ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

INSTITUTE OF BIOTECHNOLOGY

GENETIC DIVERSITY AND NATURAL SELECTION OF A MALARIA VACCINE CANDIDATE GENE IN THE ETHIOPIAN *PLASMODIUM VIVAX* POPULATION

A Thesis submitted to the School of graduate studies of Addis Ababa University, Institute of Biotechnology, in partial fulfillment of the requirements for the Degree of Masters of Science in Biotechnology

By: - Alebachew Messele

December, 2017

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

INSTITUTE OF BIOTECHNOLOGY

GENETIC DIVERSITY AND NATURAL SELECTION OF A MALARIA VACCINE CANDIDATE GENE IN THE ETHIOPIAN PLASMODIUM VIVAX POPULATION;

A Thesis Submitted to Addis Ababa University, Institute of Biotechnology in partial fulfillment of the Degree of Masters of Science in Biotechnology

By: - Alebachew Messele

Advisors;

Adey Feleke (PhD), Institute of Biotechnology, AAU
Lemu Golassa (PhD), Aklilu Lemma Institute of Pathobiology
Endalamaw Gadisa (PhD),Armauer Hansen Research Institute
Fitsum Taddese (MSc, PhD fellow), Institute of Biotechnology, AAU,

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

INSTITUTE OF BIOTECHNOLOGY

This is to certify that the thesis submitted by Alebachew Messele entitled: "Genetic Diversity and Natural Selection of a Malaria Vaccine Candidate Gene in the Ethiopian *Plasmodium vivax* Population;" submitted in partial fulfillment of the requirements for the Degree of Master of Science in Biotechnology complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

_____ _____ _____

Dr. Tesfaye Sisay (Examiner) Signature Date

_____ _____ _____

Dr. Tileye Feyissa (Examiner) Signature Date

_____ _____ _____

Dr. Adey Feleke (Advisor) Signature Date

_____ _____ _____

Dr. Lemu Golassa (Advisor) Signature Date

_____ _____ _____

Dr. Endalamaw Gadisa (Advisor) Signature Date

_____ _____ _____

Mr. Fitsum Girma (Advisor) Signature Date

_____

Chair of Department or Graduate Program Coordinator

Dedication

For my Mom Mrs. Aster Zewdie and my Dad Mr. Messele Kebede

‘’And if the wine you drink, the Lip you press,
End in the Nothing all Things end in-Yes-
Then fancy while Thou art, Thou art but what
Thou shalt be – Nothing – though shall not be less.’’

Edward FitzGerald, Rubaiyat of Omar Khayyam

Acknowledgment

It's been a quite arduous journey to arrive at this point. All of which wouldn't have been possible had it not been for the help and support of the individuals I am about to mention. I have been privileged enough to have worked with not one but four exceptional advisors. I would like to extend my utmost gratitude to Dr. Adey Feleke at Addis Ababa University, Institute of Biotechnology for all the invaluable inputs into my study as well as helping me in acquiring primers for this project. I am indebted to Dr. Endalamaw Gadissa for his assistance starting from proposal development, all the way to supervision of my laboratory activities and finally for critically reviewing my thesis. I express my thanks and appreciation to Mr. Fitsum Girma, for mentoring and supervising my laboratory activities, for orchestrating the logistics of the study from start to finish, and for his major contribution during draft write up of the thesis. It was truly an honor working with Dr.Lemu Golassa. I would also like to say thank you to him for giving me the chance to work on this project and for his detailed and critical review during draft write-up of this thesis. Finally I thank my advisors for their invaluable comments and suggestions which have fast-tracked my progress be a better researcher overall.

I am not indebted only to these contributors but also many other individuals, such as Mrs. Meseret Abebe for helping me out in the laboratory. My sincere thanks also go to the entire staff of the Armauer Hansen Research Institute for their support and cooperation. I would also like to thank the Institute of Biotechnology at Addis Ababa University for giving me this platform to bring forth a pressing issue in Ethiopia. In particular I thank Dr. Tesfaye Sisay and Dr. Kassahun Tesfaye for their support throughout.

It would have been impossible to finish this project without the support and guidance of my family, specifically my dad; there aren't enough words in the English language to express my gratitude so I will simply resign to thank you. My brother and sisters thank you for setting the standard of excellence, for being supportive and an inspiration to follow.

Table of Contents

List of Tables

List of Figures

List of Appendices

Acronyms and Abbreviation

ACT:	Artemisinin combination therapy
AHRI:	Armauer Hansen Research Institute
AL:	Arthemether Lumifantrine
AMA1:	Apical membrane antigen 1
BLAST:	Basic local alignment search tool
CQ:	Chloroquine
CQR:	Chloroquine resistant
CRT:	Chloroquine resistance transporter
CSP:	Circumsporozoite
DARC:	Duffy antigen receptor for chemokines
DBP:	Duffy binding protein
DBS:	Dried blood spot
G6PDd:	G6PD deficient
GPI:	Glycosylphophatidylinositol
HRP2:	Histidine rich protein 2
IRS:	Indoor residual spray
ITN:	Insecticide treated net
LAMP:	Loop amplified isothermal amplification
LLIN:	Long lasting insecticidal net

MCL: Maximum composite likelihood

MDR-1: Multidrug resistant 1

MCI: Multiple clone infections

MSP: Merozoite surface protein

PBS: Phosphate buffer saline

PF: *Plasmodium falciparum*

Pfcrt: *Plasmodium falciparum* chloroquine resistance transporter

PNG: Papua New Guinea

PQ: Primaquine

PV: *Plasmodium vivax*

Pvmdr1: *Plasmodium vivax* multidrug resistant 1

PvMSP: *Plasmodium vivax* merozoite surface protein

RBP: Rhoptry binding protein

RDT: Rapid diagnostic test

RFLP: Restriction fragment length polymorphism

SNP: Single nucleotide polymorphism

SP: Sulfadoxine Pyrmethamine

TRAP: Thrombospondin related adhesion protein

UPGMA: Unweighted Pair Group Method with Arithmetic mean

WHO: World Health Organization

Abstract

Genetic Diversity and Natural Selection of a Malaria Vaccine Candidate gene in the Ethiopian *Plasmodium vivax* Population

Alebachew Messele

Addis Ababa University, 2017

The burden of P.vivax in Ethiopia is amongst the highest in the world. However, P. vivax diversity, particularly that associated with antigens, such as Plasmodium vivax Merozoite surface protein 3α (PvMSPα), has rarely been studied in Ethiopia. These studies are fundamental for tracking important phenotypic variants of the parasite such as drug resistance genes or antigenic variants in different transmission settings, as this aid in designing future control and elimination strategies. *In the present study the genetic polymorphism in the defined target was assessed by examining genes encoding two blocks of this antigen locus. Finger prick blood samples spotted onto filter papers were collected from microscopically and RDT confirmed malaria patients attending health facilities of Shewa Robit, Melka Oda, Abure, Aje Haposto and Ilala. DNA was extracted by Chelex Saponin extraction method, the genomic DNA was used for confirmation of P.vivax infection by targeting the 18S rRNA gene. Positive samples were subsequently evaluated by Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) for identification and assessment of the genetic polymorphism of the MSP3α gene. Further single clone infections were then analyzed using Sanger sequencing. Three size variants were amplified from the 50 isolates, Type A, B and C with frequencies of 82.97%, 12.7% and 4.2% respectively. Further details of diversity were attained by Hha I RFLP, with 11 alleles and 12% multiple clone infections. The sequence analysis showed that size polymorphisms were results of insertions and deletions in the block I component of the gene, which also had higher nucleotide diversity (π) (0.10565) than the block II (0.014). The relatively conserved block II was evolving under positive selection, but a select region that encodes a predicted B cell epitope in these block is under balancing selection (Tajima's D 2.64 (P>0.05), Fu and Li s F 1.7621 (P>0.05); Furthermore a peak diversity was recorded at this site (π=0.65) with low inter-population F_{ST} estimates . The conserved nature of PvMSP3α block II makes it an ideal vaccine candidate. However future vaccine design strategies targeting PvMSP3α block II should put into consideration the identified antigenic polymorphism from this study, as they might constitute an immune/vaccine escape mechanism. In contrast the polymorphic nature of PvMSP3α block I make it more suited for use as a rapid genotyping tool.*

Key words; *Plasmodium vivax; Merozoite surface protein 3α; PCR-RFLP; Multiple clone infections; Sequence polymorphism; Balancing selection; Antigenic polymorphism; Ethiopia*

1. Introduction

1.1 Background

Malaria is an infectious disease mainly caused by four *Plasmodium* species: *P. falciparum, P. vivax, P. ovale* and *P. malariae.* Quite recently, however, *P. knowlesi* has also been indicated(1). Malaria is one of the leading causes of morbidity and death worldwide with 212 million cases and 429,000 deaths in 2015 alone(2). Of these reported deaths, 92% were from the African continent, specifically Sub Saharan Africa which carry much of the disease burden 92% (3). Two of the species, *P. vivax* and *P. falciparum* are the prominent causes for much of the infection and carnage reported in malaria. *P. falciparum* is responsible for the majority of these fatalities, whereas, *P. vivax* is the most geographically widespread of the *Plasmodium* species putting 2.85 billion people at risk(4). In sharp contrast, *P. vivax* is largely absent from Africa, with a significant exception in countries such as, Ethiopia, Eritrea, and south Sudan (5,6).

In east Africa, Ethiopia holds a superior geographic presence and a large population number; of which, three quarter of the country where 67% of the population lives in, is malarious (7). Here it is reported that *P. falciparum* and *P. vivax* account for 66% and 34% of the malaria infections, respectively(8). Malaria transmission in Ethiopia is highly unstable and heterogeneous owing to the varying climate and altitude, and accordingly the country is classified into four eco-epidemiological zones; the malaria free highlands (above 2500 m), the highland fringe areas (1500-2500m) affected by frequent epidemics, lowland areas (below 1500m) with seasonal transmission and finally areas with stable transmission all year round regardless of elevation (9). In most areas of Ethiopia, malaria transmission has a seasonal pattern following the two rainy seasons; the major transmission season is from September to December following the major rainy season of July to September , and the minor transmission season is from April to May following the short rainy season from February to March (10,11).

Depending on seasonal patterns, lowland areas are frequently afflicted by *P. falciparum* infections whereas *P. vivax* in comparison is prevalent in highland areas (12). However, in most parts of Ethiopia, *P. vivax* and *P. falciparum* co-exist albeit difference in

proportion. A common theme in the past was that during high malaria transmission season/epidemics, *P. falciparum* infections dominate, whereas in the low transmission period, *P. vivax* prevails (13). Moreover, a shift in species prevalence is quite common in most areas of Ethiopia (14). Recently, studies have indicated a new precedent, a consistent rise of *P. vivax* prevalence in different parts of the country regardless of seasonal patterns or eco-epidemiological zones. Three independent studies in particular reported a prevalence of 41.6% in the south western part of the country (15); 45.2% in the regions including Amhara, SNNPR, Tigray (16); and 76.4% in the Oromia Region (17). In congruence, Ethiopia was also one of the four countries in the world whom together contributed to 73% of global *P. vivax* infection burden (8). This rise of *P. vivax* in Ethiopia has largely been accredited to the national malaria control and prevention strategies which have placed huge emphasis on *P. falciparum*, while neglecting *P. vivax* infections (18). In most of Africa where *P. falciparum* is the sole source of malaria, control and prevention programs have indeed reduced case incidence by as much as 40% from the years 2000-2015 (19). On the contrary, Ethiopia is a country which is co endemic for both *Plasmodium* species; as such the ground gained from reducing *P. falciparum* is under immediate threat from *P. vivax* infections.

P. vivax, even with the prospect of proper attention and disease management is quite hard to control, and this is down to several factors. However, the most notable feature of the *P. vivax* with respect to control measures is its dormant liver stage or hypnozoite stage which has been attributed to relapses (20). Hypnozoites are primary source of relapses; re-occurrence of *P. vivax* infections weeks, months or years after the primary infection had subsided. This dormant forms remain a major obstacle in attempts to control the parasite, as current drug regiments for *P. vivax* don't affect this stage and those that do such as Hypnozoiticidal drugs (Primaquine, Tafenoquine) are contraindicated in G6PD deficient individuals where they cause hemolysis and are therefore prohibited from use (21). Furthermore there are no diagnostic procedures capable of detecting this stage of the parasite, worryingly then is the fact that the hypnozoite forms contribute to up to 80% of all *P. vivax* blood stage infections (22). Naturally, there has been growing concern about the lack of chemo-preventive strategies, but even researches on *P. vivax* are severely limited. A major reason behind this is because *P. vivax* peculiarly favors invading

reticulocytes; which are notoriously hard to grow continuously in in-vitro cultures, therefore delaying any progress in research (23).

Evidently due to the lack of progress in studying the parasite, the use of alternative analysis methods such as population genetic analysis, which provide a simpler and less expensive route have grown popular (24). For instance population genomic studies are used to investigate patterns of transmission dynamics and population structure which are important indicators for control and elimination programs (25). Interestingly, these studies have revealed that *P. vivax* is quite diverse as compared to *P. falciparum* (26,27). Apart from studying the extent and distribution of diversity of the parasite, however, studies like this are also used to design future preventive strategies. For instance, they are applied in designing malaria vaccines.

Indeed, vaccines provide the missing link in malaria prevention/cure, by ‘’filling in the gaps’’ left by current vector and chemo-preventive regimens. For instance, pre-erythrocytic vaccines could block entry of the invasive sporozoites to the liver, effectively preventing formation of hypnozoites, or transmission blocking vaccines that would prevent spread of the hypnozoite to an onward transmission (28). However, the extent and distribution of antigenic diversity and the strain dependent manner in which humans develop immunity, present a colossal challenge and has hindered any progress toward a broadly efficacious vaccine (29). Particularly for *P. vivax* it has been shown to be more prevalent, more diverse and less structured than *P. falciparum* making it harder to study (27). Previously molecular biology , biochemistry and immunological methods have been used to identify potential vaccine candidates i.e. antigens (29), The use of population genetics however has fast tracked efforts and enabled to identify antigenic diversity pertinent to immunogenicity i.e. those involved in immune escape (24). This has allowed to narrow the list of potential vaccine candidates as well as to minimize costly analysis (30).

Currently blood stage vaccine candidates take precedent in pre-clinical efficacy tests. That is mainly because clinical symptoms are evident during this stage and antigens produced at this stage represent an opportunity for prevention of this invasive stage (29,31). As such merozoites which have been implicated in invasive stage of

erythrocytes are obvious targets (32). Accordingly several surface proteins of the *P. vivax* merozoite have been identified as ideal targets. Merozoite surface protein 3α (PvMSP3α) is one of those vaccine candidates. The surface protein has also previously been used as rapid population genetic marker, genotyping *P. vivax* isolates via a Polymerase chain reaction-Restriction fragment length polymorphism (PCR-RFLP) procedure(32–38). The dual potential of this gene is down to its two contrasting blocks. Block I is a relatively polymorphic portion of the gene, complete with large number of insertion and deletions; which result in shifts in recognition sites of restriction enzymes ,and lay the basis for the polymorphisms visualized as restriction patterns (32,33). In contrast, block II region of the merozoite surface protein 3α is relatively conserved worldwide, and has been shown to trigger pronounced antibody response in clinical infections in two *P. vivax* endemic countries (39,40,41).

In Ethiopia studies concerning *P. vivax* have so far been limited to drug resistance, submicroscopic infections, diagnostic studies and G6PD deficiency (42, 43, 44, 45, 46). Studies concerning *P. vivax* molecular phylogenetics, transmission dynamics and population structure have so far been few (43,47,48). Even worse the population genetics of *P. vivax* vaccine candidate antigens have not been assessed at all. Studying the molecular evolution responsible for variations, by analyzing the nucleotide diversity of *P. vivax* merozoite surface protein 3α, enables to guide rational vaccine design and to identify immune escape mechanisms of *P. vivax* in this antigenic locus. Furthermore, this antigen locus is also used as a rapid genotyping tool for population genetic studies due to its polymorphisms. However, there are only a handful of studies that have formally assessed reasons behind its variability; henceforth interpretations of results have been limited to observing the number of allele's present and comparing RFLP patterns. Interpretations of results are further complicated by multiple insertion deletion events and/or convergent evolution which can give rise to alleles having similar RFLP patterns but different at the sequence level. Here, using Ethiopian isolates, the complete sequence of the PvMSP3α is used to study both blocks of the genes with their respective use as a vaccine candidate antigen and a molecular marker.

1.2 Significance and justification of study

The aim of the study is to generate and analyze sequence diversity of PvMSP3α, a potential vaccine candidate and molecular marker using samples collected from clinical patients. By studying the extent and distribution of polymorphisms in this gene, we will gain an understanding of the mechanisms and reasons behind variation patterns and doing so we will be able to define the reasons behind diversity of the candidate gene, enabling us to predict polymorphisms that contribute to its genetic and antigenic diversity which are pertinent for future vaccine design studies; it will also allow us to get a picture of the geospatial distribution of the predicted haplotypes. The data generated here will be compared to other published global sequences of the gene with respect to rational vaccine design. Furthermore, the gene has also been of interest regarding its potential as a possible cost effective and simple molecular marker to track and analyze genetic diversity with respect to *P. vivax*. In Ethiopia, the high prevalence, morbidity and high diversity of *P. vivax* has been documented(15). However, the parasite is still considered as a non-African problem. Despite this, Ethiopia remains a major contributor to the morbidity of this parasite population globally. The emission of such facts risks the underestimation of the Ethiopian *P. vivax* parasite population and ultimately the global population. The repercussions will be dire as future prevention and intervention programs would risk the efficacy and proper management of the parasite.

Population genetic studies are cheaper and valuable tools in surveillance of diversity and its effect on several control measure undertaken to prevent and treat malaria. One of the control measures is development of vaccines. Unlike previous pre-clinical efficacy studies, population genetic studies are very important in predicting vaccine resistant strains before the actual administration of the vaccine, making such studies invaluable in reducing costs; furthermore, such studies help in determining which region is under immune selection and which strain would be most effective for the strategy. While Ethiopia is endemic for *P. vivax*, it has so far been exempted from studies of vaccine efficacy. The main reason being the underestimation of the prevalence of the parasite, and because of the global narrative that dictates *P. vivax* as a non-African problem. While studies should include samples from all representative *P. vivax* populations, the above

mentioned facts inhibit the inclusion of Ethiopian *P. vivax* samples for consideration in future studies. The risk of doing so could derail the study as well as the future use of vaccines in Ethiopia. The strain specific nature of the malaria vaccine and the high nucleotide diversity in combination could alter the efficacy of any vaccine, the main reason being that administering vaccines without prior information could risk the selection of vaccine resistant strains.

The current road map is to develop vaccines which have 75% efficacy against severe and mild malaria by the year 2030. Ethiopia, while on track to meet its millennium development goals against *P. falciparum*, still requires giving due attention to *P. vivax*, which cannot be eliminated by using the same techniques as *P. falciparum* largely due to its peculiar nature. While drug resistance and insecticide resistance are on a high, the need for alternative strategies has become evident. *P. vivax* merozoite surface protein 3 alpha block II is a potential blood stage vaccine candidate, while population genetic studies have conclusively shown the conserved region throughout global sequences and its ability to trigger the immune system in areas endemic to the parasite, that study has not been extended to endemic parts of the African continent, specifically Ethiopia. Therefore, the study aims to cover this gap; particularly, it aims to identify polymorphism of this antigen, explore the extent and distribution of diversity, identify polymorphisms under immune selection, construct haplotypes, measure haplotype frequencies and finally perform network analysis of the haplotypes with worldwide sequences. In addition, validation of the gene as a molecular marker will be elucidated by comparing In-silico and gel RFLP products, to determine if the cost effective marker covers the extent and distribution of nucleotide diversity.

2. Literature Review

2.1 *Plasmodium vivax* biology

The peculiar nature of *P. vivax* has remained a mystery for much of the quest to identify its biology. Furthermore, researches have been focused on its much fatal relative *P. falciparum*. However, current studies have elucidated several facts such as its severity (49) fatality (50) and the extent of its diversity (26). This is in addition to the threats of drug resistance (51) and relapse (52) have brought *P. vivax* to the forefront of studies investigating prevention, control and ultimate elimination strategies for malaria.

Worldwide, it is estimated that 2.5 billion people are at risk, with annual infections in the range of 80-300 million cases(53). Asia and the Americas present up to 70% of this infections(54), and as a result *P. vivax* is labeled as a non-African problem. The near fixation of Duffy antigen receptor for chemokines (DARC) negativity i.e. absence of Duffy receptor, in the African continent has been pointed out as the main reason for the near absence of *P. vivax.* Presumably DARC is the only receptor used during invasion of reticulocytes(23). This concept has however been challenged, Duffy negative *P. vivax* patients have been reported in several countries(55,56). Furthermore, the prevalence of *P. vivax* in countries where Duffy negativity has reached 95% fixation has led to suggestions that it doesn't confer resistance to the infection (57) .Interestingly, it also might mean that the Duffy receptor is not the only receptor involved in invasion. In support of this argument it has been suggested that since *P. vivax* strongly favors reticulocytes and not mature red blood cells which also express the Duffy antigen, other membrane receptors might be involved(23).

The strong preference of reticulocytes which represent 1-2% of circulating red blood cells with a 24 hour life span has also perhaps enabled it to control hyper-parasitemia which are very rare for *P. vivax* infections (23,53). Besides the strong inclination for reticulocytes, *P. vivax* have several unique characteristics, many of which have enabled the parasite to thrive in wider geographical ranges and remain less amenable to control measures (58). Perhaps the most eminent characteristic is its ability to form dormant hypnozoites, which are dormant liver stage forms able to remain inactive for weeks,

months or even years upon which they can cause relapses. The relapsing periodicity of *P. vivax* is attributed as a survival tool as a way to ‘’hibernate’’ between seasonal changes, as studies have shown that this event coincides with vector emergence (20). This would indicate that relapses are not random, in addition stress and the illness itself has been described as triggers for activation of the hypnozite (59). Relapses pose a unique challenge for malariologists, primarily hynozoites are quite distinct from the clones that caused the first infection (60), allowing them to have low antigenic profile which in turn enables them to infect the individual even if the patient had received medication. Relapses also serve as a source of gametocytes, and are responsible for the minimal population structuring and high diversity observed globally (61,62). This is mainly because of the possibility of distant dissemination of parasite strains by mobile humans (63). Another important unique feature of *P. vivax* is its gametocyte, which sometimes appears before clinical symptoms emerge (53).This feature also has strong implications, the slow progression towards drug resistance as compared to *P .falciparum* is due to this aspect (49). The presence of the trophozoite, shizont and gametocyte in the blood at the same time and the presence of caveole-vesicles are further distinctive features of *P. vivax* under the microscope (53). Caveole vesicles particularly form what is known as the Shuffners’ dots, very distinctive mass of pink dots, a result of Giemsa staining (64). *P. vivax* is also characterized by a higher infectivity to mosquitoes and a shorter development cycle in the vector (54). Its clinical symptoms include, severe anemia, thrombocytopenia, rigor, and fevers reaching 41°C (54,65).

2.1.2 Overview of the lifecycle of *Plasmodium vivax*

The lifecycle of *P. vivax* (Figure 1) inside its human host begins when an anopheles mosquito injects into the skin of its host and releases sporozoites. Experiments have shown that sporozoites are able to remain in the skin for up to 6 hours, however most sporozoites migrate to the liver within minutes after inoculation (22,66,).The sporozoites move by gliding motility at 1-2µm $s^{-1,}$ taking either the lymphatic or blood vessel route (67,68). They encounter several barriers of the host, all of which they breach using a remarkable cell traversal technique until they reach their target hepatocyte (69). Using their circum-sporozoite (CSP) protein they then bind and invade the

hepatocyte(70).This begins what is known as the pre/exo –erythrocytic stage. Inside the liver they differentiate into either tissue shizonts or a dormant hypnozoite (67). The liver stage is then subject to asexual replications, which are responsible for the production of 10,000 to 30,000 merozoites that grow into the shizont (66). The stage culminates by the release of these merozoites into the blood stream where the erythrocytic stage ensues. The duration of the pre-erythrocytic cycle is between 7 to 10 days (67). For hypnozoites, they also follow similar cycle causing relapse weeks, months or even years after the initial infection (71). While relapses may appear random, it has been suggested they are evolutionary mechanisms used to bypass unfavorable climate for the vector (72). Evidently relapse periodicities have a frequency pattern of 3-6 weeks in tropical and up to a year for temperate zones, in congruence with the prevalence of the vectors for that season (20,49,52).

