysis-of-malaria-prevalence-inn.
37. Amambua-Ngwa A, Tetteh KKA, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, *et al*. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS Genet. 2012;8:e1002992.

Publisher's Note