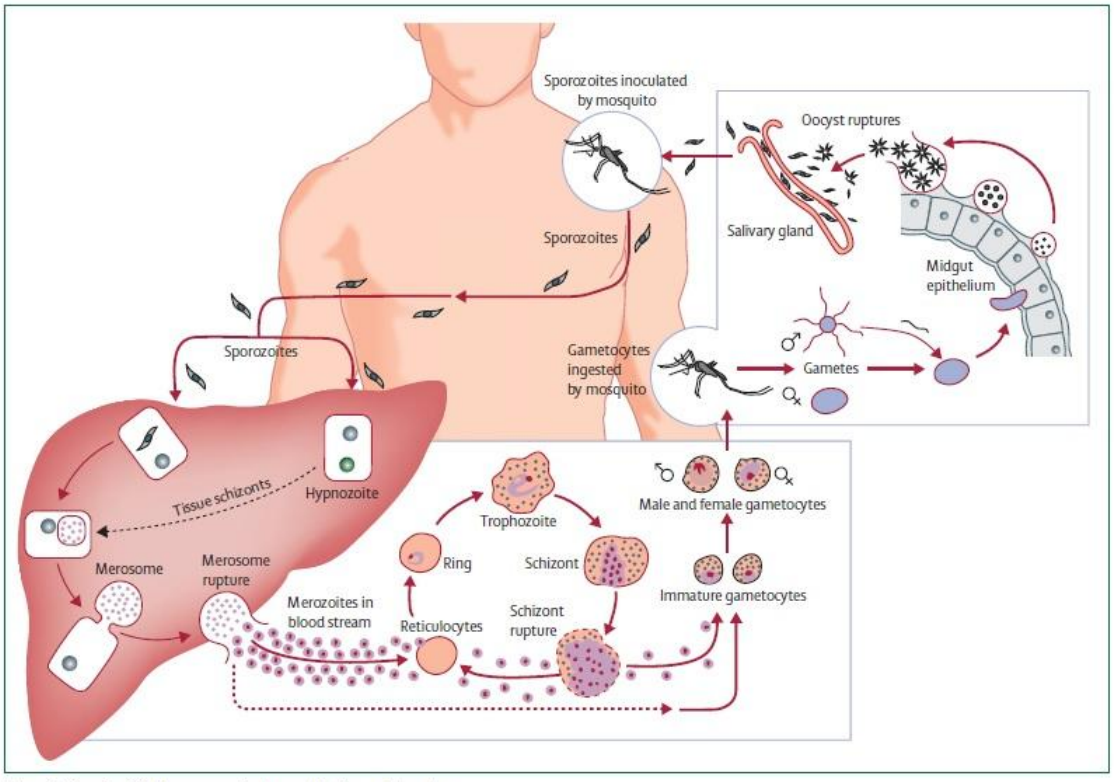


Figure: Life cycle of the human malaria parasite *Plasmodium vivax*

Figure 1 The life cycle of *Plasmodium vivax*(53).

The erythrocytic stage is initiated by the merozoites that invade red blood cells. This stage is also responsible for much of the clinical symptoms manifested in *P. vivax*

infections. For *P. vivax* it predominantly prefers reticulocytes, although it still is not understood why (23). So far a single blood surface receptor, the Duffy antigen Receptor for chemokines is implicated as a requirement for the infections of these reticulocytes in complement with its Duffy binding protein (73). The overall merozoite invasion takes 30 to 60 seconds and the process of invasion is divided into four phases (74). Identification of the host cell, reorientation of the merozoite so as to align its apical projection to the reticulocyte membrane, tight junction formation and finally entry (74,75). Inside it undergoes changes, maturing through ring, trophozite and shizont stages (76). Each cycle is culminated when the infected reticulocytes rupture to release more merozoites. This particular step is responsible for the anemic symptom (58). After each cycle a small number of blood stage parasites take male and female gametocyte forms, sexual forms of the parasite (77). Mature gametocytes can then be ingested by an anopheles mosquito during its blood meal. Uniquely for *P. vivax* infections the appearance of mature gametocytes before onset of clinical symptoms has been noted (49). *P. vivax* infected reticulocytes have also been shown to enlarge their host cells, making it more deformable, further characteristics such as cytoadherance and rosseting have also been suggested, all of which are important strategies to avoid clearance by circulation (54,78) The overall cycle is expected to take 48 hours (53).

Inside the Anopheles mosquito, ingested gametocytes are activated into gametes which fuse to form the motile zygote known as the ookinate (79). The ookinate is responsible for crossing the gut epithelium where it differentiates into its replicative form, the oocyst (68). The oocyst contains thousands of sporozoites, which are released into the salivary gland of the mosquito, to continue the cycle (68).

2.2 *Plasmodium vivax* treatment and control

Diagnosis of infections in malaria and indeed *P. vivax* are the basis for downstream control strategies, while also being informative for prevention and survey studies (80). The ability to detect and discriminate between *Plasmodium* species are termed as their sensitivity and specificity respectively (81). As we shall see subsequently both of these qualities are responsible for much of the success and failure of *P. vivax* treatment and control. For convenience, diagnostics purposes can be classified into three categories,

those for diagnosing clinical infections, mass blood surveys for detecting asymptomatic infection/carriers and clinical trials for evaluating efficacy of anti-malaria's (80).Diagnosing clinical infections are usually carried out using microscopy, rapid diagnostic tests(RDT) and/or ''clinical'' presumptive diagnosis(82). The latter form of diagnosis is based on the clinical manifestation of malaria i.e. by observing symptoms such as fever headache, and chills alone(82). This form of diagnosis is the least expensive and is perhaps more practiced in areas that are malaria prone (83). Their use however is riddled with errors, particularly because they lack specificity and consequently have been attributed to the over diagnosis of malaria in African countries (as high as 84%) (82,83). As such the risk of overusing antimalarial drugs and subjecting them to unnecessary selective pressure, unnecessary cost and masking other infections is very high.

The second method is Giemsa microscopy, based on Gustav Geimsas' methodology of using Methylene Blue and Eosin to stain *Plasmodium* parasites. It is perhaps the most widely used diagnostic procedure (82,84). It is relatively inexpensive with the ability to discriminate between the different *Plasmodium* parasites as well as detecting 10-100 parasites/µl of blood depending on the skills of the individual using the microscopy (85). Evidently the reliance in human skills is perhaps its weakest link, human errors such as incorrect staining, smear preparations and problems with interpretations of results can negatively affect the outcome of the diagnosis (80,82,85). For instance, false positive results are a making of poor blood film preparation (82). Another challenge to microscopy is low parasitemia, which are particularly hard to detect using microscopy and as such properly labeled as submicroscopic infections (86). These submicroscopic infections are prevalent in areas of low transmission, where microscopy has been shown to miss an average of half of all infections compared with other sensitive diagnostic procedures such as PCR (87). Further complications such as the limitations in detection of mixed infections, particularly in areas co endemic with *P. falciparum* and *P. vivax* have been observed (88). *P. vivax* is characterized for its low density infections, that combined with high parasitemia infections of *P. falciparum* make it easier to miss in microscopic slides, and this feature has also been indicated as the reason for underestimation of *P. vivax* (84,88,89). Worryingly the limitations of microscopy have implications towards drug administrations; misdiagnosis of species, false negativity and

positivity, which all play a role in the rise of drug resistance. Moreover, unreliable datasets can derail overall proper care for the patient and mislead future intervention strategies (5).

As a result, the limitations of microscopy required further diagnostic methodologies. Rapid diagnostic test (RDT); an immune-chromatographic assay which detects malaria antigen in the blood was introduced (82). It was an upgrade to microscopy because of its ease to use and speed, but its sensitivity remained an issue (82,90,91). Constant improvements have been made to improve its detection resolution in *P. falciparum*; However, same can't be said for *P. vivax* RDT'S, which might be in part because they are commonly occurring as low density infections (80,92). Even in the case of *P. falciparum*, RDTs have been responsible for erroneous results. RDTs targeting the histidine rich protein antigen of *P. falciparum* (HRP2) have been shown to give false negative results as results of deletion or mutation in the target gene. In contrast, false positive results have also been documented in patients that had cleared the parasite, mainly because the antigen remains for weeks after successful treatment (90). Moreover, false positive results have also been reported as a result of cross reactivity with rheumatoid factors (93).

Mass blood surveys, particularly with the intent of detecting low density (submicroscopic) and/or asymptomatic infections need a quite sensitive diagnostic mechanism. None are more sensitive and specific than genotyping of *Plasmodium* species by using their nucleic acids; which is achieved by a PCR amplification procedure(94). The sensitivity and specificity of this procedure is ideal for all types of diagnostic procedures but the relative high cost and need for static space have put a challenge to its use in point of care settings (95). Indeed, progresses have been made to make it portable via methods like loop amplified isothermal amplification(LAMP) , the cost however is still an issue (96,97). Clinical trials also use this procedure to evaluate efficacy of drugs, particularly using molecular markers to genotype resistances associated with their presence and prevalence (98,99). Further studies are underway to use serological biomarkers, like ones that can assess and differentiate clinical manifestations like asymptomatic infections, uncomplicated malaria and severe malaria (100).

Following diagnostics, particularly in clinical patients, a proper disease management is required. Overall two methods are used to prevent/treat and control malaria, chemoprevention, and vector control (2). Antimalarial drugs represent the chemoprevention management protocol. Several drugs with differing mode of action are available to treat malaria. As a result different countries implement different types of drugs according to their national malaria guideline. The most common drug however is chloroquine(CQ), synthesized in 1934, it is a class 4-aminoquinlines which affects haemozin synthesis (detoxification of waste from hemoglobin) (80,101,102). It has been effective in malaria endemic countries. Its low cost, longer half-life and its lower toxicity to humans are some of its favorable characteristics (102). However, the rise in drug resistance has forced several countries to review and replace this drug. Particularly in *P. falciparum* the effect has been most pronounced, with studies showing *P. falciparum* was resistant to CQ for 30 years before similar trends appeared for *P. vivax* (103,104). In *P. falciparum*, two genes *Plasmodium falciparum* chloroquine resistance transporter (PFcrt) and *Plasmodium falciparum* multidrug resistance transporter-1(PFmdr-1) are responsible for conferring resistance, similarly studies have tried to associate orthologues *P. vivax* genes, crt and mdr-1 with very limited success(99,105–107). However, chloroquine resistance in *P .vivax* was evident in in-vivo efficacy tests in several endemic countries (54,116). Alternatives for CQ include but are not limited to artesunates, mefloquinine, atovaquone, proguanil, halofantrine and piperaquine, accordingly each are used alone or in combination in different parts of the world to treat uncomplicated malaria(88). In the case of severe malaria, characterized by severe anemia and respiratory distress, WHO recommends similar treatment as *P. falciparum* (2). Accordingly arthemether (3.2mg/Kg IM), artesunate (2.4mg/Kg IV/IM) and quinine (20mg/Kg IV) can alternatively be used (88). Hypnozoites which are responsible for causing *P. vivax* relapses pose further problems, currently the only drug capable of clearing them is Primaquine(PQ) (110). PQ prevents relapse when given as a 14 day course, and in combination with other drugs can be effective for treatment as well as a transmission blocking entity (111). Parallel to its use, however, it has limitations, debate in its dosage regimen, contraindication on pregnant or lactating mothers and life threatening hemolysis in G6PD deficient (G6PDd)individuals are some of its drawbacks (21,80,112). The latter,

G6PDd is a common enzymopathy caused by a loss of an enzyme (of the same name) that is responsible for the catalysis of the first reaction in the pentose phosphate pathway (21). Similar to the 'null' duffy antigen, G6PDd is also associated with malaria resistance, and so it is prevalent in areas endemic to *P. vivax* (21,46). Hence, the people most in need of the drug are the ones most susceptible to its adverse effects. Worryingly recent studies have also shown G6PDd individuals afflicted by *P. vivax*, perhaps indicating a shift from its association with malaria resistance (46,86). PQ has also been subject to resistance, with a study reporting relapses after a treatment with the drug, although that could be because of low dosage (113).

The second method of malaria control is vector control. Vector controls represent activities with regards to preventing the vector anopheles mosquito from biting humans (2). Two insecticide based methods are commonly used, indoor residual spraying (IRS) and insecticide treated nets (ITN). IRS specifically has been responsible for the control and/or elimination of malaria in most parts of Asia, Europe, and South America (114). In retrospect ITNs are the major contributors (68%) for the decline of the clinical incidence of malaria by 40% in Africa between the years 2000-2015(19). As such, the two of them are responsible for much of the success against malaria. They are, however, like most control strategies susceptible to a level that can lead them to incapacitation. Both of the methods are indoor based and dependent on the vectors behavior, for instance early biting or outdoor biting by the vector can bypass these preventive measures easily (115). Insecticide resistance is also a pressing issue, insecticide Permethrin and Pyrethroids , most commonly used in IRS and ITN respectively have both reported resistance (2,114). Overall, 60 of the 73 malaria endemic countries have reported resistance in at least one insecticides (2). The combination of all of these factors has increasingly become a serious problem, with WHO warning a compromise in vector control could lead to a resurge in vectors and ultimately resurgence of malaria.

Vaccines or at least the potential of vaccines represent the final and most important tools against malaria. Currently the only vaccine to have completed phase 3 trials is RTSS,S/ASO1, based on the CSP protein of the *P. falciparum* (116). The vaccine is shown to reduce clinical incidence by 39% and severe malaria by 31.5% (2). The

feasibility of malaria vaccines have been further indicated by individuals living in endemic areas who gain immunity after successive exposure to the parasite, as well as passive immunity which has also been shown to be successful, i.e., the transfer of antibodies from an immune individual to a naïve individual (118). Accordingly several projects are under way to develop efficacious vaccines (119). However, the research focus is mainly on *P. falciparum* with the larger number of promising candidates identified as compared to *P. vivax*(24).

2.3 *Plasmodium vivax* in Ethiopia

Ethiopia is a country co afflicted by *P. falciparum* and *P. vivax*, with previous reports indicating a dominance of *P. falciparum* (120).The country has made strides in controlling *P. falciparum*, with as much as 66% decline in malaria cases (45).The progress for *P. vivax*, however, is almost antiparallel, and by the year 2015, Ethiopia was the second highest contributor of *P. vivax* worldwide after India (Figure 2) (2).This is in agreement to the increase or predominance of *P. vivax*, in countries co-infected by and on the verge of eliminating *P. falciparum* (89,121). This is perhaps the clearest sign that the same treatment and control regimens that worked for *P. falciparum* may not be as effective. Nevertheless the progress achieved in controlling *P. falciparum* has motivated the country to target several low transmission areas for elimination of malaria by 2020 (122). Several factors are amiss from this plan, such as the prevalence and overall relevance of submicroscopic infections, and what to do with hypnozoites. It has been established previously that *P. vivax* occurs as low density infections, with the ability to relapse, as demonstrated in studies done in Ethiopia (44,45,86,123). A study using a mass treatment and testing with the intent of clearing parasites from a low transmission setting indicated several shortcomings, such as low sensitivity of diagnostic methods (RDT) that impacted reaching submicroscopic carriers(124). Another potential problem is the lack of diagnostics for hypnozoites. While primaquine is quite effective for hypnozoites, its administration is not only unethical, but also fatal without prior testing for G6PD deficiency. Different qualitative and quantitative kits are available for this; although they are still fraught with errors(80). For instance because the gene is X linked; in females it can be absent, homozygous, or heterozygous. In the case of the latter, the interpretation of

the data would be complicated, as half of the red blood cells could be susceptible to hemolysis. Furthermore the cost is a pertinent issue (5). As a consequence in Ethiopia particularly along with Libya and Somalia ,primaquine is ‘’not recommended under any circumstance ’’by the WHO (80). For now, under the national malaria guideline chloroquine is the first line drug for treating uncomplicated *P. vivax* malaria, while severe cases are treated with artemisinin lumifantrine. This is despite numerous reports of chloroquine resistance from different parts of the country, although other parts of the country still sustain high efficacy (109,125–128).

In Ethiopia, 68% of the population lives in areas at risk of malaria (120). The country’s topographic arrangement is quite heterogeneous in altitude, leading to varying spatio-temporal meteorological conditions in different parts of the country (129). The transmission season and intensity also vary accordingly. Previous studies have made broad distinction to associate malaria transmission using altitude ranges and labeling them as lowland and the highland regions (130,131). Whereby the lowland experiences stable malaria transmission and acts as a reservoir for the mostly malaria free highland areas (13). Indeed malaria has been negatively associated with increasing altitude, which is in sharp contrast to the frequent severe epidemics that occurred in 1958 up to recent ones in 2002-2004 in highland areas (130,131). This is not surprising because malaria transmission is influenced by a multitude of factors (132). In Ethiopia malaria has also been positively associated with increase in temperature, and rainfall (peak monthly rainfall), as a result most parts of the country experience a short transmission season followed by long interval of low or no transmission (10,13,130,133).

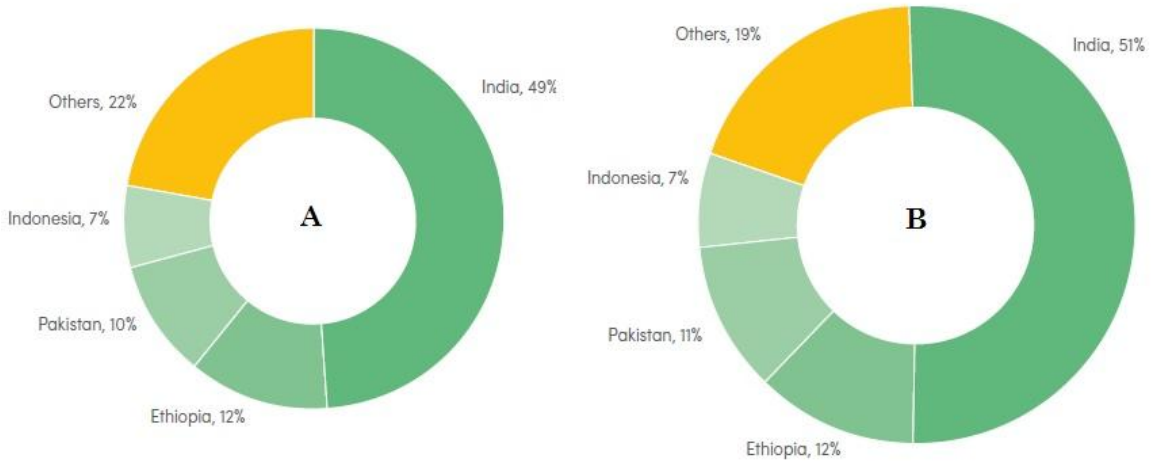


Figure 2 Estimated country share of A) *Plasmodium vivax* malaria cases B) *Plasmodium vivax* malaria deaths in 2015(2)

The dependence of climactic factors is directly related to the breeding behavior of the malaria vectors. Worldwide 40 significant species of the anopheles mosquito have been identified, of which 34 species have previously been found in Ethiopia (136). Currently 4 species are prevalent, the primary vector *Anopheles arabiaensis*, secondary vectors *Anopheles funestus*, *Anopheles nili* and *Anopheles pharoensis*(135). Accordingly a study has shown the strong association of the prevalence of any of the four mosquitos to the main rainy season (Sep/Oct) (136).However, temperature is also a significant factor, previously abnormal rise in temperatures have been associated with malaria epidemics in highlands, in parallel recent studies have shown that climate warming is significantly increasing malaria burdens in areas that were not previously malaria prone (129,131,137). An Increase in temperature is associated with spatial distribution of malaria cases, and in Ethiopia an increase of 35%-64% malaria cases per 1°C have been observed (131). The risk of malaria is also associated with other factors such as mobility, and control measures (138). Indeed the much successful indoor vector controls implemented in Ethiopia which have previously brought about a reduction in malaria infections, have

recently been shown to be susceptible to *Anopheles arabiaensis* and *pharoensis* species which can feed outdoors or during early part of the evening (115).In the past vector controls such as ITNs have provided remarkable results worldwide, and in the wake of the 2002 severe epidemic, Ethiopia distributed 46 Million Long lasting Insecticidal nets(LLINs) and sprayed IRS to 70% of the targeted households (139). As a result a rise in ownership and utilization of LLINs has been observed in several parts of Ethiopia(138,140). Henceforth, recent reports of susceptibility are quite alarming, and should be a cause for concern for future malaria intervention strategies.

2.4 Molecular Evolution of the *Plasmodium vivax* parasite

P. vivax is the most geographically widespread of the *Plasmodium* species(5). The origin however remains a topic of debate. Previous studies suggested that *P.vivax* originated in southeastern parts of Asia mainly because of *Plasmodium cynomolgi*, the closest known relative of the parasite, which in combination with other *Plasmodium* species in its clade infect south east Asian primates(141). This origin theory however doesn't explain the prevalence of malaria resistant variants like the near fixation of Duffy negativity in Africa. Recently the prevailing consensus from several studies is that *P.vivax* originated from Africa and that it ''escaped out of Africa'' before the spread of Duffy negativity(141,142). Indeed it should be noted that these conclusions are reliant on comparative population genetics studies, an absence of a continuous ex vivo culture has inhibited further insight. Concurrently genomic studies have been instrumental in providing information on *P. vivax*. The advancement in this areas and the recent sequencing of the vivax genome has brought much needed insight into the biology, diversity, epidemiology and pathogenesis.

The sequencing of the first reference genome 'Salvador I' in 2008, enabled scientists to study the genome and the genetic structure of *P. vivax* (143,144). This study showed that the vivax nuclear genome contained 5433 genes in 14 chromosomes, with an estimated size of 26.8 megabases (Mb) and a G+C content of ≈42% (144). The study demonstrated distinctive differences between *P. vivax* and *P. falciparum*, such as G+C content (*P. falciparum* has 19.4% coverage) and differences in microsatellite numbers with fewer ones found in *P. vivax*. Indeed a recent study has used some of these microsatellites for

the development of SNP barcodes to genotype *P. vivax* infections globally (43). Another interesting question before the complete sequencing of the whole vivax genome was the prevailing comparative diversity of the duo. In two very recent studies , *P. vivax* demonstrated it was more diverse where even ‘’the least variable *P. vivax* population was more diverse than a sample of diverse *P. falciparum* isolates’’(26) Both studies indicated that *P. vivax* was evolving locally, responding to local endemic settings; concurrently hypervariable regions were detected at ends of the chromosomes, where most genes involved with antigenicity are located (145).

2.4.1 Genetic Diversity

The basic evolutionary processes of mutation, recombination, natural selection, genetic drift, and gene flow, interact to affect natural populations(146). Mutation and recombination are primary and secondary sources of genetic variation respectively; consequently the effect of natural selection, genetic drift and gene flow dictate the frequencies of these alleles. Selection can be directional, i.e. positive and negative, which cause fixation or loss of alleles in a population, or they could be balancing whereby different alleles including rare alleles are maintained in a population(147). Selective neutrality in contrast dictates that genetic variation is a result of random mutations which have no effect for the organism(148). Therefore, this serves as a null hypothesis for what are known as tests of neutrality, whereby patterns of diversity are compared and lay the basis for measuring selection pressures. Accordingly, several methods have been developed such as Tajima’s D, Lu and Fu’s F and Wrights F_{ST} that collectively calculate departure from neutrality(149).

Measures of genetic diversity are good indicators for the effect of natural selection. For instance positive selection is observed in drug resistant genes, exemplified by reduction of diversity around resistance determinant loci, in response to drug pressure (150). Genetic diversity can therefore be used as an indicator of a pathogens response to interventions (151). Genetic diversity in *P. vivax* is extensive, although varying in scale in its gene families .As an example , a genome wide study on *P. vivax* isolates from South east Asia, revealed that five gene families had the highest diversity namely MSP3,VIR,MSP7 SERA and RBP (152). Most of these are genes coding for membrane

associated proteins, antigenic loci's of *P. vivax*. Quite similar to *P. falciparum* these antigen loci are highly polymorphic, and studies on them have so far shown that these are the parasites adaptation to evade the immune response through a frequency dependent selection (balancing selection) (153).Apart from assessing the effects of natural selection, this antigen loci are also used to study the parasite's diversity and population structure. Particularly in investigating allelic diversity and frequency distribution, multiplicity of infections and differentiating relapse from new infections (154). Most use PCR-RFLP genotyping methods for *P. vivax,* and/or sequencing these antigen polymorphic genes. Use of PCR-RFLP is the earliest methodology; it uses variations in recognition sites for restriction enzymes as the basis to differentiate between isolates, accordingly restriction fragments of different lengths can be separated in an agarose gel and visualized by several methods. Similar to the genome wide studies, PCR-RFLP results showed that antigen loci like MSP3α and MSPβ are highly polymorphic gene (155). As a result, they have been widely used worldwide to give information on molecular variation of *P. vivax* isolates. Although without sequencing data, it is quite impractical to associate molecular evolutionary reasons behind observed variations or indeed use the data to infer relationship between the parasites response to its environment.

The extent and distribution of genetic diversity in *P. falciparum* has been strongly associated with its transmission, where high transmission settings experience high diversity of isolates (103). The same type of association couldn't be made for *P. vivax* where even in low transmission, diversity can be quite high (156,157,158). This could be a result of relapse, early gametocyte production or underestimation of *P. vivax* in sub microscopic infections (143). Regarding the population structure, a study by Hupalo and colleagues grouped global *P. vivax* populations into three, the Americas, Asia and a quite distinct population in Papua New Guinea (26). The African population represented by isolates from Mauritania and Madagascar were grouped with the Indian cluster, supporting the theory that African and south Asian *P. vivax* populations are similar. However this evidence is applied to only two African countries and can't be used to cluster all African *P. vivax* population to Asian. In fact a study using SNPs to detect population divergence revealed the highest pairwise divergence between Ethiopian and Sri Lankan Isolates (43). Furthermore both countries representing Africa in that study

setting have a negligible prevalence of *P. vivax* (0%). Contrastingly the burden of *P. vivax* is high in East African countries mainly, Ethiopia ,Sudan and Eritrea with a prevalence of 10%,36% and 39% respectively for the year 2014(7). The extent of *P. vivax* prevalence in Eritrea was further revealed by a study that investigated migrants arriving in Germany, where it was reported that 95% of overall *P. vivax* detected in the refugee camp belonged to this Eritrean migrants. Ethiopia and Sudan have also reported *P. vivax* as a public health concern, further aggravated by a recent report which detected 13.67% and 18% severe *P. vivax* cases in Ethiopia and Sudan respectively (6,159). Interestingly, *P. vivax* diversity in Sudan was low according to a study using PvMSP3α and PvCSP markers, additionally no multiple clone infections were found indicative of a low transmission setting (6). In contrast, a study in Ethiopia found a high level of Polyclonality in Hawassa, Interestingly, however, it failed to associate it with transmission intensity, although that could be because of higher mobility of people from neighboring towns (47). Polyclonal *P. vivax* infections are quite common in malaria endemic areas, where they have been associated with disease severity, drug resistance and transmission intensity with varying success (160). The former two are attributed to within host competition of clones, and along with this, drug resistant *P. vivax* has been associated with severe and fatal malaria (50). In Ethiopia while no formal inquiry has been made regarding the association of polyclonal infections with drug resistance and severity, CQ resistance has been reported in various parts of the country together with the report on severity along the Sudanese border (109,125,128).

2.4.2 Effect on prevention and control of *Plasmodium vivax*

The extent and distribution of *P. vivax* diversity is of great consequence to the prevention and treatment programs. Molecular approaches to study the local and regional transmission pattern, migration and evolution are detrimental to our understanding of how effective these control activities are. Because of the long history of interaction with the *Plasmodium* and vice versa, the genomes of both have been greatly influenced, each to protect itself from the other. This ‘’evolutionary arms race’’, is a result of the parasites adaptation to evade the immune system as well as various prevention and control

measures; conversely several resistant variants of malaria are maintained by the human genome in endemic areas (147,150). This dynamic is mainly maintained by selection, particularly in antigens of the parasite, drug resistant genes and resistant variants in humans (21,150,161). As such, they are a good place to start to monitor for changes.

The greatest threat with regards to prevention and control of malaria remains drug resistance. This is perhaps best exemplified by the fact that almost all of the antimalarial drugs used so far have developed resistance. Disturbingly that also includes the current first line drug artemisinin and its derivatives(162). One of the ways to curb resistance is to monitor and surveil drug resistance genes that confer selective advantage for parasites. By doing so areas with significant number of drug resistant genes for a specific antimalarial, will instead be treated by a new antimalarial with a new mode of action. This enables to get on top of the situation as opposed to waiting for overwhelming evidence of failure of drugs, which would in fact cause unnecessary suffering for patients as well as further spread of resistance among many other things. Including artemisinin most antimalarial with drug resistance have molecular markers to track and monitor their resistance, like Pfcrt for chloroquine or *Plasmodium falciparum* kelch propeller domain (k13 propeller) gene for artemisinin (98,99,163). There are, however, significant exceptions for drugs used to treat *P. vivax*, despite numerous reports of treatment failure in chloroquine, no single molecular marker has been established to monitor its resistance(106,107) An overall subsequent in reduction of diversity is observed around these genes, a result of ‘’hitchhiking’’ by unrelated markers close to this loci(164). This ‘’sweeping effect’’ has larger implication on the overall genetic diversity, as evidence the impact of control measures on *P. falciparum* are major factors for its decrement of diversity. Conjointly the extensive diversity of *P. vivax* might indicate that past control efforts have less impact on them.

2.5 Malaria vaccines; The last frontier

In light of the WHO goal to reduce malaria case incidence and malaria mortality to 90% by 2030, it is now clear that *P. falciparum* and *P. vivax* are two different beasts. The approaches that gained substantial success in decreasing *P.falciparum* don't seem to affect *P.vivax*. In addition to the threats imposed by *P.falciparum* like multi-drug resistance, *P.vivax* are 'helped' by several factors such as lower density infection, and our diagnostic inability to detect or control hypnozoite stage parasites. Apart from mitigating resistance in all fronts, including vector control and antimalarial drugs, it is important to add further tools to fight the different features of *P.vivax* which weren't an issue with *P.falciparum*. Vaccines provide this much needed platform, supplying services such as preventing infection, preventing blood stage disease or blocking onward transmission from humans to mosquito.

2.5.1 The need for *P. vivax* vaccines

The need for a vaccine is based on the limitations of the current diagnostic prevention and control methods. Their limitation is further exploited by *P. vivax*, which enjoys a widespread geographic presence as compared to its relative (49). The ability to produce hypnozoites which later cause relapses complemented by lack of diagnostics to detect this stage(72); Large number of sub-patent or submicroscopic infections which are below or near detections are some of the limits of the currently used tools (165). Moreover, most *P. vivax* vectors bite outdoors, rendering indoor control methods ineffective (135,166). Ideally vaccines aim to fill this vacuum, for instance, pre erythrocytic vaccines that target sporozoites, prevent infection as well as potential hypnozoite reservoirs. Similarly, blood stage vaccines prevent diseases, but also reduces transmission potential of the infected individual (22). The impending threat of drug resistance is also another reason to develop vaccines, drug pressure is always going to be an issue especially in areas that have targeted elimination, as this favors drug resistant strains. Furthermore, studies on the fitness of wild type versus resistant type strains have indicated that in the 'therapy free' environment, where there is no uniform drug pressure, the wild type is advantageous prevailing at larger numbers (161,167,168). Ideally a vaccine can complement drugs by limiting their exposure in areas that have developed multidrug resistance, until resistance

subsides. Alternatively, in tandem with PQ, they can also aid in elimination settings, where PQ is contraindicated, they can prevail, providing protection to G6PDd individuals, pregnant or lactating mothers. In line with this, in 2006 WHO launched a malaria vaccine technology roadmap, later revised by 2013 to include *P. vivax* vaccines; it aims to develop vaccines with high efficacy against clinical malaria and vaccines that can achieve malaria elimination in all transmission settings by the year 2030.

2.5.2 Approaches and challenges of developing *P. vivax* vaccines

Despite a catalog of efforts throughout the years only one vaccine has progressed to phase 3 trials, RTSS/AS01(169).This vaccine is unsurprisingly designed for *P. falciparum*, in going with this trend, most of the identified vaccine candidates and most clinical trials are also for *P. falciparum*. *P. vivax* has in comparison fewer candidates and even less clinical trials, much of this can be attributed to the lack of knowledge towards its biology (170). Despite this, studies using related species like *P. knowlesi* and P. *cynomolgi* have progressed research in *P. vivax* vaccines (25). The approach to *P. vivax* vaccine research is similar to *P. falciparum* vaccines. However vaccines appear more feasible for *P. vivax*, as immunity appears more rapidly for *P. vivax* than *P. falciparum* in both high and low transmission settings. Furthermore, studies have shown that long lasting immune response can be induced for *P. vivax* (171). Like *P. falciparum, P. vivax* vaccines have been classified into three, based on the lifecycle they target.

Pre erythrocytic vaccines; primarily target the infective sporozoite to prevent entry into the liver but are also used to target liver stage hepatocytes. The sporozoites are an appealing target for a number of reasons, chief among them is that they represent a bottleneck for the parasite, only a few sporozoites are injected by the vector into the blood stream so blocking this stage of the cycle would completely prevent the infection(172). Indeed an escape of even a single sporozoite can potentially result in thousands of merozoites into the blood stream. The challenge of developing a vaccine for this stage is also a result of low number of sporozoites in addition with the small amount of time it spends in the blood stream(1-2 minutes), which might not be enough to elicit immunogenic

reaction(173). Although this can also be turned to an advantage, because of the lack of reaction from the immune system, the parasite is not expected to have immune evasive strategies(170). So far two candidates have been extensively studied, Thrombospondin-related anonymous protein (TRAP) and CSP

Blood stage vaccines; are designed to prevent disease. The blood stage is where the pathophysiological manifestation of *P. vivax* is displayed. Therefore, most of the vaccine candidates of *P. vivax* have been designed for this stage. Vaccines at this stage aim to prevent invasion of reticulocytes to reduce parasite density in the blood stream including gametocytes (174). Therefore, blood stage vaccines can prevent both disease and onward transmission (175). So far the most extensively studied vaccines are based on the duffy binding protein(DBP) and merozoite surface protein 1(MSP-1), but large number of candidates have also recently entered development phase, such as rhoptry binding protein(RBP)), merozoite surface protein 3α(MSP3α) and merozoite surface protein 9(MSP9) (176,177).

Transmission blocking; The last of these three are not actually involved in preventing infection nor disease but are specifically tasked in obstructing transmission. Vaccines target antigens expressed in the mosquito phase of the lifecycle, such as oocysts or ookinete(178). Most notably antigens Pv25s and Pvs28s expressed on surface of the zygote and ookinete respectively have shown good promise in preclinical trials(179).

The challenges for developing vivax vaccines are numerous, most notably however; the lack of continuous culture which inhibit ex vivo studies of vaccine candidates, limited investment and the high cost of developing vaccines; Antigenic diversity and the strain specific immune response need to be addressed (29,180,181). Recently progress and new approaches in developing short term continuous cultures have been reported, the cost of developing vaccines however is still reliant on philanthropic efforts(23,182). The latter of the problems however is perhaps the biggest of them all. Antigenic diversity maintained by balancing selection allows parasite to evade immune mechanisms which can only

detect few strains at a time and as a result develop a strain specific response(24,183). Similarly vaccines developed without putting the extent of diversity in perspective will only be efficacious against a subset of population while selecting for vaccine resistant strains(184). This was witnessed in a clinical trial targeting the combination B vaccine; using the polymorphic antigen MSP2.The antigen has two allelic families 3D7 and FC27,by using only the 3D7 form it was found that it could only elicit antibodies specific to 3D7 while remaining neutral for FC27(185). Similarly, a study was conducted to evaluate PfMSP3 based vaccine, using its two allelic forms independently to elicit antibodies, it found strain specific antibodies, furthermore frequency based statistical analysis revealed this allelic forms were maintained by a balancing selection(186). It is already established that *Plasmodium* species antigens are polymorphic with extensive diversity, similarly vaccine candidate antigens of *P. vivax* have shown large amount of variability(144). However, it is important to note that the mere presence of genetic diversity in vaccine candidate antigens does not warrant antigenic escape, rather it is of significance to study polymorphisms relevant to immune selection (187). So far however only a single study targeting the *P .falciparum* apical membrane antigen, has been able to analyze worldwide allelic diversity, and rigorously identify which polymorphism were responsible for mediating antigenic escape (188). Studies like that however first require information on the circulating allelic diversity of the antigen worldwide, as such vaccine development programs include surveys as part of their studies.

2.6 Merozoite Surface Protein 3α

Merozoites are unicellular forms of *Plasmodium* species, expressed during the late shizogony stage in the liver and red blood cells, which are responsible for erythrocyte invasion (74). Previous studies have suggested that in invasion, proteins located in the surface of the merozoite are involved, shedding as soon as irreversible attachment is established between merozoite and erythrocyte (177). This has clearly peaked interest in studying merozoite surface proteins, and so far 10 distinct MSPs have been identified (26).MSP3 is one distinct multigene family, located on the Chromosome 10 of *P. vivax*, on a 60kb region with 11 members of its own (189). Previously *P. vivax* merozoite surface protein 3 was considered a homolog of the *P. falciparum*s' own MSP3, and hence

its nomenclature (190). Despite certain similarities in structure, and both being immunogenic However, they are quite distinct from each other, further illustrated by weak similarity of their sequences(144). PvMSP3 has a central alanine rich region, with heptad repeats predicted to form coiled coil structures. Unlike the other MSPs, MSP3 is not attached to the merozoite via a Glycophosphatidylinositol (GPI) anchor, but rather through this coiled coil structures which form a peripheral association with other membrane bound proteins(190,191). From PvMSP3s 11 members, two have been widely studied and used as a population genetic marker, these are PvMSP3α and PvMSP3β.

PvMSP3α is a highly polymorphic loci, a result of insertions, deletions, recombination and point mutations in its central alanine rich domain (36). Despite its overall high diversity, higher degrees of polymorphisms are limited to specific domains (Figure 3)(32). In his study in 2002, Rayner and his team conducted a molecular evolutionary analysis by dissecting MSP3α'S nucleotide sequence into four regions, amino terminal region (Positions 229-309*), Block I(310-1188),Block II(1300-2058) and Carboxy terminal region(2059-2151).They found extensive diversity in the amino terminal and block I regions while block II and C terminal were relatively conserved. Subsequent studies found a similar trend where the block I region was found to have large number of in-del events, whereas the block II was conserved and found to be under positive selection; non-synonymous polymorphisms observed in this block were attributed to recombination events (36,151). In spite of the high polymorphism exhibited in PvMPSα, it was suggested to be a potential vaccine candidate because it's a component of the merozoite and studies on its orthologue PfMSP3 indicated immunogenicity and protectiveness (186). In tandem, it has also been established as a reliable cost effective population genetic marker in countries such as Thailand, Bangladesh, Papua New Guinea and several other *P. vivax* endemic countries.

*Positions are based on Belem reference sequence, genbank accession number AF093584

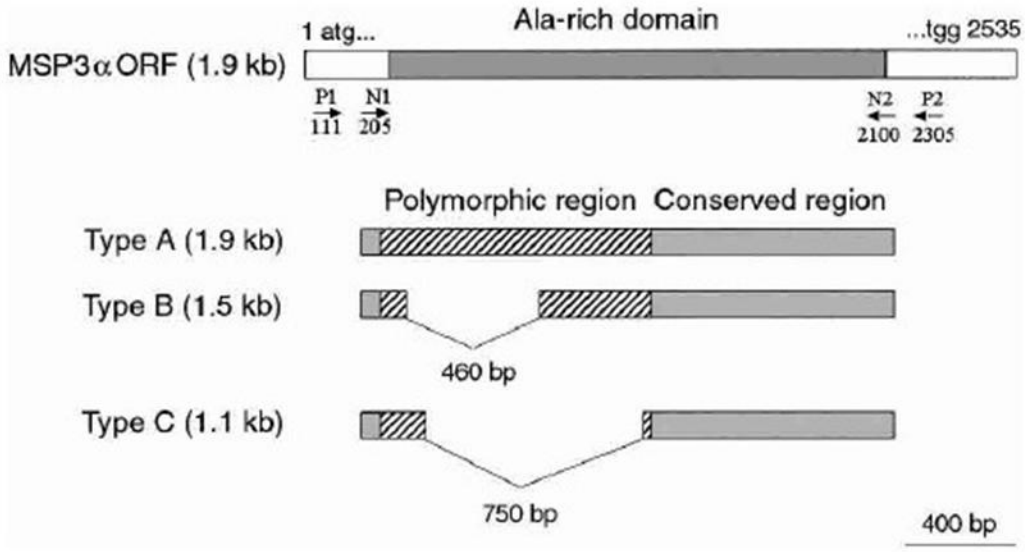


Figure 3 Plasmodium vivax Merozoite Surface Protein 3α and its three size variants(192)**. Positions are based on Belem reference sequence, genbank accession number AF093584.**

2.6.1 A Molecular Marker for Population Genetic analysis

Genes encoding polymorphic antigens have been widely used to type *P.vivax* populations, such as apical membrane antigen(AMA1) ,circumsporozite antigen(CSP)and merozoite surface protein 1(MSP1)(193–195).Like most of this markers,PvMSP3α has also been widely used in tandem with other antigen loci's or alone to detect allelic variations and multiple infections(33,196). Its application is fundamentally based on its size variations of PCR products and consequent application of restriction enzymes to digest these products .PvMSP3α has three size variants, Type A which is 1.8-1.9Kilobase (kb) in size, Type B with 1.4-1.5kb and Type C with 1.1-1.2kb ,a fourth variant D has also been reported with a size of 0.5kb although its prevalence is negligible(Figure 3) (197).From worldwide studies type A is the most prevalent, while type B and C have variable prevalence in different areas, and type D has only been reported in India. The reason behind these size variances is down to a deletion in the block I component, where

type B has approximately 460 bp deletions whereas Type c has 750 bp deletions(133). Apart from analyzing size variants, PCR products of PvMSP3α can also help detect multi-clonal infections, where the presence of more than one band of the PCR product is indicative of a multiple infection. To further discern diversity, restriction enzymes *Hha I* and *Alu I* are applied to produce patterns that are then used to calculate the amount of alleles(33). The basis of this RFLP procedure is the result of insertion deletion events, and point mutations which shift recognition sites of these enzymes, and as a result forming unique set of patterns for the different alleles(36,155). The RFLP procedure also allows to further detect multiple clone infections, this is accomplished by adding the size of RFLP fragments and comparing with uncut PCR product, where the total size of fragment exceeding uncut PCR product is indicative of a multiple infection. It is also important to note that the enzyme *Hha I* 's recognition sites are usually found in Block I, whereas *Alu I* has sites in both blocks and as such has more restriction fragments i.e. number of alleles for *Alu I* are almost always higher than for *Hha I*(32). Although PvMSP3α is a widely used genetic marker, only a handful of studies have analyzed the molecular evolutionary mechanisms behind the variation patterns, or indeed to quantify the genetic diversity observed between alleles. Indeed a study executed in 2005 using 11 endemic vivax populations has estimated the nucleotide diversity(π) in the complete sequence to be 0.05 for Asia, 0.056 for Europe and 0.049 for the global sequence set(32,198).Nevertheless it is important to note that, the study was undertaken a decade ago and doesn't reflect the current *P.vivax* population dynamics ,furthermore the study doesn't reflect the population genomics of African isolates.

2.6.2 PvMSP3α as a potential *Plasmodium vivax* vaccine candidate

Ever since its discovery, PvMSP3 has been implicated as a potential vaccine candidate due to its similarities with its ortholog PfMSP3. Although subsequent studies have revealed its distinctiveness from PfMSP3 or indeed MSPs of *P. cynomolgi* and *P. knowlesi*, PvMSP3 is still strongly associated with immunogenicity and protectiveness(199). Accordingly, the two key elements required in potential vaccine candidates, immunogenicity (the ability to elicit pronounced number of antibodies) and association with protectiveness have been assessed.

Four independent studies were executed in Brazil and Papua New Guinea, two countries with a predominant prevalence of *P. vivax.* The studies tested the full length as well the four regions of PvMSP3α for their antigenic potential. Overall the PvMSP3α gene proved to be highly immunogenic (78%) in all the studies comparable to other vaccine candidates such as AMA1 and CSP, with the full length of the antigen being the most immunogenic (77-78%). For the C terminus, two of the studies in Brazil reported a 54%, while a third reported 68% immunogenicity, while the study in PNG found it at 65%. For the N terminus, it recorded the lowest immunogenicity with a range of 38-39% responsiveness. The two blocks were highly immunogenic with block I having 64% and block II reporting a 53% Immunogenicity(31,40,41,199). Interestingly, another study characterized the elicited antibodies of block II and discovered them to be cytophilic IGg1 and IGg3 antibodies which have previously been associated with protectiveness from clinical episodes malaria (200). Furthermore, the study also found 15 antigenic determinants/B cell epitopes in the two central blocks. Previously the block II component has been found to be highly conserved (π=0.019), and its low diversity indicates it is unlikely to be involved in vaccine escape(39). Curiously however a small region in this block has been found to be highly variable among different isolates, these variants have generated two motifs, motif I: MSELEK/**L**S**K**LE**E** and motif II: TAANVVKD/**KE**A**TAA**K**L** (39)**.** In the previous study these motifs have been found to be one of the 15 B cell epitopes, hence their variability has been suggested as a mechanism to escape immune system. Evidently the collusion of this studies indicate that block II is the more ideal candidate, as the extensive polymorphism exhibited in the gene is limited to the N terminus and block I, which although immunogenic are under balancing selection and hence more likely to be involved in immune evasion and potentially vaccine escape. Block II in contrast is highly conserved but still highly immunogenic, Furthermore, the pattern of variation in the B cell epitopes in this block have been found to be equally prevalent in worldwide populations (32,35,36). All of this facts point to PvMSP3α as a potential vaccine candidate.

3. Objectives of the Study

3.1 General Objective

- To investigate genetic polymorphism and signatures of selection in the merozoite surface protein 3^{α} gene, in Ethiopian *Plasmodium vivax* population, a potential malaria vaccine candidate antigen.

3.2 Specific Objectives

- To assess genetic variation of *Plasmodium vivax* isolates at MSP3α gene locus by using PCR-RFLP and examine phylogenetic relationship within the Ethiopian RFLP haplotypes, using full sequence data (block I and II) of the gene for representative isolates.

- To analyze the extent and distribution of diversity in the block II component of the gene by dissecting sequence data

- To investigate departure from neutrality and signs of balancing selection, in the block II component (to determine if the gene region is influenced by immune selection) and conjointly investigate the presence and extent of recombination in this block

- To establish the phylogenetic relationship of the PvMSP3α block II within the Ethiopian population and among selected global population sequences and analyze haplotypes derived from Ethiopian and global sequences by performing network analysis.

4. Materials and Methods

4.1 Study Area and Population

The study was conducted on stored samples collected in 2012-13. Sample size determination took substantial amount of dataset(n=50) as recommended by two studies (201,202).Sample collection and study was conducted in Shewa Robit town in the Amhara region and Shala district located in the west Arsi zone of the Oromia region(Figure 4). Malaria transmission is unstable and seasonal in the two districts following the two rainy seasons. The town of Shewa Robit is Located at 9° 59' 40.6''N and 39° 53' 48.9''E with an elevation of 1280m above sea level. The city experiences a warm climate with heavy rainfall in its peak rainy season. In 2013 Shewa Robit was estimated to have a population of 44,726 individuals with 5.03% of its population reported to be afflicted with malaria for that calendar year(203).

In the West Arsi zone samples were collected from the Shala District at four health centers (Aje, Bure, Haposto, Ilala health centers) and Melka Oda hospital. Aje is located at 7° 17' 34.2''N and 38°21' 46.3''E, with an elevation 1852meters above sea level. Melka Oda is located at 7°13' 7.2167°N 38°29' 38.4833°E with an elevation of1937 meters above sea level; Bure is located at 7°15' 7.25°N and 38° 29' 38.4833°E with 1696meters above sea level; Haposto is located at 6.7377904 and 38.3691932; and Ilala is located at 8°55' 28.27''N,39° 50'35.90''E. Shala District was estimated to have a population of 149,804 in the year 2013. In all study sites malaria is caused by both *P. falciparum* and *P. vivax.*

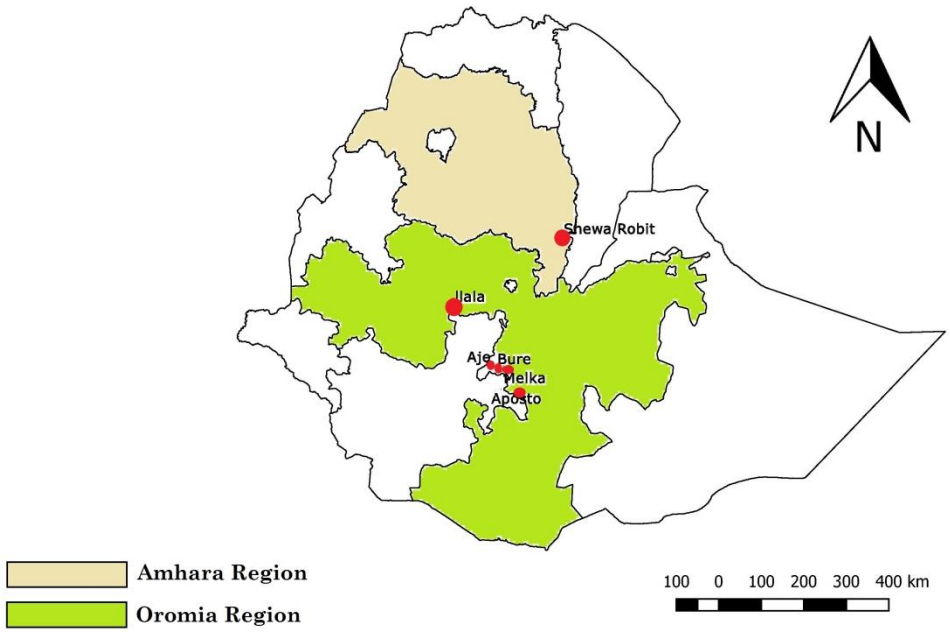


Figure 4: Map of the study sites, red dots indicate locations of specific sites

4.2 Sample collection

Stored dried blood spot (DBS) samples (n=146), ~0.3ml finger prick blood samples on Whatman 3MM (Whatman, maidstone, UK), at -20 ^{o}C were retrieved from AHRI biobank. The DBS samples were collected from self-presenting individuals for malaria diagnosis and treatment after the objectives, risks and benefits of the study were explained and written informed consent was obtained from participants or guardians/parents of minors. The blood spots on the filter papers(Whatmann 3MM) were air dried and transported at room temperature and stored in a freezer in zippped plastic bags, containing self-indicating silica gel desiccant beads(Geelay chemical Ltd). Finger pricked blood samples were also used for malaria diagnosis using RDT (SD BIOLINE Malaria Ag P.f/P.v POCT test kits ,Standard Diagnostic, Inc, Germany) that was performed in the field settings and blood films were then sent to the Adama (formerly Nazareth) malaria control center for microscopic examination. Two experienced laboratory technicians who were blind to the RDT as well as each other's result analyzed the blood films using thick and thin blood smears.

4.3 Laboratory Procedures

The 146 DBS samples that were RDT and microscopically *P. vivax* positive were used for the downstream process. From this 146 DBS samples, 50 DBS samples that represented all six study sites were randomly selected, next parasite carriage was reconfirmed for this select samples using nested PCR amplification targeting the small subunit 18S RNA gene(94). PCR positive single species *P. vivax* samples were then genotyped for MSP3α. PCR products with single bands were indicative of monoclonal infections; products with more than one band were a result of mixed infection. Accordingly only monoclonal samples were then further analyzed using the RFLP procedure; this enabled to generate patterns pertinent to genetic diversity, samples were then PCR amplified for sanger sequencing targeting the block I and block II components of the MSP3αgene.

4.3.1 DNA Extraction

Genomic DNA was extracted by a modified Chelex Saponin extraction method (134,204). Briefly, 6mm DBS was punched and immersed in 820µl of a 0.5% saponin (Sigma Aldrich) in-1x Phosphate buffer saline (PBS) solution and left on horizontal shaker overnight. The next morning it was centrifuged for 1 minute at high speed (12,000 rpm). After aspirating the supernatant (using automated vacuum suction system), 1ml of cold (4°C) 1xPBS was added to each tube and incubated at 4°C for 30 minutes. Next the tubes were shaken for 30 minutes and the PBS was discarded from the tubes. Chelex (Chelex®100 resin, Bio-RAD Laboratories) 150µl of 6% was added to elute the DNA. The gDNA was eluted by incubating the tubes for 7 minutes, four times in water bath at 95°C. Finally the tubes were centrifuged at a maximum speed for 5 minutes (12,000 rpm) and 80µl of the supernatant was carefully transferred to sterile (DNase/RNase free) tubes, and stored at -20°C until further use.

4.3.2 *Plasmodium* species confirmation by nested PCR

Genus and species level identification of *P.vivax and P.falciparum* was carried out by a nested PCR procedure using primers and PCR conditions as described before(205,206) (Table 1).The reaction conditions were 68.75 µl of each 0.25µM primers,0.25 µl of 0.25mM dNTPs, 2µl of 2mM of $MgCl_2$ with 0.2µl of 1 unit of Taq polymerase(GOTaq DNA Polymerase ,Promega) in a 25µl total reaction volume, the primary reaction volume contained 5µl (5ng/µl) of the gDNA eluate and the species specific PCR used 2µl of the genus specific PCR product as template, finally for amplification the T100 Thermal cycler (BIO-RAD) was used. Positive controls, *P. falciparum* NF54 cultures (Radboudmc, Nijmegen, The Netherlands) and *P. vivax* malaria reference laboratory Positive control (London School of Hygiene and Tropical Medicine, London, UK), were included on every PCR plate. For negative control, DNA/RNA free water was included. To minimize risk of contamination, the following three procedures were undertaken in separate rooms designated specifically for each action: - DNA template preparation (extraction), master-mix preparation and addition of DNA template for amplification.

To visualize products of amplification, 2% Agarose (HI Res standard, AGCT Bio-products Ltd) gel was prepared in 1x Tris Borate EDTA buffer (TBE); the solution was

then mixed and boiled in a microwave; next 0.3µg/ml of ethidium bromide was added to the agarose but only after it had gotten lukewarm. After the gel was casted and cooled for 30 minutes, 3µl of 100bp DNA ladder (Promega) was added. Next amplified products alongside positive and negative controls were added to their independent wells. Voltage was set at 120V and results were subsequently visualized by UV illumination, followed by photographing in a Gel-doc system (BIO-RAD).

Table 1 PCR master mix for primary (N1) and nested (N2) amplification respectively

Number of reactions					50				
	Stock concentration		Final concentration in reaction		For 1 reaction (µL)		MM set-up		
PCR Buffer	5	X	1	X	5	µL	275	µL	
Mg Conc of Buffer	0	mM	2	mM Mg					
Separate Mg Solution	25	mM			2	µL	110	µL	
dNTPs	25	mM	0.25	mM each	0.25	µL	13.75	µL	
Forward Primer	5	µM	0.25	µM	1.25	µL	68.75	µL	
Reverse Primer	5	µM	0.25	µM	1.25	µL	68.75	µL	
Taq Polymerase	5	U/µL	1	U/reaction	0.2	µL	11	µL	
PCR grade water					10.05	µL	552.75	µL	per reaction
MM final volume					20		1100	µL	20
Template	1	ng/µL	5	ng/µL	5	µL	275	µL	
Total volume					25				
	N1: Aliquot 20µL + 5µL input								

Number of reactions					50				
	Stock concentration		Final concentration in reaction		For 1 reaction (µL)		MM set-up		
PCR Buffer	5	X	1	X	5	µL	275	µL	
Mg Conc of Buffer	0	mM	2	mM Mg					
Separate Mg Solution	25	mM			2	µL	110	µL	
dNTPs	25	mM	0.25	mM each	0.25	µL	13.75	µL	
Forward Primer	5	µM	0.25	µM	1.25	µL	68.75	µL	
Reverse Primer	5	µM	0.25	µM	1.25	µL	68.75	µL	
Taq Polymerase	5	U/µL	1	U/reaction	0.2	µL	11	µL	
PCR grade water					13.05	µL	717.75	µL	per reaction
MM final volume					23		1265	µL	23
Template	1	ng/µL	5	ng/µL	2	µL	275	µL	
Total volume					25				
	N2: Aliquot 23µL + 2µL N1 product								

4.3.3 Genotyping of *P. vivax* by PCR-RFLP by targeting PVMSP3 α

To determine details in genetic diversity and multiplicity of infection, samples were checked with PCR followed by RFLP procedure as described below

4.3.3.1 PCR amplification of PvMSP3 α

The alanine rich region of the block I and block II of the PVMSP3α gene was amplified using primers and PCR conditions as described by Bruce *et al.*(1999)(33)and briefly outlined in table 2. Nested PCR reaction was carried out in final reaction volume of 20µl using 2µl of DNA extract for Primary round (N1) and 1µl of N1 products for the nested round (N2). One unit of Taq polymerase was used per reaction with oligonucleotide primers at final concentration of 0.1µM, dNTPs at 0.15mm and 2.5mM $MgCl_2$. PCR products were visualized on 0.8% agarose gel containing 0.25µg/ml ethidium bromide after a running for 90minutes at 80V. Products were identified using 1kb plus molecular ladders (Invitrogen). PCR product sizes were then used to differentiate samples as well as to determine the multiplicity of infections.

Table 2: Primer sequences and thermo-cycling conditions for PCR RFLP of the MSP3α gene

PCR	PRIMERS(5' →3')	CYCLING CONDITIONS Temperature Time
Primary round	**P1:**CAGCAGACACCATTTAAGG **P2:**CCGTTTGTTGATTAGTTGC	Initial denaturation 94°C 3 min Denaturation 94 °C 30 sec Annealing 56°C 30 sec **35 cycle** Extension 68°C 2.5min
Nested round	**N1:**GACCAGTGTGATACCATTAACC **N2:**ATACTGGTTCTTCGTCTTCAGG	Initial denaturation 94°C 3min Denaturation 94°C 30 sec Annealing 57°C 30 sec **30 cycle** Extension 68°C 2.5m

4.3.3.2 RFLP analysis of PvMSP3 α

Restriction fragment length polymorphism was done using two restriction enzymes *Hha* I and *Alu I* (Table 3) using Multi Core Buffer (Promega) and *Alu I* using SuRECUT/ Buffer A (Sigma Aldrich) .

Each nested PCR product (N2) 4µl(5µg/µl) was digested using 0.5µl (5unit) of the restriction enzyme and 10x reaction buffer at a final volume of 20 µl and incubated at 37°C for 5 hours on a heat block. The digested products were then subjected to gel electrophoresis using 1.8% agarose containing 0.25 µg/ml of ethidium bromide at 120v for 1 hour. 1kb plus molecular ladder (Invitrogen) was used for estimating sizes of products.

Table 3: Restriction Enzyme recognition sites

Restriction Enzyme	Cleavage site	Source Microorganism	Pattern of cutting
Hha I	5'…GCGC. 3' 3'…CGCG..5'	*Haemophilus haemolytics*	3' Overhang —
Alu I	5'…AGCT..3' 3'...AGCT..5'	*Arthobacter luteus*	Blunt ends —

4.3.3.3 Sanger sequencing of the PvMSP3 α Gene

Nested PCR products were first purified using the QIAquick PCR purification kit(QIAGEN) and then used as a template to sequence the MSP3α gene using primers N1 and N2 as well as internal primers F2 and R2 as described(Appendices 10.6) (207). Consequently samples targeting the full sequence (block I and block II(1896 bp)) and further samples targeting the partial sequence(block II(758 bp) were sequenced twice, in forward and reverse direction, using the Big Dye terminator sequencing kit(Applied

Biosystems) and the ABI Prism 310 Genetic analyzer(Applied Biosystems) (BASE CLEAR, The Netherlands).

4.3.3.4 PvMSP3α sequence and phylogenetic analysis

Raw nucleotide sequence data was first inspected visually to ensure correct base calls of the chromatogram data. The data was then trimmed for low quality regions and then assembled ,manually edited and individually aligned using ClustalW (208) to two reference sequences the Belem and Salvador strain Msp3α sequences (genbank accession numbers AF093854 and PVX_097720 respectively), for comparison and identification of polymorphisms. Ultimately 14 full consensus sequences with a single continuous read corresponding to the 205-2100 bp (of the Belem strain) were extracted for this study.

In a similar fashion sequences corresponding to the block II region only (1300bp-2017bp of the Belem strain) were assembled, manually edited and aligned to both references sequences for comparison. In order to analyze the dataset generated in this study to a global context, 126 MSP 3α sequences from 17 *P. vivax* endemic countries were retrieved from genbank (http://www.ncbi.nlm.gov/genbank/).The sequences included in the current study were isolates from Brazil, India, Thailand, Sri-lanka, North Korea, Papua New Guinea, Ecuador, Bangladesh, Venezuela, South Korea, Malaysia, Pakistan, Indonesia, Myanmar, Panama, Mauritania and Vietnam (Appendices 10.8). The 717bp region corresponding to the PvMSP 3α block II of the reference Belem strain, in all 167 sequences (including the 40 generated in this study) was used for the final analysis.

To infer parameters of sequence polymorphism, genetic diversity and phylogenetic analysis, the following procedures were performed. To analyze sequence polymorphisms behind molecular variation of RFLP alleles, in-silico restriction mapping was carried out using the full Sequences generated in this study. This also allowed for comparing gel images of PCR-RFLP haplotype diversity to the In-silico map. In-silico digestions below 100bp were not applied to compare the two as they are hard to discern in gel electrophoresis; however the fragments were documented.

To extrapolate data on sequence diversity , various statistics were computed including the number of polymorphic sites (S) , within population and overall nucleotide diversity (π), number of haplotypes(H) and haplotype diversity(Hd) (209).To determine genetic differentiation, Wrights Fixation index F_{ST} (210) was tested through 1000 random permutations; to assess departure from neutrality and examine if regions were under selection, the number of synonymous substitutions per synonymous site(d_S) and non-synonymous substitutions per non-synonymous site(d_N) was calculated using the modified Nei Gojobori method (211) , the null hypothesis of neutrality (H_0: $d_N = d_S$), and alternative hypothesis of positive selection($d_N > d_S$) and purifying selection($d_N < d_S$) were tested using a two tailed Z test for neutrality and one tailed Z test for either of the alternative hypothesis, standard errors were computed through 1000 bootstrap replicates. Furthermore, Tajima'sD(148) Fu and Li's F(212) were applied using a sliding window approach to investigate signatures of balancing selection. Recombination events(213) were also analyzed through seven methods namely, RDP, GENECONV, Bootscan, MaxChi, Chimiera, Sis Scan and 3Seq(214–219). To further study genetic diversity pertinent to antigenic diversity, haplotypes were constructed from all of the 167 MSp3α Block II sequences, using only non-synonymous polymorphisms. Consequently haplotype network was drawn using a median Joining algorithm(220).

To study the phylogenetic relationship among the RFLP haplotypes as well as among the 167 block II sequences, both aligned nucleotide and deduced amino acid sequences were used to construct phylogenetic trees. Trees were constructed using the maximum likelihood tree with Tamura and Nei model of nucleotide substitution for nucleotide alignments (221) and the Jones Taylor Thronton model of amino acid substitution method for amino acid alignments(222), with 1000 bootstrap replicate support for both. Accordingly, bootstrap support (>50%) is shown as percentage in all the phylogenetic trees presented in this study.

4.4 Ethical consideration

The umbrella project under which this project was executed was reviewed and approved by the institutional ethics review boards of Aklilu Lemma Institute of Pathobiology, Addis Ababa University (IRB/11/2011/2012), Armauer Hansen Research Institute

(PO22/12), the National Research Ethics (310/109/2016) and the London School of Hygiene and Tropical Medicine(10628). DNA samples were sent for sequencing according to the Material transfer agreement stated under 10628 to London School of Hygiene and Tropical medicine.

Under the procedure approved, prior to sampling, informed written consent was sought in written format from all participants, in the case of minors, legal guardians or parents provided the consent. Under the consent, participants' agreed for the storage and further study of their samples. Patient confidentiality was kept throughout the study with patient names substituted by individual codes. Patients positive for malaria by RDT were treated according to the Ethiopian Malaria Guideline(223).

4.5 Data Analysis

Socio-demographic data was entered and analyzed using Excel. Chromatogram data was analyzed using Chromas (version 2.6.4, Technelysium LTD.) Sequence trimming, manual editing and assembly were performed using SeqMan Pro 14 (Lasergene 14 software, DNASTAR Inc.) sequence alignment and construction of phylogenetic tree as well as tests of D_N/D_S was performed using the MEGA7 software (Version 7.0.21), further population genetic parameters such as genetic diversity, population differentiation (Wright's Fixation index), recombination and measures of departure from neutrality (Tajima's D and Fu & Li's F) were examined using the Dnasp Software(Version 5.10.01,Universitat de Barcelona), finally Dnasp was also used for haplotype construction. Snapgene software (Version 3.3.3, GSL Biotech LLc) was used to construct in-silico restriction map and consequent restriction site analysis. Genalex (Version 6.501, Australian National University) was used to convert data between programs. POPART (Version 1.7, Allan Wilson Centre) and NETWORK (Version 5.0.0.1, Fluxus Technology Ltd) were used to create and visualize haplotype network.RDP4 program was used to detect recombination signals. To maximize success Ugene (Version 1.26.1, Uni Pro) was also used as a control/alternative performing assembly, alignment and construction of phylogenetic trees.

5. Results

5.1 Proportion of malaria in the samples from the study sites

A total of 146 patient blood samples detected to be harboring the malaria parasite by both RDT and microscopy diagnostic methods were selected for this study. From the 146 samples, 26 were from Aje, 33 from Shewa Robit, 32 from Ilala 13 from Melka-oda , 24 from Haposto and 18 from Abure. Most of the samples, 63.3% (94) were positive for *P. vivax*, 28.6% (40) were of *P. falciparum* species and the remaining 8.4% (12) were mixed infections. *P. vivax* was detected at higher proportion in samples from three of the six study sites, namely Ilala, Abure and Haposto with 60%, 80% and 95% of the malaria infections respectively. In samples from Aje and Shewa Robit, *P. vivax* contributed to 50% of the overall infections; only in samples from Melka Oda did it contribute at a lower proportion to *P. falciparum* (40%).Mixed infections were only detected in samples from three of the study sites; Melka Oda , Ilala and Shewa Robit, with 10% ,15% and 20% proportion Respectively(Figure 5).

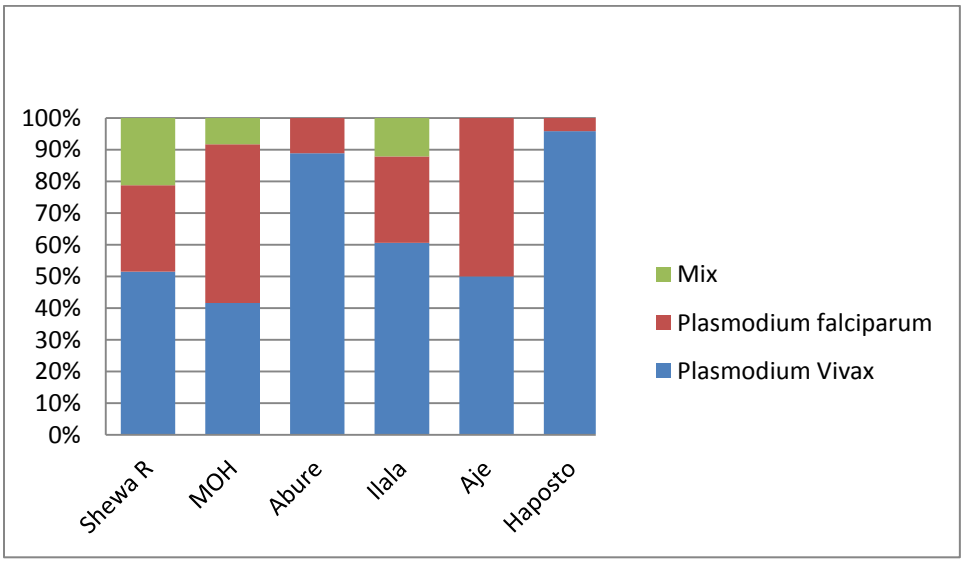


Figure 5: Retrospective proportion of *Plasmodium vivax* in the samples from the six study sites

Based on these 146 positive samples, group of individuals with ages above 15 were the most reported cases, which contributed to 64.38% of the observed infections. This

observation was particularly evident in samples from the study sites of Shewa Robit and Haposto where the majority of the infections fell into that group. Furthermore, males were more infected by both *P. vivax* and *P. falciparum* (64.38%) than females (36.98%), and more likely to be infected by *P. vivax* (45.3%) than their female counterparts (25.3%). It was also a similar case with *P. falciparum*, with males being infected at a higher proportion 18.4% than females 10.7% (Table 4).

***Table 4: Demographic information based on the positive samples from the six study sites (2012/2013)**

	Health Facilities						
Age Group	**Aje**	**Ilala**	**Melka Oda**	**Haposto**	**Bure**	**Shewa Robit**	Percent (Malaria infections)
<5	2	6	0	0	4	0	**8.21%** (12/146)
5-15	11	9	4	6	8	2	**27.39%** (40/146)
> 15	12	18	9	18	6	31	**64.38%** (94/146)
Gender							
Male	14	16	11	16	12	23	**63.01%** (92/146)
Female	13	11	3	8	8	11	**36.98%** (54/146)

*Table is solely based on the 146 DBS samples that were deemed positive by both RDT and Microscopy

5.2 PCR confirmation of *Plasmodium vivax*

A total of 50 samples collected from the six study sites were amplified, with 100% success rate (Figure 6). A cross examination was carried out using primers specific for *P. falciparum*, and no mixed infections were detected.

.

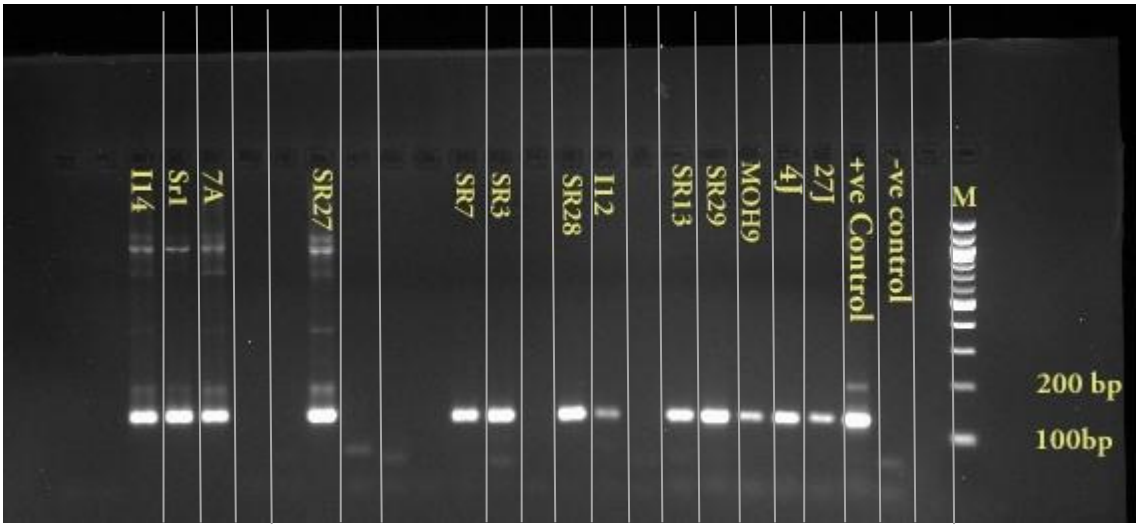


Figure 6: *Plasmodium vivax* genotyping using 18sRNA Gene; Nested PCR amplified products from the six study sites M=100bp molecular ladder

5.3 PCR-RFLP MSP3α

The 50 samples chosen to represent all six sites were analyzed for the MSP3α gene, among which 47 were successfully amplified. Three size variants were observed type A (1.9kb), type B (1.5kb) and type C (1.1kb) (Figure 7). The type A was the predominant variant with 39(82.97%), type B had 6(12.7%) and type C with 3(4.2%) frequencies. A single multi-clonal sample was also observed because of the presence of more than one band.

To get further diversity analysis, the nested products were subjected to independent restriction digestion analysis using *Hha I* and *Alu I* , excluding the multi-clonal sample. Consequently, further multi-clonal samples were detected, 4 from Ilala, and 1 from Haposto. All age groups were represented in the multi-clonality (Ages; 3, 15, 18, 25, 32)

except those above 40 years of age. Overall from the 50 samples analyzed, 12% of the infections were multi-clonal *P. vivax* infections as detected by PvMSP3α.

Using the *Hha I* enzyme, from the 47 samples analyzed, 11 different allelic patterns were observed, type A size variants showed many number of patterns than type B or Type C. Type A size variants had 8 different size combinations (A1-A8); Type B had 2(B1, B2) while Type C had only one size variant (C1) as observed using the *Hha I* digestion(Figure 8). *Alu I* on the other hand did not produce clear restriction pattern in most of the isolates (Figure 9).

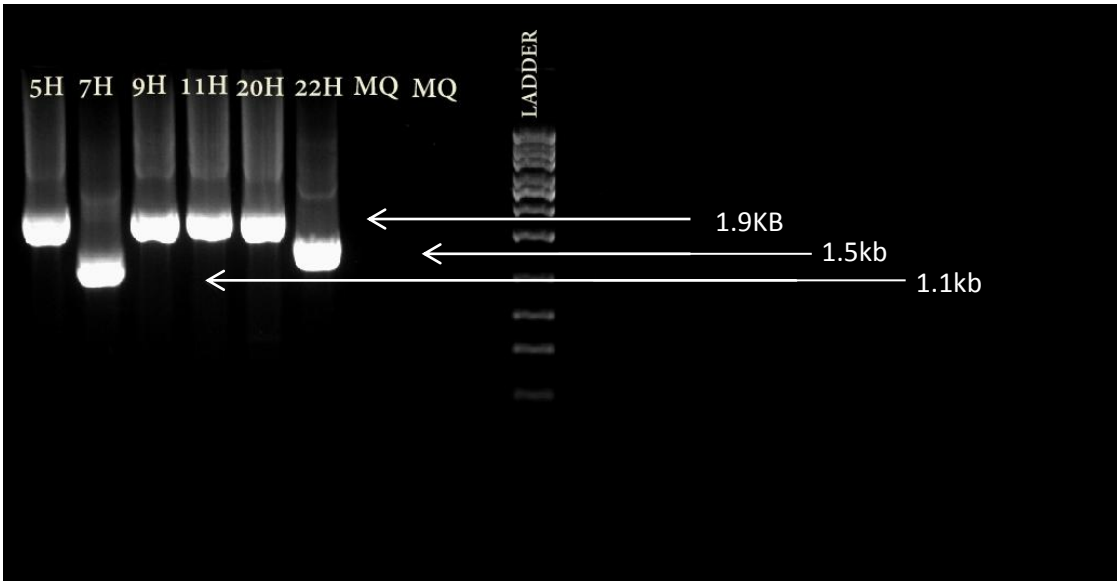


Figure 7: Nested amplified PCR products of the PvMSP3α gene, in the field samples; with 1kb molecular ladder.

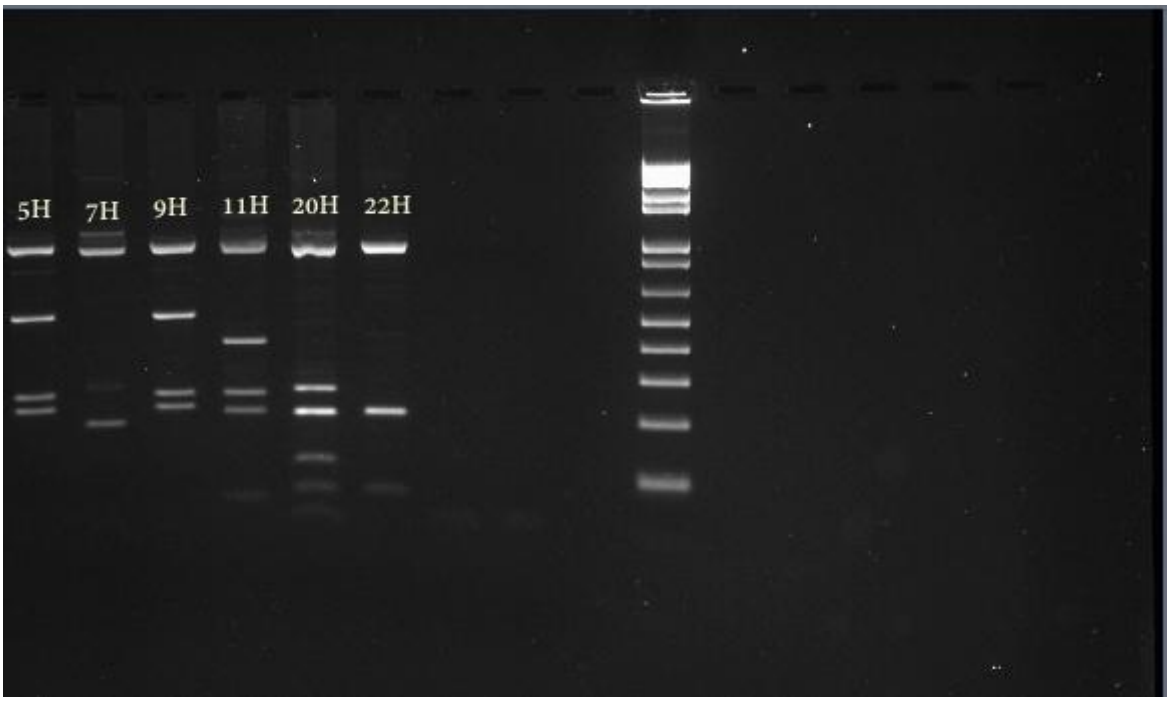


Figure 8: PCR-RFLP products of the *Hha I* digestion of the MSP3*α* gene, in the field samples; with 1kb plus molecular ladder.

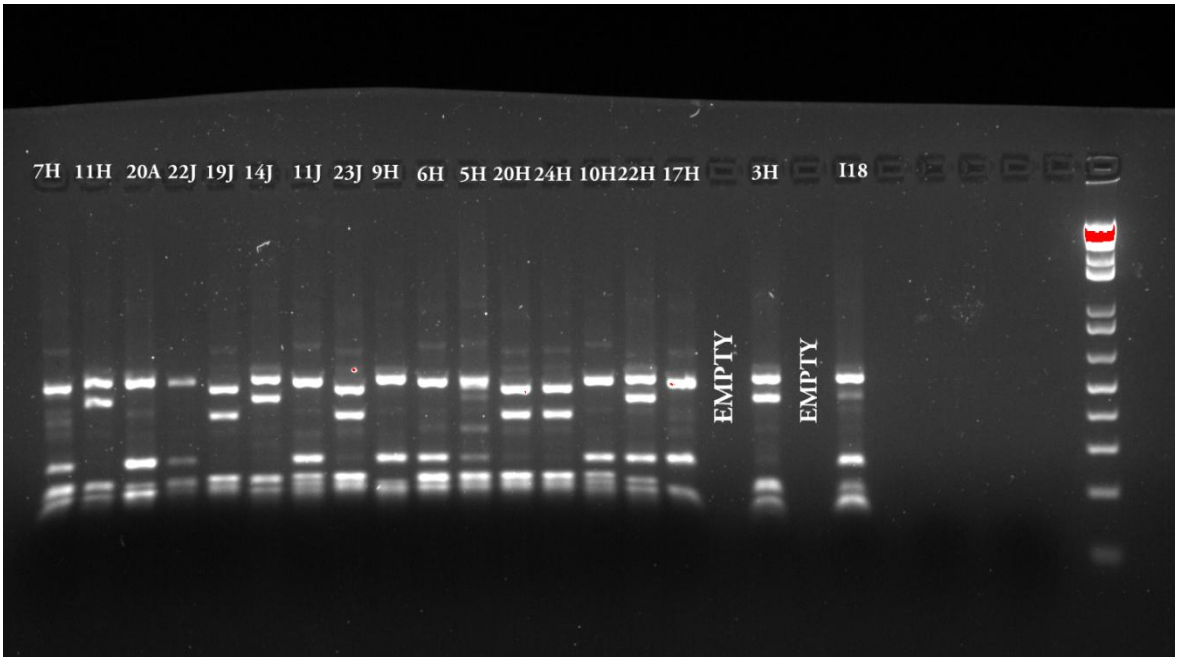


Figure 9: PCR-RFLP products of the *Alu* I digestion of the MSP3*α* gene, in the field samples; with 1Kb plus molecular ladder

5.4 Sequence polymorphism

In total 15 samples were sequenced for MSP3α entire region i.e. including both block I and block II. However one of the samples was not of sufficient quality to obtain the full length data. The remaining 14 samples were used for the rest of the downstream analysis regarding the complete region. Accordingly the three size variants observed during the PCR amplification of the MSP3α gene were all represented; type A size variants ranged from the smallest 1815bp to the largest 1925 bp . Type B ranged from 1407 to 1437, and Type C's two representatives had a size of 1173 bp. Type A variant sequences had several insertion deletions amongst that class itself, as evident in the length between the smallest and the largest sequences which amounted to 110bp difference amongst the two. Regarding differences between the three classes, type B variants had an intact 281 bp at the start of the sequence followed by a deletion of 435 bp and ending with an intact block II compartment. Type C variant had a similar trend with type B regarding the start of the deletion, but it has a deletion of 748 bp which is then also followed by an intact block II compartment.

Overall the block II component was relatively conserved unlike the block I component in the 14 full sequences. More samples were sequenced to investigate, the PvMSP3 α block II accordingly 40 sequences were generated. They had 39 single nucleotide polymorphisms, of which 29 were parsimony informative (Site positions: 6 19 122 136 154 160 175 269 275 284 292 319 322,400 403 411 412 414 420 423 426 431 432 433 439 442 510,520 and 706), and 23 were non synonymous mutations. MSP3α block II was peculiarly characterized by the presence of two motifs as observed in the amino acid alignment. Motif I had either MSELEK or LSKLEE alleles; Motif II had either TAANVVKD or KEATAAKL. These variations were observed in Ethiopian populations with motif I s MSELEK being highly prevalent (Figure 10). Motif II s KEATAAKL/TAANVVKD were found to be equally prevalent in the Ethiopian isolates (Figure 11). Apart from this almost all other non-synonymous polymorphisms were

shared between isolates, whereas singleton polymorphism were not observed in the isolates

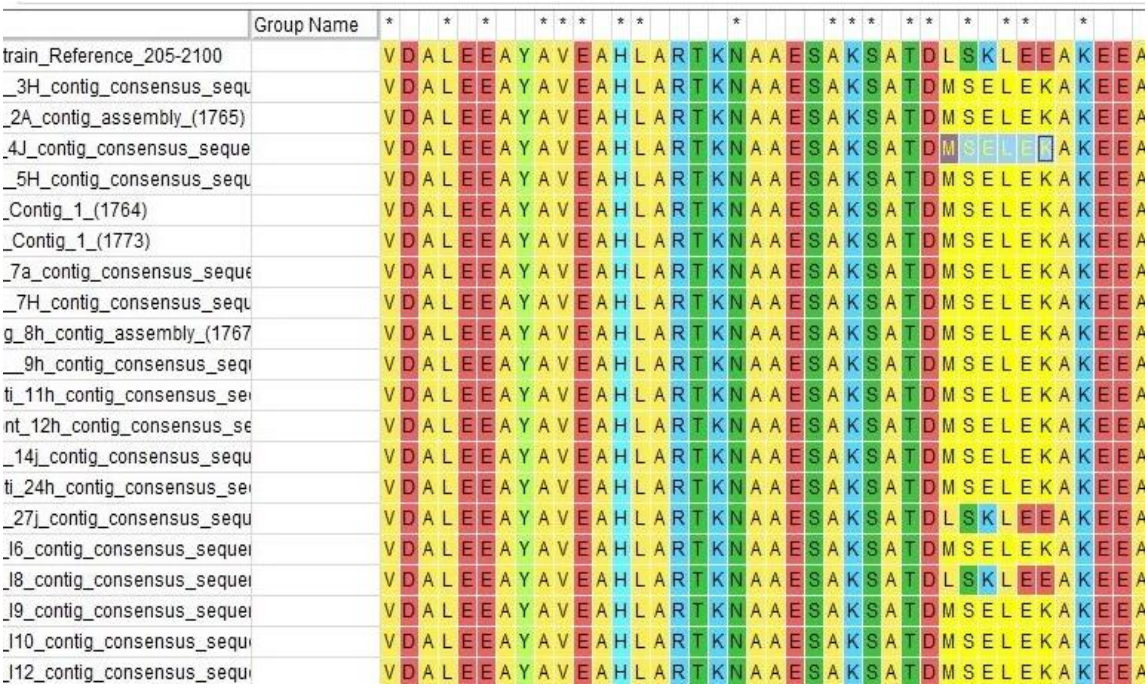


Figure 10 MSP3α block II motif variations in the field samples; frequency between LSKLEE AND MSELEK in Ethiopian Isolates

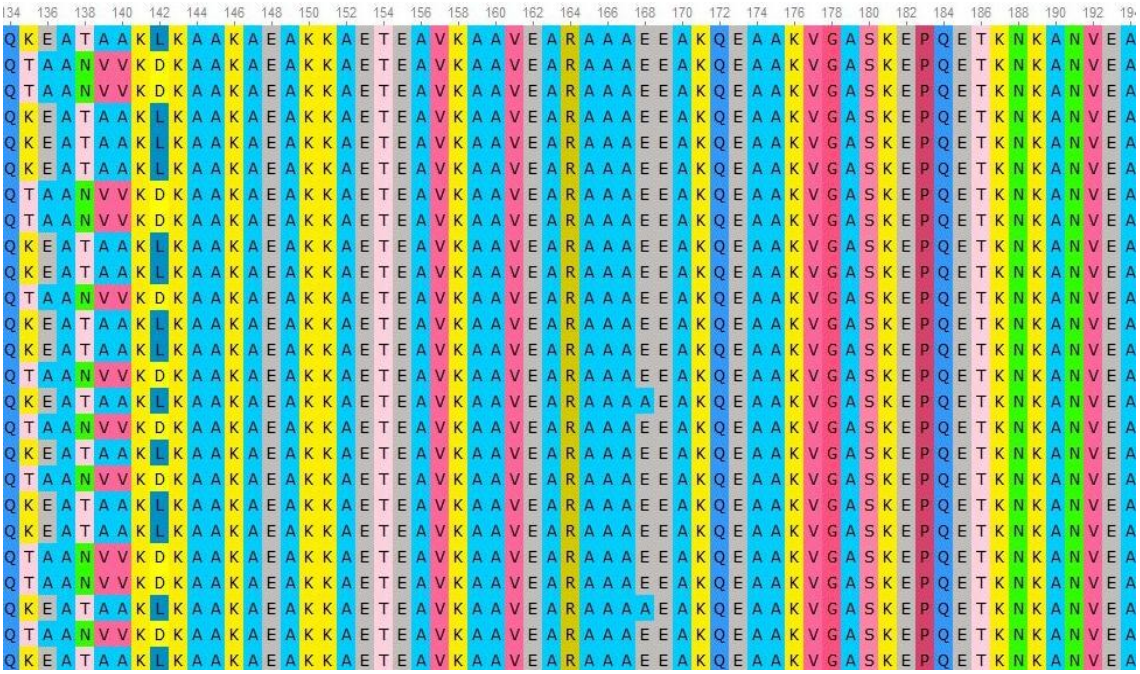


Figure 11 MSP3 α block II motif variations in the field samples; frequency between TAANVVKD/KEATAAKL in Ethiopian Isolates

5.5 In-silico RFLP and phylogenetic relationship of the block I

The shift caused by insertion deletions causes the restriction patterns that are observed as a result of the movement of the recognition sites. In comparison between the three size classes, type A had more allelic combinations as a result of superior number of recognition sites in its length, followed by type B and type C which only had one allelic pattern. Interestingly, it was found that for the 14 sequences analyzed, all of the recognition sites for *Hha I* were exclusively found in the block I component of the gene, whereas *Alu I*'s recognition sites could be observed in both components of the gene. Due to the difficulty in distinguishing *Alu I* restriction patterns in the RFLP gel images for the complete 50 samples, the analysis was solely based on the in-silico pattern. Accordingly the 11 allelic variants observed in *Hha I* digestion were similarly used, overall *Alu I* digestion produced 6 to 12 fragments, some of which were hard to discern using the RFLP gel procedure (Figure 9 &12).

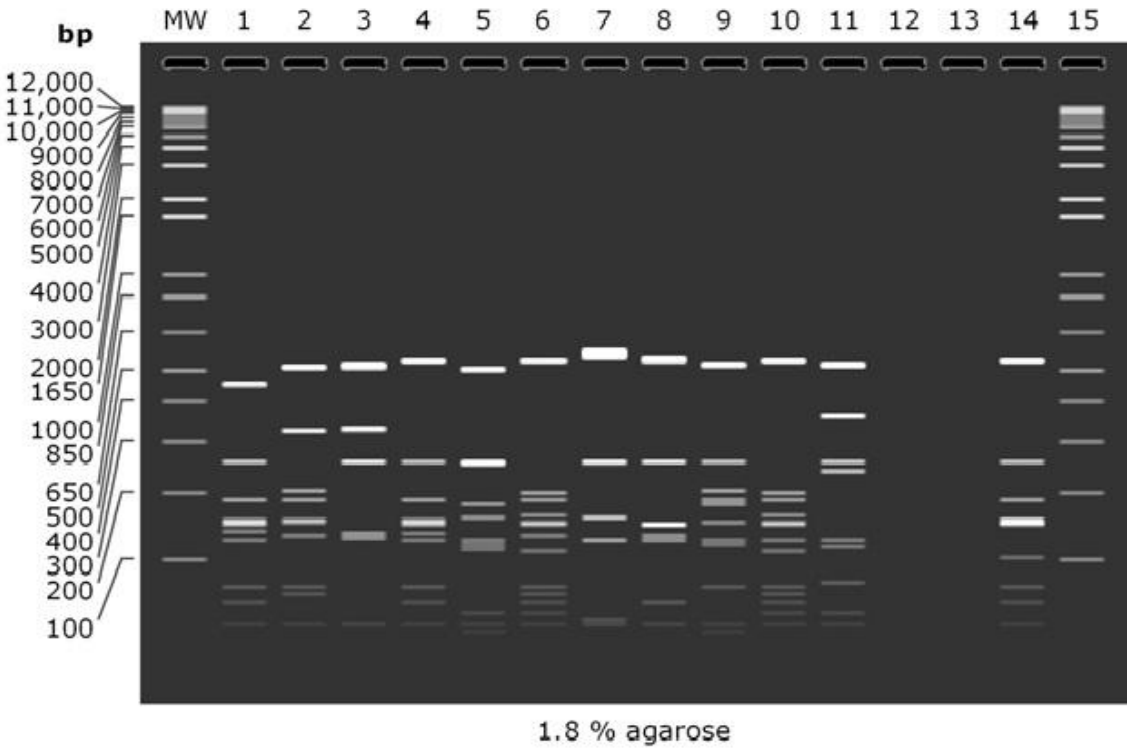


Figure 12 In-silico PCR using *Alu I* restriction fragment; The 11 allelic variants displayed here from lane 1 to lane 11 and the reference Belem Strain in Lane 14, Ladder= 1kb plus

Similarly, by using the In-silico RFLP, the restriction sites and resulting fragments were observed using the *Hha I* enzyme (Figure 13). As in the RFLP gel images all size classes and size variants shared similar 1kb fragment in the beginning of the patterns. However the analysis using In-silico digestion indicated slight polymorphism, as sizes ranging from 917 to 982 were seen. Overall the in-silico RFLP digestion and results from the actual RFLP digestion using the *Hha I* were similar. Although In-silico analysis allowed the visualization of bands that were hard to tell apart in the gel RFLP analysis, furthermore it also allowed to examine fragments that were below the 100bp mark, and hard to ascertain by the naked eye. In silico *Hha I* digestion revealed 4 to 6 fragments for each allelic variant, whereas the RFLP gel could only show 4 fragments. To infer the relationship between the 11 haplotypes observed by RFLP procedure (*Hha I*), 11 representative sequences corresponding to the block I of the MSP3α from each haplotype were used to construct a phylogenetic tree (Figure 14).

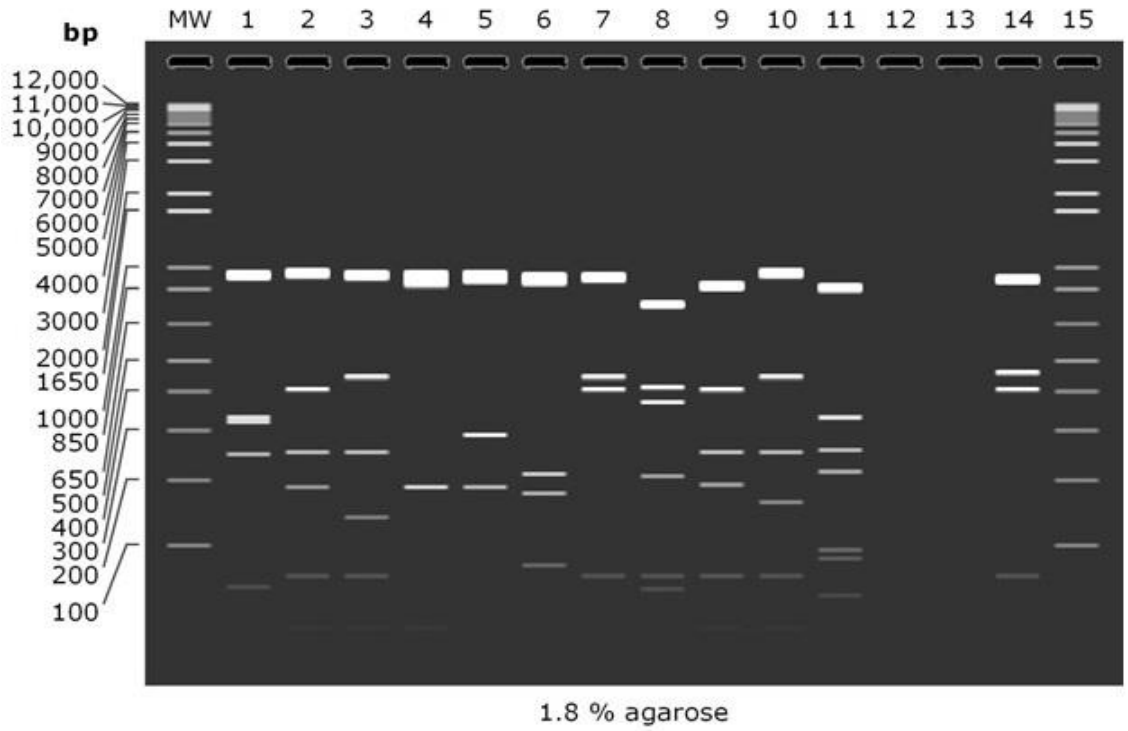


Figure 13: In-silico PCR using *Hha* I restriction Fragment; The 11 allelic variants are displayed and the reference Salvador strain (PVX_097720) in Lane 14, Ladder= 1kb plus

The block I sequences were also estimated for their nucleotide diversity (π) which stood at 0.10565.Although the study on block I was limited to local phylogenetic analysis, it is also important to note that 5 of the 8 type A allele sequences observed in this study showed 98% identity with South Korean (KU 893826.1 and KU893851.1), Indian (KC935446.1), Myanmar (EU430577.1) and Bangladesh (AF491951.1) isolates. The two type B alleles showed a 97% and 99% sequence identity to Venezuelan (AJ864953.1) and Peru (FJ612091.1) isolates respectively. The single type C size variant sequence showed a 99% identity with a Myanmar isolate (EU430598.1).

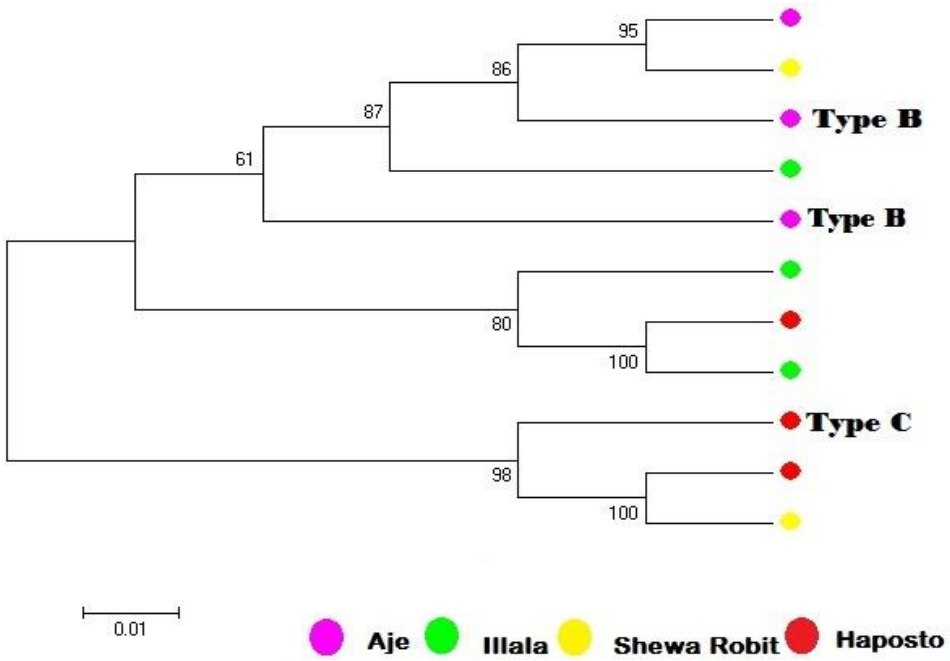


Figure 14 Phylogenetic relationship of the 11 haplotypes in the Ethiopian Isolates, based on the block I sequence of MSP3α, (1000 bootstrap for the confidence level) by maximum likelihood method. scale bar represents 1 substitution per 100 nucleotides

5.6 Genetic diversity analysis of the PvMSP3α Block II in Ethiopian isolates

To analyze the genetic diversity of the MSP3α block II component of the gene, nucleotide diversity (π) was calculated. Accordingly the nucleotide diversity for the Ethiopian population was 0.01479 with 19 haplotypes and a haplotype diversity of 0.953. To visualize the extent of diversity in the Ethiopian sequences, a sliding plot with a window length of 100 and step size of 25 sites was used. It revealed a peak plot of π= 0.065 between the position of the 390bp and 432bp of this block.

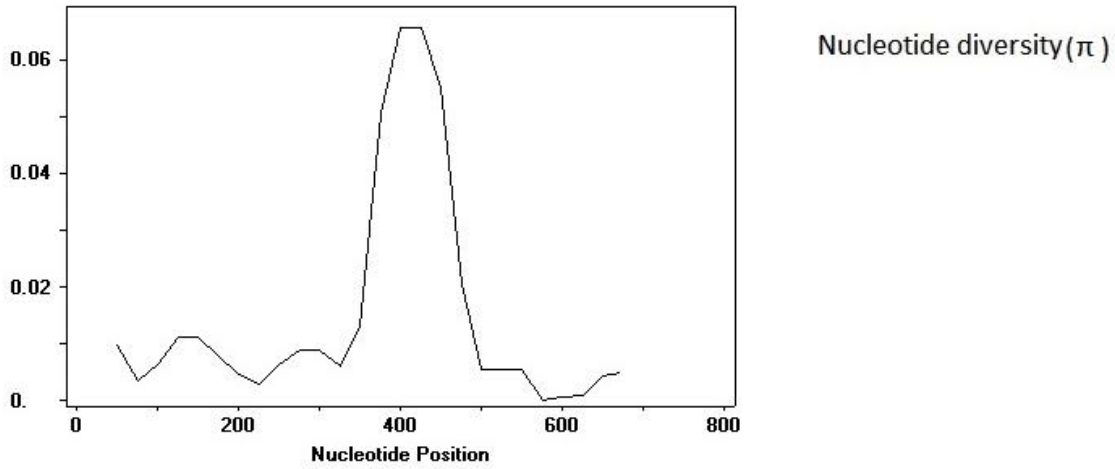


Figure 15 Average Pair wise Nucleotide Diversity, using Ethiopian MSP3α block II gene sequences

In order to put the results in a perspective the representative isolate sequences from Ethiopia were used for constructing a phylogenetic tree; to achieve this, the 717 bp corresponding to the block II of the MSP3α was used. Subsequently minimal population structuring was observed in the representative isolates; in fact isolates were clustered around the structural motif II (KETAAL/TANVVD) (Figure 16).

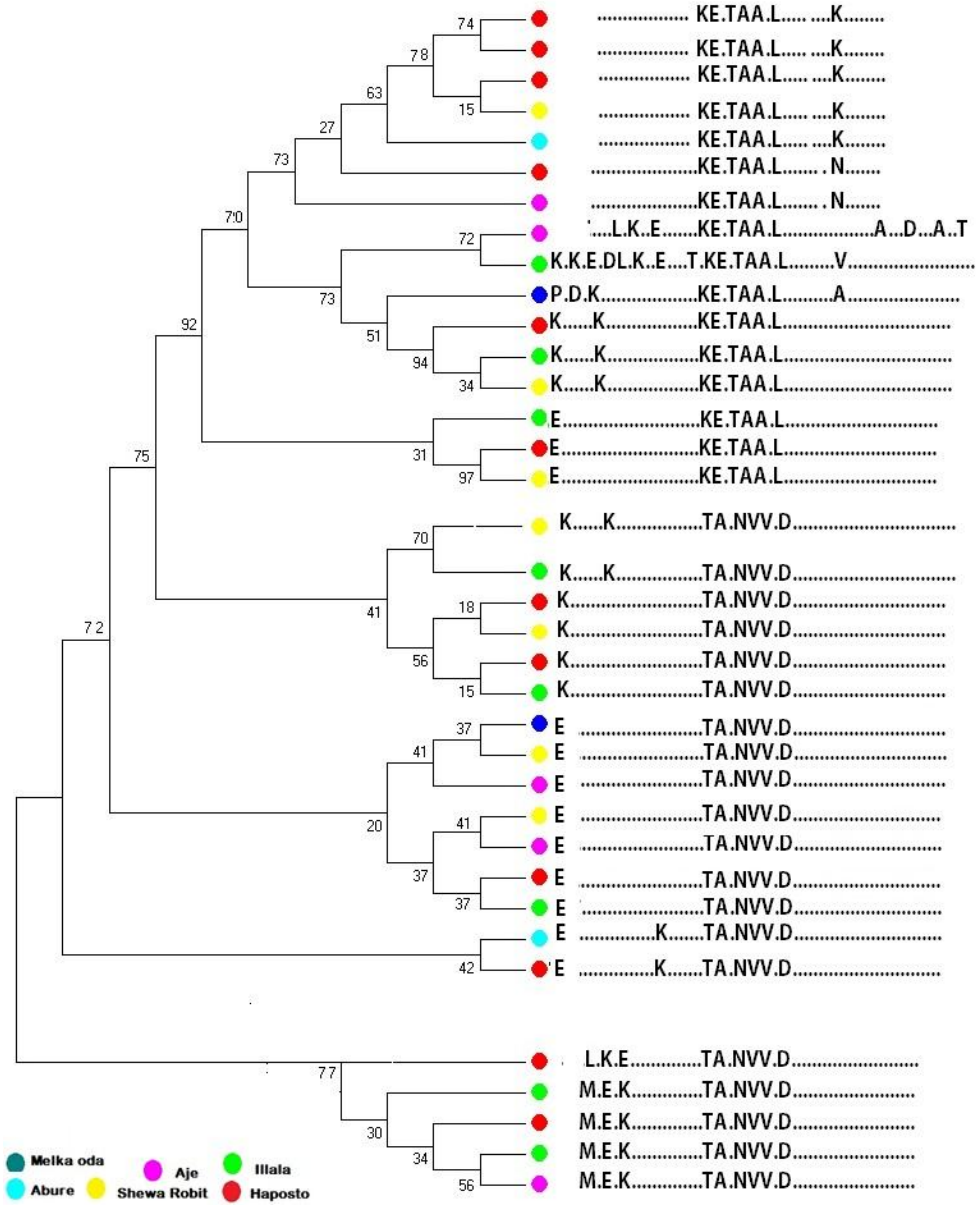


Figure 16 Phylogenetic relationship of the Ethiopian isolates based on block II amino acid sequences, (1000)bootstrap for the confidence level) by maximum likelihood method, study sites are represented by the different colors.

5.7 Genetic differentiation

To calculate population differentiation between Ethiopian and the remaining global *P.vivax* MSP3α population sequences, Wrights fixation index (F_{ST}) was used. Ten populations with representatives of more than one sequence were chosen, namely ;South Korea, Venezuela, Myanmar Sri-lanka, Brazil ,India ,Thailand ,Papua New Guinea and Indonesia. The values of Wrights fixation index indicated a low degree of differentiation between the isolates of Ethiopia, India (F_{ST}=-0.02) ,Brazil(F_{ST}=-0.02),Indonesia(F_{ST}=0.05),Myanmar(F_{ST}=0.07),and Sri-Lanka(F_{ST}=0.025), and relatively higher estimates in Thai(F_{ST}=0.11), South-Korea(F_{ST}=0.27),Papua New Guinea(F_{ST} =0.34) and Venezuelan (F_{ST}=0.36) isolates.

5.8 Selection and recombination

To calculate departure from neutrality, i.e if the block was under selection, the number of synonymous substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N) was calculated. Accordingly the null hypothesis (H_0: $d_N = d_S$) and the alternative purifying selection($d_N < d_S$) were rejected at significant values of P=0.049 and P=0.026(P<0.05) respectively. Additionally, frequency based tests of , Tajimas' D and Fu and lis' F tests were calculated for Ethiopian populations. The tests also determine whether polymorphic sites are under balancing/immune selection, where positive values with significant deviations indicating the presence of balancing selection (immune selection) whereas negative values would indicate directional selection.

Tajima s' D value for the Ethiopian MSP3α block II population was 0.4693 although deviations were not significant (P >0.10). Fu and Li's F value was 0.12946. Similarly, deviations were not significant (P>0.10). Interestingly, however, window plot analysis with window size of 11bp and step length of 1 revealed significant values of Fu and Li s F between the 327bp and the 449bp region (Figure 18); similarly the Tajima s D has significant values at said region (Figure 17); in this region Fu and Li's F was 1.7621 (P>0.05), Tajima's D was 2.64 (P>0.05).The observations indicate that the small region with significant values of the parameters were under balancing selection. Comparative analysis of the 24bp region encoding motif II (AAAGAAGCAACCGCTGCAAAATTA/ACTGCAGCAAACGTTGTAAAAGAT) revealed significant values of Fu and Li's F 2.29103(P<0.02) and Tajimas D 3.23023 (P<0.001),

whereas the rest of the block minus the motif II sequence(693 bp) revealed significant values of -5.32358(P<0.002) of Fu& li's F and -2.56670(P<0.001) of Tajima's D.

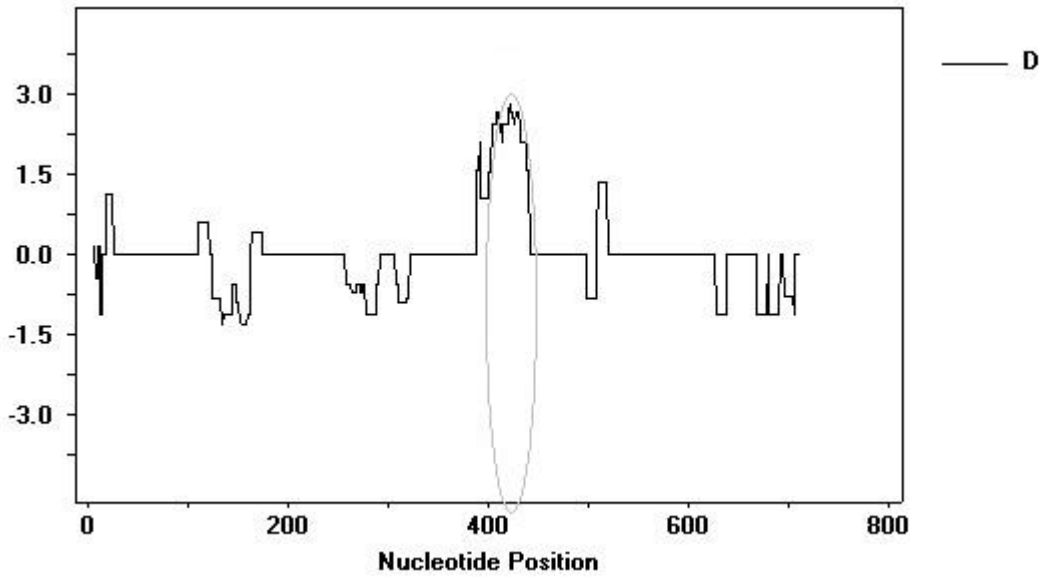


Figure 17 Tajima s' D value for Ethiopian Isolates, the region with significant values of D is circled (grey) at (P<0.05)

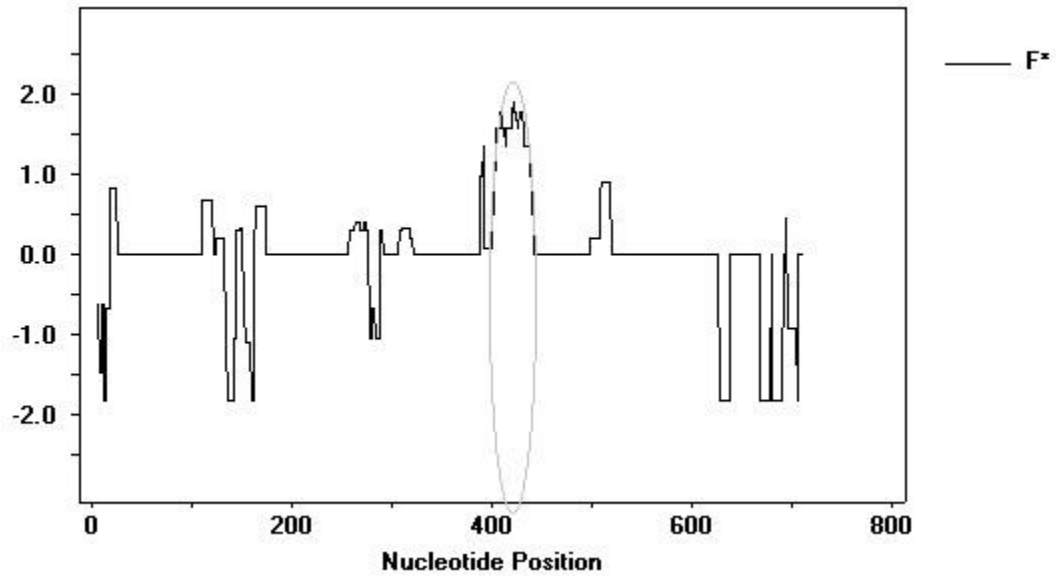


Figure 18 Fu and Li's F value for Ethiopian Isolates, the region with significant values of F (P<0.002) is circled (Grey)

Recombination has previously been indicated as a driving force in the diversity of the MSP3 gene. Accordingly 6 recombination events were detected by Dnasp in the Ethiopian population between sites; (19,122) (122,154) (154,175) (175,269) (322,400) (442,520). The sites under balancing selection mentioned above were also included in the sites where recombination occurred. This indicates that recombination and natural selection affect the MSP3α block II component of the gene. Furthermore recombination could also be detected by eye when analyzing the amino acid alignments. Apart from this however, further seven different methods were performed to detect recombination ; RDP(R), GENECONV(G), Bootscan(B), MaxChi(M), CHIMAERA(C), SisScan(S) and 3SEQ(T). Accordingly 4 recombination events were unambiguously identified by six of the seven methods (G, B, M, C, S, and T).

5.9 Global population structure and haplotype network for PvMSP3α block II

To assess and visualize the geo spatial diversity of the MSP3α block II gene globally; 161 sequences from 19 countries were used. An unrooted phylogenetic maximum likelihood tree using Tamura and Nei s' model of nucleotide substitution were constructed by performing 1000 bootstrap replicates (Figure 19). Each population was colored a specific color and ordained a specific shape to distinguish them. The Ethiopian population was represented by the color red with circle shape. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. All positions containing gaps and missing data were eliminated.

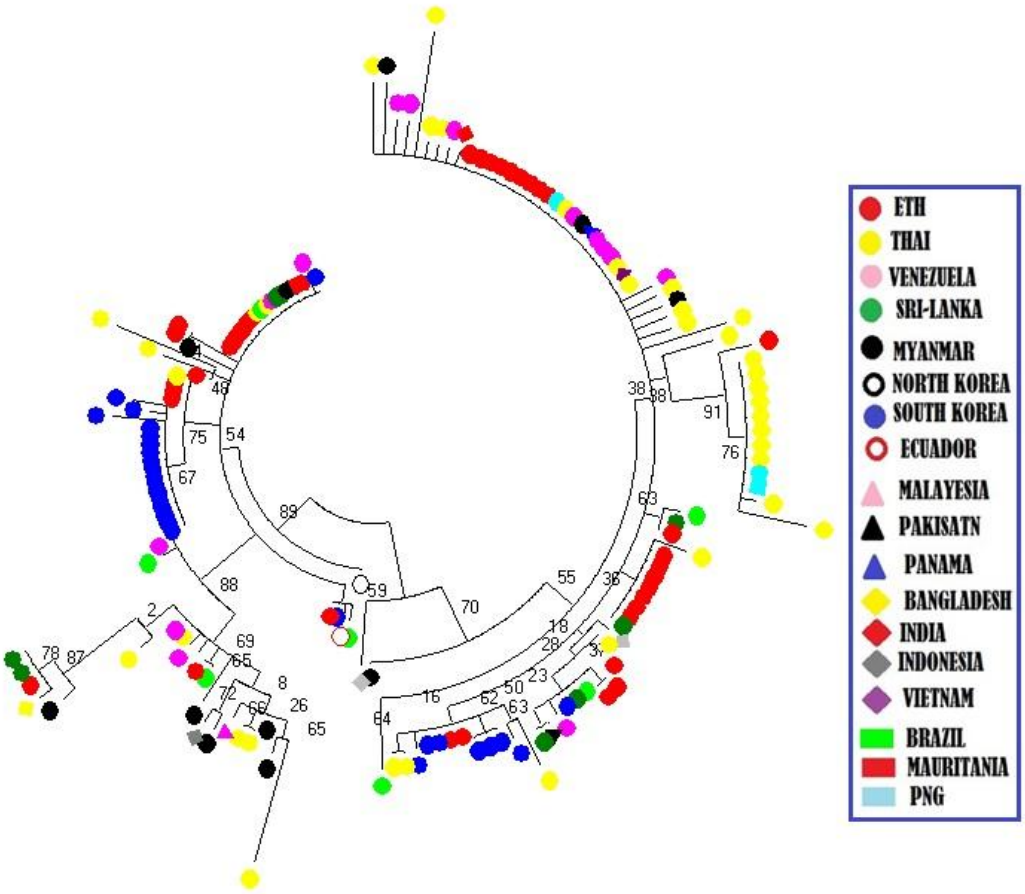


Figure 19.Maximum likelihood unrooted phylogeny of PvMSP3α block II; 161 worldwide sequences using Tamura and Nei's model of nucleotide substitution and 1000 bootstrap of the confidence level.

The absence of population structuring was evident in the global population from the phylogenetic tree. In combination with the results of population differentiation, this indicates that geographic distance has low impact in the structuring of the populations. Accordingly, the Ethiopian population was found in clusters with seemingly distant locations such as populations arising from Asian (South Korea-blue circle), African (Mauritania-red square), and South American (Brazil-green circle) countries.

To infer the relationships between the haplotypes of the 19 populations, a haplotype network was drawn by using the median joining algorithm (Figure 20). To focus on the haplotypes that were frequent in the world and relevant to vaccine design, only non-synonymous variations that were seen in more than two isolates were used to construct the haplotypes. While the focus of the study was primarily the Ethiopian population, the haplotype network was derived from the 39 (non-synonymous) haplotypes of the 19 populations. However, seven of the populations had prominent number of haplotypes and

are very well represented, accordingly the 7 populations were the Ethiopian (h=19. Red), South Korea (h=10, Blue), Venezuela (H=12, Pink), Myanmar (h=11, black), Sri-Lanka (h=7, deep green), Thailand (h=25, Yellow) and Brazil (H=7, light Green). Accordingly 9 haplotypes with prominent frequency were commonly observed in the populations. Three of the haplotypes Hap 1, Haplotype 2 and Hap 4 were observed in 51% of the isolates in this study, meanwhile Hap 13, Hap 9, Hap 8, Hap 24 and Hap 33 were also readily observed with descending frequency, respectively. Haplotype 2 included sequences from populations of Ethiopia, Sri Lanka, Papua New Guinea, Venezuela, Thailand, Myanmar, Vietnam and Panama. The second most frequent haplotype 1 also included sequences from the above populations as well as from Brazil, Ecuador, South Korea, India and Mauritania. Between the two haplotypes they represent 13 of the 17 Endemic *P. vivax* countries included in this study.

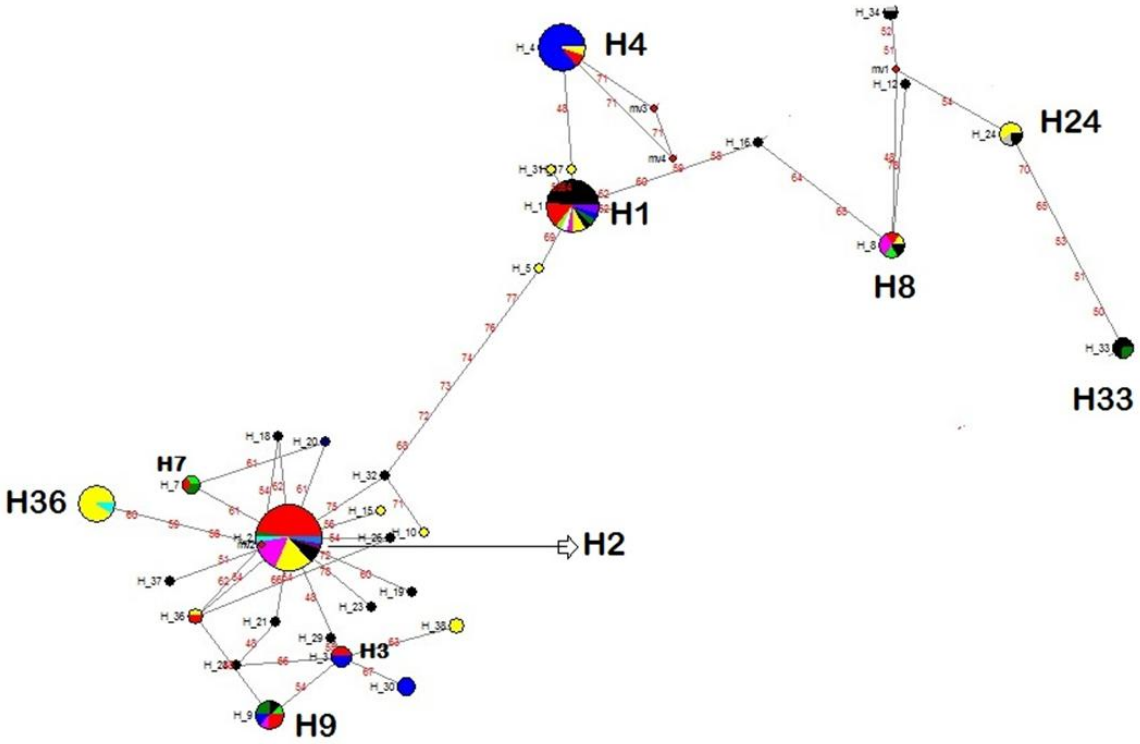


Figure 20 Haplotype network constructed using only non-synonymous variations. Thirty nine haplotypes composed of 23 common amino acid polymorphisms were analyzed using the Median Joining algorithm. Nodes represent the haplotypes and lines indicate connections between them. The size of each node indicates haplotype frequency.

6. Discussion

Plasmodium vivax is the most geographically widespread of the *Plasmodium* species that afflict human beings. The prevalence of *P. vivax* is, however, most felt in South America and Asian countries where it is responsible for as much as 70% of malaria infections (54). The parasite is noticeably absent from much of African countries with the prominent exception of the East African corridor. Chief among the East African countries is Ethiopia, having the largest geographical presence as well as highest population number in the region. Concurrently, Ethiopia also has one of the highest rate of *P. vivax* prevalence both in the horn of Africa and globally; in fact, by the year 2015, it was the second highest contributor to *P. vivax* infections worldwide next to India (224), In parallel it also had the second highest deaths(12%) associated with *P. vivax* infections, preceded only by India(2). While the omnipresence and prevalence of *P. vivax* infections have been reported from many segments of the country; much remains in terms of understanding the population genetics of the parasite. Indeed two genetic studies have previously analyzed population structure, genetic diversity and multiple clone infections of the parasite in Ethiopia, using microsatellite markers(48,225). However, these are not enough given the extensive diversity and dynamic nature of the parasite, rather it is of utmost importance to continue tracking and analyzing the population genomics of the parasite in different parts of the country. Furthermore, these studies are fundamental for understanding transmission dynamics and for tracking important phenotypic variants of the parasite such as drug resistance genes or antigenic variants in different transmission settings, as this aid in designing future control and elimination strategies. Therefore, our study investigated the extent and dynamics of sequence diversity in the vaccine candidate component of the antigen (block II only), and examined evidence of balancing selection (immune selection) and recombination to identify immune escape mechanisms in this antigen loci for *P. vivax*. Finally, we used the complete sequence of the antigen loci to investigate genetic diversity and frequency of multiple clone infections using an established rapid PCR-RFLP procedure. To understand the mechanism behind the molecular variation observed in the RFLP alleles and to infer phylogenetic relationship between identified local haplotypes, representative isolates were sequenced and studied.

In this cross sectional study, the proportion of *P. vivax* was significantly higher in four of the six study sites. Our results maybe a reflection of previous studies, which found a similar trend(higher proportion of *P. vivax*) in different parts of the country(15,115,226).

In various studies, age distribution has been an indicator of endemicity in malaria infections; for *P. vivax* it usually afflicts infants in areas that are hyper-endemic, as immunity is attained faster as compared to *P. falciparum*(227). Contrastingly in low transmission settings where most infections appear as submicroscopic infections, all age groups are at risk(228). In the current study all age groups were indeed infected to some extent, although it was higher in age group of 15 and above, (64.38%). Moreover two previous studies in the same study sites have indicated the prevalence of a high number of submicroscopic infections as reflected by PCR(123,229).Evidently the demographic results are what is expected under the characteristics of a low transmission setting(228).

The MSP 3α gene was used to assess *P. vivax* diversity and multiplicity of infections as has been performed in much of the *P. vivax* endemic countries(198). In most of these studies, three size variants Type A(1.8kb), Type B(1.5kb) and Type C(1.1) were the most frequently observed with descending frequency respectively; whereas a fourth variant(0.5kb) is very rarely reported. The current study found a similar trend with 82.97%, 12.7% and 4.2% prevalence for Type A, B and C respectively. This result was most notably similar to that of a study done in Sri-Lanka that reported 84.4%, 13.1% and 2.2% for the three size variants(230). However, it was quite higher in prevalence of the type A size variant than those reported from Venezuela (59.3%), Brazil (68%), Thailand (74.8%) and India (75.4%) but much lower from that reported in Colombia (96%)(38,151,192,207,232).The prevalence of Type B size variants on the other hand was lower than those reported from Venezuela (21.9%) and Thailand (18.7%) but comparable to those from Brazil (15%) and India (14.3%). The prevalence of the Type C variants observed in this study was uniformly lower as compared to much of the results reported from other countries with the exception of Sri-Lanka (2.2%), but significantly differed from a study in India which reported 70% prevalence(34). Although all of the studies included in this review reported a lower prevalence of the type C variant to that of

the Indian study, they were still higher than the one observed here, particularly in Venezuela (18.8%), Brazil (17%) and India (10%) (151,207,231).

Given the higher prevalence of the type A variant, it has been postulated that it might confer a selective advantage over the other variants, although the occurrence of the smaller size variants in every studied population so far, seems to indicate that they can counterbalance the fitness cost associated with deletions in their genes; albeit reducing fitness to some extent, as they are found at a smaller prevalence(192,151). Relationships between the three size variants has been previously analyzed in Thai, Sri-Lankan and Venezuelan isolates with differing results(151,192,233). The phylogenetic tree results from the Sri Lankan and Venezuelan studies indicated that both type B and type C variants were derived from type A, and given the very low sequence diversity in type B and type C variants both studies concluded that the type A was the ancestral form, this is in contrast to the Thai study where two of the C variants clustered to different parts of the phylogenetic tree(151,233).Our study is quite similar to the former two studies where type A size variant formed three separate clades, and the type B and type C variants clustered on different clades formed by the Type A size variant. The exceptional case of the Thai study can be explained by the fact that samples were originated from different parasite populations(151).Although this study limited phylogenetic analysis to local isolates, a previous study has reported that similar sequences don't come from similar geographic origin, a case confirmed by subsequent studies that reported a lack of geographic grouping(32, 36,198,).Similarly in this study, BLAST analysis revealed that each of the 11 allelic genotypes identified by PCR-RFLP were consistently paired with sequences from 11 different countries. Two opposing explanations have been presented to explain this phenomenon: the first is that ''phylogenetic relationship between isolates is masked by the recent extensive polymorphism as a result of similar selective pressures operating on geographically different lineages to produce similar outcomes; or the ancestral PvMSP3a is polymorphic and extensive recombination events have obscured the phylogenetic relationships of the ancestral lineages''(32,36).

Furthermore, we investigated into the MSP3α diversity and multiplicity of infection using restriction enzyme *Hha I*. Using this restriction enzyme, in total 11 different haplotypes were observed, which are quite similar to the findings of studies done in

Brazil(11) Papua New Guinea(11) and French Guiana(11) but higher than those reported from Venezuela(9) and Columbia(9)(63,151,231,232,234). However, three other countries reported a higher haplotype number, particularly Peru(17) and two more studies from India and Thailand that reported 14 and 13 haplotypes respectively(34,192,235). Interpretation of this results have, however, so far been limited to observations of how variable the antigen loci is, and comparison of RFLP gel images to identify similar alleles in different geographic regions. Even in the case of the latter it has been pointed out that two alleles with similar *Hha I* RFLP pattern may not necessarily mean they are not divergent at the nucleotide level and in fact might be a result of ‘’similarity by RFLP identity, but not by descent’’(198). However, the results generated here were to the contrary, by analyzing sequences from the same RFLP pattern categories we found that the sequences were largely similar, further supported by a quite low pairwise nucleotide diversity (π=0.002-0.003). The nucleotide diversity for block I was 0.10565, comparable to Thai isolates (0.10621) nucleotide diversity of the complete sequence was 0.43(π), higher than the one in Brazil (0.34), but otherwise lower than those observed from Venezuela (0.049) and Thailand (0.065). To put the results in global context; the South American continent had a 0.056 and the Asian continent 0.050 whereas the global diversity stood at 0.049 (32, 36,231).

In silico RFLP using the 14-representative sequence revealed that all of the *Hha I* restriction recognition sites were located on the block I ;and the ≈1000bp fragment located at the top was in fact the intact block II segment. This is quite important as previous studies have used *Hha I* RFLP patterns and the complete sequence of the gene collectively to study phylogenetic relationships when in fact restriction fragment analysis are only limited to the block I region only. Furthermore the two blocks of the MSP3α (block I & block II) together and separately give rise to three quite different topologies in cladograms. This could be a result of an extensive recombination or different parts of the gene evolving at different paces(36). Nevertheless any molecular phylogenetic study planning on using genotypes identified by *Hha I* RFLP and sequencing data in tandem should consider using only the cladogram that actually involves recognition site of the restriction enzyme in that particular block. On the other hand in-silico digestion using *Alu I* showed that it might be a good alternative to *Hha I* digestions, as almost every band of

every pattern revealed that they were highly polymorphic. But distinguishing this restriction patterns is quite hard as reported from studies in Thailand and Papua New Guinea(192,234). A similar case was detected in this study where fragments below the 500 bp fragment appeared unresolvable using the normal agarose gel electrophoresis; in-silico RFLP analysis revealed that some of these fragments were actually 10 to 23 base pairs apart and as a result hard to discern. Hence a revision in the gel electrophoresis procedure, or a change in the type of gel might be required to clearly visualize this particular restriction fragment pattern given that all of the studies use a quite similar strategy.

A common observation on *P. vivax* apart from its remarkably high diversity however, is the incidence of multiple clone infections, which are infections caused by more than one genotype of the parasite. Multi-clonal *P. vivax* infections are frequently observed regardless of the transmission settings and are associated with increased virulence, drug resistance and transmissibility(160,236,237). In this study, of the 50 samples analyzed 12% were identified to be multi-clonal infections; in a similar manner a Sri Lankan study has reported a 13% prevalence of multiple clonal infections, but again this is quite lower than that reported from Columbia (36.4%) ,Papua new Guinea(23%) and Thailand (19%) but it is higher than the one reported from India (8.2%) (33, 38, 192, 207).The results here are much lower than those reported from recent studies that used microsatellite markers to investigate multi-clonality, in south western parts of Ethiopia. Three sites in the Southern Nations Nationalities and Peoples region, namely Hawassa, Arbaminch and Halaba had significant levels of multi-clonality at 67%, 44% and 21% polyclonal infections whereas the fourth site Badawacho (8%) had lower polyclonal infection as compared to our study(47). This might be a result of sample size, sampling site (geographical difference) seasonality of infections or detection methods, as microsatellite markers are relatively free from strong selective constraints, although we can't rule out amplification bias in microsatellites, where PCR artefacts can lead to an overestimation of multiple clone infections(160).

Regardless of the levels observed in different parts of Ethiopia, multiple clone infections are a public health concern. Although quite hard to distinguish relapses from primary

infection, based on seasonal patterns ,a study in east Shoa zone has indicated that as much as 40% of the infections can be attributed to relapses following the major rainy season(226). Particularly in *P. vivax*, its relapsing nature provides an opportunity for increasing clonal diversity, where a genetically distinct new infection joins a dormant hypnozoite from a previous infection (160).Among other things relapses can also lead to higher multiplicity of infections, and within the host this can result in a competition for resources, where traits such as virulence and drug resistance are favored (160,236).Drug resistance has already been reported for CQ in most areas of Ethiopia. For *P. vivax* this includes the currents study sites, particularly in Shala district(42,125,238–240). These all are disconcerting to efforts that aim to control or interrupt *P. vivax* transmission. Further complications such as contraindication of the hypnozoiticidal PQ in G6PDd individuals, pregnant or lactating mothers mean that hypnozoites will remain a major hindrance for *P.vivax* unless the issues are resolved or alternatives are sought.

The most convenient, cost effective alternative to yet be included in the current armamentarium of anti-malarial's are vaccines. This is in part because researches have been hampered by the lack of *in vitro* culture to study the *P. vivax* parasite. Instead population genetic studies in complement with molecular and immunological studies take center stage in advancing knowledge towards a broadly efficacious vaccine. So far population genetic studies have revealed a quite extensive diversity in most potential vaccine candidate antigens. Furthermore, in complement with molecular studies for some candidate antigens like *P .vivax,* Apical Membrane Antigen(PvAMA-1), an association has been made between their high polymorphism and how they use it as an immune escape mechanism(188).Concurrently other *P. vivax* endemic countries have included population genetic analysis in pre-clinical trials of candidate antigens. Ethiopia, while a major contributor for *P. vivax* has so far remained uninvolved in efforts of developing a vaccine.

Here, for the first time the sequence diversity of a vaccine candidate antigen, the PvMSP3α block II, where due in large part, to its highly conserved nature and a remarkable immunogenicity has been recognized as an ideal vaccine; is assessed (39,40). So far studies in 11 *P. vivax* endemic countries have revealed nucleotide diversity (π) in

this block, ranging from 0.015(Brazil) to 0.023(India) while the global diversity was 0.019. (39). This is comparable to other vaccine candidate antigens like PvAMA-1(π=0.01653), *P. vivax* Thrombospondin Related Anonymous Protein(PvTRAP) (0.0059) *P. vivax* Merozoite Surface Protein 1(PvMSP1) 0.0129 and *P. vivax* Duffy Binding Protein (PvDBP) 0.01103(241).In the current study nucleotide diversity (π) for Ethiopian PvMSP3α isolates was 0.01479 with 19 haplotypes and haplotype diversity of 0.953 .Consistent with previous reports, block two was more conserved than block I (π=0.10565)(36,39,242,243).It has been suggested that the reason behind this peculiar difference lies in the structural arrangement of the gene, where the block II component is under functional constraint and necessary for MSP3α function such as providing peripheral interaction with other membrane bound proteins; whereas the divergent block I appears dispensable(32)(191). Despite this, however, a sliding window approach revealed a remarkable nucleotide diversity (π=0.65) between 390bp-432bp of the block II. Similar observations have been reported in PvMSP3α worldwide isolates with a peak diversity of 0.069(39). This region encodes structural motif II ,which has dimorphic alleles TAANVVKD and KEATAAKL, indeed another region with a dimorphic allele was also identified(MSELEK and LSKLEE) at distant part of the gene, although its dimorphism was at a lesser extent. In contrast, the dimorphic alleles of motif II were equally prevalent (1:1) in the isolates; this also has been a constant theme in different populations studied over the past decade(32,36,151,198).

In a previous in-silico study, both motifs have been predicted to be B cell epitopes with >75% specificity, particularly motif II has been reconfirmed as a B cell epitope in previous *in vitro* studies(http://www.iedb.org/) (39).Hence it was quite interesting that both dimorphic alleles of motif II were found to be equally prevalent ,since a conserved pattern of variation in antigenic sequences is associated with either structural constraint or a strong frequency dependent immune selection(39,150). In particular diversity in genes encoding antigens especially on the sporozoite and merozoite are results of natural selection imposed by the immune system. Accordingly test of neutrality were performed and, the null hypothesis H_0: $d_N = d_S$ (that polymorphism is neutral) versus alternative hypothesis of purifying or positive selection was rejected at a significant value (P=0.049,P<0.05).Similarly the negative selection hypothesis was rejected

(P=0.026P<0.05),suggesting that positive selection might be operating on block II. This finding is quite similar to the ones observed in a study by Rayner and colleagues in this antigen loci(32) using worldwide isolates. In positive selection, genetic variants favored by this pressure will either increase in frequency or be maintained, as has been the case for *P. falciparum* vaccine candidate antigens such as MSP-1 MSP-2 TRAP and AMA-1(147). Although in the current study, positive values were observed for both Tajima's D and Fu and Li's F across the length of the block, highly significant values(Tajimas D 2.64,P>0.05;Fu &Li's F 1.7621,P>0.05) were observed only in one specific region, the region encoding motif II. Further comparative analysis of the structural motif II (24bp) and rest of the block (692bp) also revealed a highly significant positive value for both tests in the structural motif II (Tajima's D 3.23023,P<0.001;Fu&Li's F 2.29103 P<0.02). In contrast, the rest of the block had significant negative values (Fu& li's F -5.32358,P<0.002 ; Tajima's D -2.56670 P<0.001). This would indicate that, while positive selection is operating on the entire block thus reducing diversity, the small region encoding motif II is under balancing selection (Immune selection).

As further evidence on the effects of balancing selection genetic differentiation estimates for select population sequences revealed low F_{ST} estimates. Moreover, F_{ST} values between Ethiopian and Brazilian (F_{ST} = -0.02), Ethiopian and Sri Lankan MSP3α block II isolates (F_{ST} =0.025) are lower than values attained using Single Nucleotide Polymorphism markers(SNP) for both pairs (Ethiopia& Sri lanka F_{ST} = 0.21; Ethiopia &Brazil F_{ST} = 0.31)(43). However this data should be interpreted with caution because of the small sample size in the study. Apart from F_{ST} estimates, evidence of population structuring was also minimal as observed using maximum likelihood tree for both local and global phylogenetic alignments indicative of extensive gene flow. Instead, the phylogenetic tree clades were clustered based on the structural motif I&II dimorphic alleles. This observation is further supported by four studies, each of which used isolates that represent different geographic origins(32,36,39,192). Another important observation was the low bootstrap values gained when constructing a phylogenetic tree for block II, regardless of Amino acid or Nucleotide alignment, regardless of the method used in constructing the tree (Maximum likelihood, Neighbor joining, Bayesian, Maximum parsimony, UPGMA) or indeed substitution models(Jukes Cantor, Kimura 2 parameter,

Tamura Nei ,General time reversible). A similar situation had also occurred for two independent studies (36,198). This could be a result of the small length of the sequences to be analyzed(717bp),or recombination signals that can also obscure evolutionary relationship amongst the isolates themselves(150,198). Recombination is evident by eye where any combination of the two structural motifs(motif I&II) is observed. Furthermore, DNASP estimated six recombination events whereas the RDP program estimated 4, the difference between the two results can be explained by singletons which can affect DNASP results. Nevertheless, recombination was observed from isolates of different geographic origins and between different PCR size classes and throughout the phylogenetic tree generated by the RDP program. Indeed, frequent recombination in MSP3α have been reported in previous studies where recombination in the block II can be two to five times higher than the block I component(151).Similarly, Mascorro and colleagues found increased number of recombination events in the block II as opposed to block I and found cladograms constructed from the block I more informative with better bootstrap support, similar to our study(36).In support of the evidence in recombination, multi-clonal infections(MCI) can facilitate recombination, and in our study alone we did find 12% prevalence of MCI which would give ample opportunity for meiotic recombination in the mosquito midgut (the malaria parasite is a haploid for most of its life cycle in the human host) .

Regardless of the source of genetic variation or maintenance of these variations, effective vaccines should include alleles that can cover the prevailing antigenic diversity. Therefore, it was encouraging that from the constructed 39 haplotypes (using non-singleton non-synonymous variations) three of the haplotypes were shared in 51% of the sequences included in the current study. However, it is important to bear in mind that due to the limited number of samples, results should be interpreted with caution. Nonetheless the results were more positive than the vaccine candidate antigen PvAMA-1, where only 15% of the haplotypes were shared among studied isolates(244).This, however, not to say that the other haplotypes warrant further investigations, as most of them are clustered around these three distinct haplotypes and might be involved in evasion of naturally acquired-or vaccine-mediated immune responses, hence vaccine formulations should not comprise of only this haplotypes.

7. Conclusion

- The *P. vivax* population in Ethiopia is genetically diverse even in low transmission settings

- The *P. vivax* population in Ethiopia is also expanding, with significant genetic complexity, this is in tandem with other studies which indicated a surge in *P. vivax* prevalence in Ethiopia, this suggest that current prevention/control mechanisms aren't having the required effect on the parasite.

- The PvMSP3α block I component is highly polymorphic and hence this component of the gene is an effective promising epidemiological marker through rapid size restriction analysis.

- The PvMSP3α block II is a largely conserved region and is more suitable for designing a multivalent subunit based vaccine as compared to the highly polymorphic block I component of the gene.

- A small region in PvMSP3α block II encoding a predicted B cell epitope is under frequency dependent selection and might be involved in immune escape The region under a frequency dependent selection is structural motif II(TAANVVKD / KEATAAKL) where the dimorphic alleles are equally prevalent, Therefore future vaccine design strategies targeting PvMSP3α block II should put into consideration the identified antigenic polymorphism from this study, as they might constitute an immune/vaccine escape mechanism.

- The majority of circulating MSP3α block II alleles share three haplotypes, this may serve as a starting point for designing vaccines targeting this antigen.

8. Recommendations

- Given its remarkable diversity and previous reports of a surge in prevalence, attention should be given to the *P .vivax* parasite. The Ethiopian national strategies and policy to combat malaria, should duly acknowledge this fact urgently and act accordingly

- Molecular epidemiological studies of PvMSP3 α need to be assessed in Ethiopia to complement this study with regards to studying its vaccine efficacy

- Immunogenicity and protection ability of PvMSP3α as a vaccine candidate antigen need to be assessed in Ethiopian context

- Population genetic studies need to be incorporated in efforts to prevent and control the *P. vivax* parasite. Furthermore, studies should also focus on the diversity of antigen as they represent potential vaccine candidates.

  .

9. References

1. Barber BE, Rajahram GS, Grigg MJ, William T, Anstey NM. World Malaria Report : time to acknowledge Plasmodium knowlesi malaria. *Malaria J* 2017;16(135):13–5.

2. World Health Organization. *World Malaria Report 2016*. WHO Press; 2016. 186p.

3. Autino B, Noris A, Russo R, Castelli F. Epidemiology of Malaria in Endemic Areas. *Mediterr J Hematol Infect Dis.* 2012;4:1–11.

4. Guerra CA, Howes RE, Patil AP, Gething PW, Boeckel TP Van, Temperley WH, et al. The International Limits and Population at Risk of Plasmodium vivax Transmission in 2009. *PLoS Negl Trop Dis*. 2010;4(8).

5. Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, et al. Global Epidemiology of Plasmodium vivax. *Am J Trop Med Hyg.* 2016;95(Suppl 6):15–34.

6. Abdelbagi A, Pirahmadi S, Abouie A, Dinparast N. Infection , Genetics and Evolution Molecular genetic analysis of Plasmodium vivax isolates from Eastern and Central Sudan using pvcsp and pvmsp-3 a genes as molecular markers. *Infect Genet Evol* . 2015;32:12–22.

7. World Health Organization.*World Malaria Report. 2014.*WHO Press;2014.242p.

8. World Health Organization. *World Malaria Report 2017*. WHO Press; 2017. 196p.

9. Enweji, N. 2014. Dynamics of Resistant Plasmodium falciparum Parasites. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1020. 49 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9007-2.

10. Alemu A, Abebe G, Tsegaye W, Golassa L. Climatic variables and malaria transmission dynamics in Jimma town , South West Ethiopia. *Parasite&Vectors* ; 2011;4(1):30.

11. Teklehaimanot HD, Lipsitch M, Teklehaimanot A, Schwartz J. Weather-based prediction of Plasmodium falciparum malaria in epidemic-prone regions of Ethiopia I . Patterns of lagged weather effects reflect biological mechanisms. *Malaria J.* 2004;3(41):1–11.

12. Yeshiwondim AK, Gopal S, Hailemariam AT, Dengela DO, Patel HP. Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. *Malaria J.* 2009;8(5):1–11.

13. Vajda ÉA, Webb CE. Assessing the Risk Factors Associated with Malaria in the Highlands of Ethiopia : What Do We Need to Know ? *Trop Med Infect Dis.* 2017;2(4):1–13.

14. Ramos JM, Reyes F, Tesfamariam A. Change in Epidemiology of Malaria Infections in a Rural Area in Ethiopia. *J Travel Med*. 2003;12:155–6.

15. Sena LD, Deressa WA, Ali AA. Analysis of trend of malaria prevalence in south-west Ethiopia : a retrospective comparative study. *Malaria J*. 2014;13(188):1–9.

16. Argaw MD, Woldegiorgis AGY, Abate DT, Abebe ME. Improved malaria case management in formal private sector through public private partnership in Ethiopia : retrospective descriptive study. *Malaria J.* 2016;15(352):1–11.

17. Zhou G, Yewhalaw D, Lo E, Zhong D, Wang X, Degefa T, et al. Analysis of asymptomatic and clinical malaria in urban and suburban settings of southwestern Ethiopia in the context of sustaining malaria control and approaching elimination. *Malaria J*.2016;15(250):1–9.

18. Alemu A, Muluye D, Mihret M, Adugna M, Gebeyaw M. Ten year trend analysis of malaria prevalence Kola Diba, North Gondar,Northwest Ethiopia. *Parasite&Vectors.* 2012;5(173):1–6.

19. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* .2015.

20. White NJ. Determinants of relapse periodicity in Plasmodium vivax malaria Determinants of relapse periodicity in Plasmodium vivax malaria. *Malaria J* 2011;10(297):1–35.

21. Hedrick PW. Population genetics of malaria resistance in humans. Heredity . *Nature* 2011;107(4):283–304.

22. Mueller I, Rushdi A, Chitnis CE. Development of vaccines for Plasmodium vivax malaria ☪. *Vaccine*.2015;33(52):7489–95.

23. Moreno-p DA, Ru JA. Reticulocytes: Plasmodium vivax target cells. *Biol Cell*. 2013;105:251–60.

24. Barry AE, Arnott A. Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Front Immunol.* 2014;5(359):1–16.

25. Luo Z, Sullivan SA, Carlton JM. The biology of Plasmodium vivax explored through genomics. *Ann N Y Acad Sci*. 2016;1342(1):53–61.

26. Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, et al. Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat Genet.* 2016;(May).

27. Jennison C, Arnott A, Tessier N, Tavul L, Mueller I, Barry AE. Plasmodium vivax Populations Are More Genetically Diverse and Less Structured than Sympatric Plasmodium falciparum Populations. *PLoS Negl Trop Dis*. 2015;9(4):1–20.

28. Barry AE, Beeson J, Reeder JC, Fowkes FJI, Arnott A. Using Population Genetics to Guide Malaria Vaccine Design. In: Omolade O, editor. *Using Population Genetics to guide Malaria Vaccine Design, Malaria Parasites*. InTech; 2012. p. 231–64.

29. Patarroyo MA. Vaccines against Plasmodium vivax : a research challenge. *Expert Rev*. 2012;11(10):1249–60.

30. Plebanski M, Flanagan KL. The Economics of Malaria Vaccine Development. *Trends Parasitol* . 2017;33(3):154–6.

31. Patarroyo MA. What Is Known about the Immune Response Induced by Plasmodium vivax Malaria Vaccine Candidates ? *Front Immunol*. 2017;8(126):1–23.

32. J.C.Rayner,V.Corredor,D.Feldman,P.Ingravallo,F.iderbdullah,M.R.Galinski. Extensive polymorphism in the Plasmodium vivax merozoite surface coat protein MSP-3 α is limited to specific domains. *Parasitology*. 2002;(125):393–405.

33. Bruce MC, Galinski MR, Barnwell JW, Snounou G, Day KP.Polymorphisms at the Merozoite surface protein-3 α locus of Plasmodium vivax: global and local diversity. *Am J Trop Med Hyg.* 1999;61(4):518–25.

34. Zakeri S, Barjesteh H, Djadid ND. Merozoite surface protein-3 α is a reliable marker for population genetic analysis of Plasmodium vivax. *Malaria J.* 2006;5(53):1–6.

35. Cui L, Escalante AA, Imwong M, Snounou G. The genetic diversity of Plasmodium vivax populations. *Trends Parasitol*. 2003;19(5):220–6.

36. Mascorro CN, Zhao K, Khuntirat B, Sattabongkot J, Yan G. Molecular evolution and intragenic recombination of the merozoite surface protein MSP-3 a from the malaria parasite Plasmodium vivax in Thailand. *Parasitology*. 2005;131:25–35.

37. Nam DH, Oh JS, Nam MH, Park HC, Lim CS, Lee WJ, et al. Short Report : Emergence of New Alleles of the MSP-3 α Gene in Plasmodium vivax Isolates from Korea. *Am J Trop Med Hyg*. 2010;82(4):522–4.

38. Cristiano FA, Pérez MA, Nicholls RS, Guerra AP. Polymorphism in the Plasmodium vivax msp 3 α gene in field samples from Tierralta , Colombia. *Mem Inst Oswaldo Cruz.* 2008;103(August):493–6.

39. Gupta B, Reddy BPN, Fan Q, Yan G, Sirichaisinthop J. Molecular Evolution of PvMSP3 α Block II in Plasmodium vivax from Diverse Geographic Origins. *PLoS One*. 2015;10(8):1–16.

40. Stanisic DI, Javati S, Kiniboro B, Lin E, Jiang J, Singh B, et al. Naturally Acquired Immune Responses to P . vivax Merozoite Surface Protein 3 a and Merozoite Surface Protein 9 Are Associated with Reduced Risk of P . vivax Malaria in Young Papua New Guinean Children. *PLoS Negl Trop Dis.* 2013;7(11):1–8.

41. Guimara C, Carolina M, Silva I, Jesus C, Silva M, Barnwell JW, et al. Naturally acquired antibodies to Plasmodium vivax blood-stage vaccine candidates (PvMSP-1 19 and PvMSP-3 a 359 À 798) and their relationship with hematological features in malaria patients from the Brazilian Amazon. *Microbes Infect.* 2012;14:730–9.

42. 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. *Malaria J*. 2014;13(1):1–8.

43. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, et al. Development of a Single Nucleotide Polymorphism Barcode to Genotype Plasmodium vivax Infections. PLoS Negl Trop Dis. 2015;9(3):1–18.

44. Golassa L, Enweji N, Erko B, Aseffa A, Swedberg G. Detection of a substantial number of sub-microscopic Plasmodium falciparum infections by polymerase

chain reaction: a potential threat to malaria control and diagnosis in Ethiopia. *Malaria J*. 2013;12(352):1–16.

45. Tadesse FG, Hoogen L Van Den, Lanke K, Schildkraut J, Tetteh K, Aseffa A, et al. The shape of the iceberg : quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria J*. 2017;16(99):1–11.

46. Tsegaye A, Golassa L, Mamo H, Erko B. Glucose-6-phosphate dehydrogenase deficiency among malaria suspects attending Gambella hospital , southwest Ethiopia. *Malaria Jl*. 2014;13(438):1–7.

47. Getachew S, To S, Trimarsanto H, Thriemer K. Variation in Complexity of Infection and Transmission Stability between Neighbouring Populations of Plasmodium vivax in Southern Ethiopia. *PLoS One.* 2015;10(10):1–15.

48. Gunawardena S, Karunaweera ND, Ferreira MU, Phone-kyaw M, Pollack RJ, Alifrangis M, et al. Geographic Structure of Plasmodium vivax : Microsatellite Analysis of Parasite Populations from Sri Lanka , Myanmar , and Ethiopia. *Am J Trop Med Hyg*. 2010;82(2):235–42.

49. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax Malaria : Neglected and Not Benign. *Am J Trop Med Hyg*. 2007;77(6):79–87.

50. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M, et al. Multidrug-Resistant Plasmodium vivax Associated with Severe and Fatal Malaria : A Prospective Study in Papua , Indonesia. *PLoS Med*. 2008;5(6):890–9.

51. Price RN, Seidlein L Von, Valecha N, Nosten F, Baird JK, White NJ. Global extent of chloroquine-resistant Plasmodium vivax : a systematic review and meta-analysis. *Lancet Infect Dis*. 2014;14(10):982–91.

52. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, et al. Geographical variation in Plasmodium vivax relapse. *Malaria J*. 2014;13(144):1–16.

53. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, et al. Key gaps in the knowledge of Plasmodium vivax , a neglected human malaria parasite. *Lancet Infect Dis*. 2009;9(9):555–66.

54. Adams JH, Mueller I. The Biology of Plasmodium vivax. In: Dyann F Wirth, editor. *Malaria: Biology in the Era of Eradication.* New York: Cold Spring Harbor Laboratory Press; 2017. p. 315.

55. Ménard D, Barnadas C, Bouchier C, Henry-halldin C, Gray LR. Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. *PNAS*. 2010;107(13):5967–71.

56. Woldearegai TG, Kremsner PG, Mordmu B. Plasmodium vivax malaria in Duffy-negative individuals from Ethiopia. *Trans R Soc Trop Med Hyg.* 2013;107:328–31.

57. Mendes C, Dias F, Figueiredo J, Mora VG, Cano J, De B, et al. Duffy Negative Antigen Is No Longer a Barrier to Plasmodium vivax – Molecular Evidences from the African West Coast (Angola and Equatorial Guinea). *PLoS Negl Trop Dis.* 2011;5(6):2–7.

58. Gazzinelli RT, Kalantari P, Fitzgerald KA. Innate sensing of malaria parasites. . *Nature* 2014;(October):1–14.

59. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, et al. Geographical variation in Plasmodium vivax relapse. *Malaria J*.2014;13(1):1–16.

60. Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, et al. Relapses of Plasmodium vivax Infection Usually Result from Activation of Heterologous Hypnozoites. *J Infect Dis*. 2007;195:927–33.

61. Koepfli C, Timinao L, Antao T, Barry AE, Siba P, Mueller I, et al. A Large Plasmodium vivax Reservoir and Little Population Structure in the South Pacific. *PLoS One*. 2013;8(6):1–9.

62. Liu Y, Auburn S, Cao J, Trimarsanto H, Zhou H, Gray K, et al. Genetic diversity and population structure of Plasmodium vivax in Central China. *Malaria Jl*. 2014;13(262):1–12.

63. Menegon M, Durand P, Menard D, Legrand E, Picot S, Nour B, et al. Infection , Genetics and Evolution Genetic diversity and population structure of Plasmodium vivax isolates from Sudan , Madagascar , French Guiana and Armenia. *Infect Genet Evol*. 2014;27:244–9.

64. World Health Organization. *Basic MALARIA Microscopy :Part I Learner's Guide.* Second edition. Geneva; 2010. 88 p.

65. Galinski MR, Barnwell JW. Plasmodium vivax : who cares ? *Malaria Jl*. 2008;7(59):1–18.

66. Good MF, Doolan DL. Review Malaria Vaccine Design : Immunological Considerations. *Immunity.* 2010;33(4):555–66.

67. Ménard R, Tavares J, Cockburn I, Markus M, Zavala F. Looking under the skin : the first steps in malarial infection and immunity. *Nature* 2013;11(10):701–12.

68. Niz M De, Burda P, Kaiser G, Portillo HA, Spielmann T, Frischknecht F, et al. Progress in imaging methods: insights gained into Plasmodium biology. *Nature*. 2016;Online Pub:1–18.

69. Sinnis P, Coppi A. A long and winding road : The Plasmodium sporozoite ' s journey in the mammalian host. *Parasitology*. 2007;56:171–8.

70. Singh AP, Buscaglia CA, Wang Q, Levay A, Nussenzweig DR, Walker JR, et al. Plasmodium Circumsporozoite Protein Promotes the Development of the Liver Stages of the Parasite. *Cell*. 2007;131:492–504.

71. Groger M, Fischer HS, Veletzky L, Lalremruata A, Ramharter M. A systematic review of the clinical presentation , treatment and relapse characteristics of human Plasmodium ovale malaria. *Malaria J* 2017;16(112):1–16.

72. Wells TNC, Burrows JN, Baird JK. Targeting the hypnozoite reservoir of Plasmodium vivax : the hidden obstacle to malaria elimination. *Cell*. 2010;26(3):145–51.

73. Oliveira FM, Nogueira PA, Gonc AQ, Fontes CJF, Soares IS, Brito CFA. Plasmodium vivax Duffy binding protein : baseline antibody responses and parasite polymorphisms in a well-consolidated settlement of the Amazon Region. *Trop Med Int Heal*. 2012;17(8):989–1000.

74. Galinski MR, Dluzewski AR, John W. Merozoite invasion of red blood cells: merozoite biogenesis ,erythrocytes. In: Irwin S, editor. *Molecular Approaches to Malaria*. Washington D.C: ASM Press; 2005. p. 113–168.

75. Williams TN. Human red blood cell polymorphisms and malaria. *Elsevier*. 2006;9:1–7.

76. Cowman AF, Healer J, Marapana D, Marsh K. Review Malaria : Biology and Disease. *Cell*.2016;167(3):610–24.

77. Bousema T, Drakeley C. Epidemiology and Infectivity of Plasmodium falciparum and Plasmodium vivax Gametocytes in Relation to Malaria Control and Elimination. *ASM*. 2011;24(2):377–410.

78. Castro-gomes T, Mourão LC, Melo GC, Monteiro WM, Lacerda MVG, Braga M. Potential Immune Mechanisms Associated with Anemia in Plasmodium vivax Malaria : a Puzzling Question. *ASM*. 2014;82(10):3990–4000.

79. Crompton PD, Moebius J, Waisberg M, Garver LS, Miller LH, Barillas C, et al. Mysteries of a deadly infectious disease. *Annu Rev Immunol*. 2014;157–87.

80. Baird JK, Valecha N, Duparc S, White NJ, Price RN. Diagnosis and Treatment of Plasmodium vivax Malaria. *Am J Trop Med Hyg*. 2016;95(Suppl 6):35–51.

81. Noedl H, Wongsrichanalai C, Wernsdorfer WH. Malaria drug-sensitivity testing : new assays , new perspectives. *Trends Parasitol*. 2003;19(4):175–81.

82. Wongsrichanalai C, Barcus MJ, Muth S, Sutamihardja A, Wernsdorfer WH. A Review of Malaria Diagnostic Tools : Microscopy and Rapid Diagnostic Test (RDT). *Am J Trop Med Hyg*. 2007;77(2):119–27.

83. Manguin S, Foumane V, Besnard P, Fortes F, Carnevale P. drugs in Angola : Consequences and effects on human health. *Acta Trop*. 2017;171(March):58–63.

84. Kasetsirikul S, Buranapong J, Srituravanich W, Kaewthamasorn M. The development of malaria diagnostic techniques : a review of the approaches with focus on dielectrophoretic and magnetophoretic methods. *Malaria J.* 2016;15(358):1-15

85. Barber BE, William T, Grigg MJ, Yeo TW, Anstey NM. Limitations of microscopy to differentiate Plasmodium species in a region co-endemic for Plasmodium falciparum , Plasmodium vivax and Plasmodium knowlesi. *Malaria J*. 2013;12(8):1–6.

86. Tadesse FG, Pett H, Baidjoe A, Lanke K, Grignard L, Sutherland C, et al. Submicroscopic carriage of Plasmodium falciparum and Plasmodium vivax in a low endemic area in Ethiopia where no parasitaemia was detected by microscopy or rapid diagnostic test. *Malaria J* 2015;14(303):1–7.

87. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: Spatial targeting of malaria for control and elimination. *PLoS Med.* 2012;9(1):1–7.

88. World Health Organization. *Treatment of Severe Malaria. Guidelines For The Treatment of Malaria*. WHO Press; 2015. 71-88 p.

89. Ferreira MU, Castro MC. Challenges for malaria elimination in Brazil. *Malaria Jl*;.2016;15(284):1–18.

90. Kozycki CT, Umulisa N, Rulisa S, Mwikarago EI, Musabyimana JP, Habimana JP, et al. False - negative malaria rapid diagnostic tests in Rwanda : impact of Plasmodium falciparum isolates lacking hrp2 and declining malaria transmission. *Malaria J.* 2017;16(123):1–11.

91. Kattenberg JH, Ochodo EA, Boer KR, Schallig HDFH, Mens PF, Leeflang MMG. Systematic review and meta-analysis : rapid diagnostic tests versus placental histology , microscopy and PCR for malaria in pregnant women. *Malaria J.*2011;10(1):321.

92. Jimenez A, Channer RRR, Perera R, Gamboa D, Chiodini PL, González IJ, et al. Analytical sensitivity of current best - in - class malaria rapid diagnostic tests. *Malaria J*. 2017;16(128):1–9.

93. Lee J, Jang W, Cho H, Kim Y, Han T, Yun G, et al. False-Positive Results for Rapid Diagnostic Tests for Malaria in Patients with Rheumatoid Factor. *ASM*. 2014;52(10):3784–7.

94. Snounou G, Viriyakosola S, Zhua XP, Jarraa W, Rosariob VE, Thaithongc S, et al. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol Biochem Parasitol*. 1993;61:315–20.

95. Bernotienė R, Valkiūnas G. PCR detection of malaria parasites and related haemosporidians : the sensitive methodology in determining bird - biting insects. *Malaria Jl*.2016;15(283):1–8.

96. Chua KH, Lee PC, Chai HC. Development of insulated isothermal PCR for rapid on - site malaria detection. *Malaria J*. 2016;15(134):1–10.

97. Vallejo Andres , Nora Martinez, Iveth Gonzalez, Myriam Herrera SH. Evaluation of the Loop Mediated Isothermal DNA Amplification (LAMP) Kit for Malaria Diagnosis in P . vivax Endemic Settings of Colombia. *PLoS Negl Trop Dis*. 2015;9(1).

98. Rogers WO, Genton B. A molecular marker of artemisinin- resistant Plasmodium falciparum malaria. *Nature*. 2014;505(January):50–5.

99. Djimde Abdoulaye , Ogobara K , Joseph Cortese KK. A Molecular Marker for Chloroquine Resistant Falciaprum Malaria. *N Engl J Med* . 2001;344(4):257–63.

100. Tahar R, Albergaria C, Zeghidour N, Ngane VF, Basco LK, Roussilhon C. Plasma levels of eight different mediators and their potential as biomarkers of various clinical malaria conditions in African children. *Malaria J*; 2016;15(337):1–15.

101. Sutherland CJ, Babiker H, Mackinnon MJ, Ranford-cartwright L. Rational deployment of antimalarial drugs in Africa : should first- line combination drugs be reserved for paediatric malaria cases ? *Parasitology*. 2013;138(12):1459–68.

102. Michael J Parnham JB. *Treatment and Prevention of Malaria: Antimalarial Drug Chemistry, Action and Use*. Henry Staines SK, editor. Basel: Springer; 2012. 326 p.

103. Talisuna AO, Okello PE, Erhart A, Coosemans M, D'Alessandro U. Intensity of malaria transmission and the spread of plasmodium falciparum-resistant malaria: A review of epidemiologic field evidence. *Am J Trop Med Hyg.* 2007;77(SUPPL. 6):170–80.

104. Sibley CH. Molecular & Biochemical Parasitology Understanding drug resistance in malaria parasites : Basic science for public health. *Mol Biochem Parasitol* . 2014;195(2):107–14.

105. Gresty KJ, Gray K, Bobogare A, Taleo G, Hii J, Wini L, et al. Genetic mutations in pfcrt and pfmdr1 at the time of artemisinin combination therapy introduction in South Pacific islands of Vanuatu and Solomon Islands. *Malaria J*. 2014;13(406):1–9.

106. Tichit M, Bouchier C, Jahevitra M, Me D. Plasmodium vivax Resistance to Chloroquine in Madagascar : Clinical Efficacy and Polymorphisms in pvmdr1 and pvcrt-o Genes . *Antimicrob Agents Chemother*. 2008;52(12):4233–40.

107. Gomes LR, Ketrin N, Oliveira A De, Lavigne AR De, Rezende S, Lima F De, et al. Plasmodium vivax mdr1 genotypes in isolates from successfully cured patients living in endemic and non - endemic Brazilian areas. *Malaria J* 2016;15(96):1–5.

108. Getachew S, Thriemer K, Auburn S, Abera A, Gadisa E, Aseffa A, et al. Chloroquine efficacy for Plasmodium vivax malaria treatment in southern Ethiopia. *Malaria Journa*l.2015;14(525):1–8.

109. Getachew S, Thriemer K, Auburn S, Abera A, Gadisa E, Aseffa A, et al. Chloroquine efficacy for Plasmodium vivax malaria treatment in southern Ethiopia. *Malaria J*. 2015;14(525):1–8.

110. John GK, Douglas NM, Seidlein L Von, Nosten F, Baird JK, White NJ, et al. Primaquine radical cure of Plasmodium vivax : a critical review of the literature. *Malaria J* 2012;11(280):1–11.

111. Thriemer K, Ley B, Bobogare A, Dysoley L, Alam MS, Pasaribu AP, et al. Challenges for achieving safe and effective radical cure of Plasmodium vivax : a round table discussion of the APMEN Vivax Working Group. *Malaria J*.2017;16(141):1–9.

112. Dewasurendra RL, Rockett KA, Fernando SD, Carter R, Kwiatkowski DP. G6PD gene variants and its association with malaria in a Sri Lankan population. *Malaria J.* 2015;14(93):1–15.

113. Verret C, Haus-cheymol R, Lafille J. Malaria Relapses after Primaquine *Prophylaxis. Emerg Infect Dis.* 2006;12(11):11–2.

114. Adams JH, Mueller I. The Biology of Plasmodium vivax. In: Dyann F Wirth, editor. *Malaria: Biology in the Era of Eradication.* New York: Cold Spring Harbor Laboratory Press; 2017. p. 315.

115. Kibret S, Lautze J, Mccartney M, Wilson GG, Nhamo L. Malaria impact of large dams in sub - Saharan Africa : maps , estimates and predictions. *Malaria Jl*. 2015;14(339):1–12.

116. Abdulla S, Adjei S, Agbenyega T, Agnandji ST, Aide P, Anderson S, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med*. 2015;373(21):2025–37.

117. Tetteh KKA, Stewart LB, Ochola LI, Amambua-ngwa A, Thomas AW, Marsh K, et al. Prospective Identification of Malaria Parasite Genes under Balancing Selection. *PLoS One.* 2009;4(5):1–9.

118. Sauerwein RW, Richie TL. Malaria vaccines getting close to clinical reality. *Vaccine* ; 2015;33(52):7423–4.

119. Schwartz L, Brown G V, Genton B, Moorthy VS. A review of malaria vaccine clinical projects based on the WHO rainbow table. *Malaria J* 2012;11(11):1–22.

120. Federal Ministry of Health.*Ethiopia National Malaria Indicator Survey Technical Summary*. Addis Ababa,Ethiopia. 2012;(September).

121. Recht J, Siqueira AM, Monteiro WM, Herrera SM, Herrera S, Lacerda MVG. Malaria in Brazil , Colombia , Peru and Venezuela : current challenges in malaria control and elimination. *Malaria J* 2017;16(273):1–18.

122. Federal Ministry of Health. *National Malaria Program Monitoring and Evaluation Plan*. Addis Ababa,Ethiopia; 2014.

123. Golassa L, Baliraine FN, Enweji N, Erko B, Swedberg G, Aseffa A. Microscopic and molecular evidence of the presence of asymptomatic Plasmodium falciparum and Plasmodium vivax infections in an area with low , seasonal and unstable malaria transmission in Ethiopia. *BMC Infect Dis* 2015;15(310):1–10.

124. Scott CA, Yeshiwondim AK, Serda B, Guinovart C, Tesfay BH, Agmas A, et al. Mass testing and treatment for malaria in low transmission areas in Amhara Region , Ethiopia. *Malaria J*.2016;15(305):1–13.

125. Ketema T, Getahun K, Bacha K. Therapeutic efficacy of chloroquine for treatment of Plasmodium vivax malaria cases in Halaba. *Parasite& Vectors*. 2011;4(46):1–7.

126. Assefa M, Eshetu T, Biruksew A. Therapeutic efficacy of chloroquine for the treatment of Plasmodium vivax malaria among outpatients at Hossana Health Care. *Malaria J*. 2015;14(458):1–8.

127. Abreha T, Hwang J, Thriemer K, Tadesse Y, Girma S, Melaku Z, et al. Comparison of artemether-lumefantrine and chloroquine with and without primaquine for the treatment of Plasmodium vivax infection in Ethiopia : A randomized controlled trial. *PLoS Med.* 2017;14(5):1–17.

128. Yohannes AM, Teklehaimanot A, Bergqvist Y, Ringwald P. Confirmed Vivax Resistance to Chloroquine and Effectiveness of Artemether-Lumefantrine for the Treatment of Vivax Malaria in Ethiopia. *Am J Trop Med Hyg*. 2011;84(1):137–40.

129. Macleod DA, Jones A, Di F, Chen Y, Morton DC, Andela N, et al. Temperature suitability for malaria climbing the Ethiopian Highlands. *Environ Res Lett.* 2017;12:1–8.

130. Abose T, Gerrit J, Borsboom G, Vlas SJ De, Habbema JDF. Spatial and temporal v ariations of malaria epidemic risk in Ethiopia : factors in volved and implications. *Acta Trop*. 2003;87:331–40.

131. Siraj A.S , Vega Santos .M, M.J Bouma, D.Yadeta, D.Ruiz Carrascal MP. Altitudinal Changes in Malaria. *Science*. 2014;343(March):1154–1159.

132. Ayele DG, Zewotir TT, Mwambi HG. Prevalence and risk factors of malaria in Ethiopia. *Malaria Jl*.2012;11(195):1–14.

133. Zhou G, Minakawa N, Githeko AK, Yan G. Association between climate variability and malaria epidemics in the East African highlands. *PNAS.* 2004;101(8):2375–2380.

134. Georges S, Viriyakbosola S, Ping X, Jarra W. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol Biochem Parasitol.* 1993;61:315–420.

135. Abduselam N, Zeynudin A, Berens-riha N, Seyoum D, Pritsch M, Tibebu H, et al. Similar trends of susceptibility in Anopheles arabiensis and Anopheles pharoensis to *Plasmodium vivax* infection in Ethiopia. *Parasites & Vectors*.2016;9(552):1–9.

136. Ribeiro JMC, Seulu F, Abose T, Kidane G, Teklehaimanot A. Temporal and spatial distribution of anopheline mosquitos in an Ethiopian village : implications for malaria control strategies. *Bull World Health Organ.* 1996;74(3):299–305.

137. Hay SI, Cox J, Rogers DJ, Randolph SE, Stern DI, Shanks GD, et al. Climate change and the resurgence of malaria in the East African highlands. *Nature.* 2002;415:905–9.

138. Graves PM, Richards FO, Ngondi J, Emerson PM, Biru E, Endeshaw T, et al. Individual , household and environmental risk factors for malaria infection in Amhara , Oromia and SNNP regions of Ethiopia. *Trans R Soc Trop Med Hyg*. 2009;1–10.

139. Jima D, Getachew A, Bilak H, Steketee RW, Emerson PM, Graves PM. Malaria indicator survey 2007 , Ethiopia : coverage and use of major malaria prevention and control interventions. *Malaria J*. 2010;9(58):1–12.

140. Tassew A, Hopkins R, Deressa W. Factors influencing the ownership and utilization of long - lasting insecticidal nets for malaria prevention in Ethiopia. *Malaria J*. 2017;16(262):1–9.

141. 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(2-3):87–97.

142. Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, Malenke JA, et al. African origin of the malaria parasite *Plasmodium vivax*. *Nat Commun*. 2014;5(3346):1–4.

143. Auburn S, Barry AE. Dissecting malaria biology and epidemiology using population genetics and genomics. *Int J Parasitol*. 2017;47(2-3):77–85.

144. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature*. 2008;455(October).

145. Cornejo OE, Fisher D, Escalante AA. Genome-Wide Patterns of Genetic Polymorphism and Signatures of Selection in Plasmodium vivax. *Genome Biol Evol*. 2014;7(1):106–19.

146. Masatoshi N. Molecular population genetics and evolution. Neuberger A, editor. Amsterdam: North-Holland Publishing Company; 1975. 290 p.

147. Escalante AA, Cornejo OE, Rojas A, Udhayakumar V, Lal AA. Assessing the effect of natural selection in malaria parasites. *Trends Parasitol*. 2004;20(8):1–8.

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

149. Botsford LW, Jain SK. Population biology and its application to practical problems. In: Botsford LW, editor. *Applied Population Biology*. Amsterdam: Kluwer Academic Publishers; 1992. p. 1–24.

150. Conway DJ. Measuring immune selection. *Parasitology*. 2002;125:1–7.

151. Ord R, Polley S, Tami A, Sutherland CJ. High sequence diversity and evidence of balancing selection in the Pvmsp 3 ␣ gene of Plasmodium vivax in the Venezuelan Amazon. *Mol Biochem Parasitol*. 2005;144:86–93.

152. Shen HM, Chen SB, Wang Y, Xu B, Abe EM, Chen JH. Genome - wide scans for the identification of Plasmodium vivax genes under positive selection. *Malaria J*2017;16(238):1–12.

153. Weedall GD, Conway DJ. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Cell.* 2010;26(7):363–9.

154. Barry AE, Waltmann A, Koepfli C, Barnadas C, Mueller I. Uncovering the transmission dynamics of Plasmodium vivax using population genetics. *Pathog Glob Health.* 2015;1–11.

155. Rayner JC, Huber CS, Feldman D, Ingravallo P, Galinski MR, Barnwell JW. Plasmodium vivax merozoite surface protein PvMSP-3β is radically polymorphic through mutation and large insertions and deletions . *Infect ,Genetics Evol*. 2004;4:309–19.

156. Li Y, Wang G, Meng F, Zeng W, He C, Hu X, et al. Genetic diversity of Plasmodium vivax population before elimination of malaria in Hainan Province ,. *Malaria J*. 2015;14(78):1–12.

157. Wangchuk S, Drukpa T, Penjor K, Peldon T, Dorjey Y, Dorji K. Where chloroquine still works : the genetic make - up and susceptibility of Plasmodium vivax to chloroquine plus primaquine in Bhutan. *Malaria J*.2016;15(277):1–11.

158. Maneerattanasak S, Gosi P, Krudsood S, Lanteri CA, Snounou G, Khusmith S, et al. Genetic diversity among Plasmodium vivax isolates along the Thai – Myanmar border of Thailand To cite this version . *Malaria Jl.* 2016; 15(75):1-15

159. Mahgoub H, Gasim GI, Musa IR, Adam I. Severe Plasmodium vivax malaria among sudanese children at New Halfa Hospital , Eastern Sudan. *Parasite& Vectors.* 2012;5(154):2–5.

160. Havryliuk T, Ferreira MU. A closer look at multiple-clone Plasmodium vivax infections : detection methods , prevalence and consequences. *Mem Inst Oswaldo Cruz*. 2009;104(1):67–73.

161. Babiker H a., Gadalla A a H, Ranford-Cartwright LC. The role of asymptomatic P. falciparum parasitaemia in the evolution of antimalarial drug resistance in areas of seasonal transmission. *Drug Resist Update.*2013;16(1-2):1–9.

162. Talisuna AO. Mitigating the threat of artemisinin resistance in Africa : improvement of drug-resistance surveillance and response systems. *Lancet Infect Dis*. 2013;12(11):888–96.

163. Imwong M, Pukrittakayamee S, Looareesuwan S, Pasvol G, Poirriez J, White NJ, et al. Association of Genetic Mutations in Plasmodium vivax dhfr with Resistance to Sulfadoxine-Pyrimethamine : Geographical and Clinical Correlates. *Antimicrob Agents Chemother.* 2001;45(11):3122–7.

164. Nash D, Nair S, Mayxay M, Newton PN, Guthmann J, Foundation S, et al. Selection strength and hitchhiking around two anti-malarial resistance genes. *Proc Natl Acad Sci.* 2005;272(June):1153–61.

165. Cheng Q, Cunningham J, Gatton ML. Systematic Review of Sub-microscopic P . vivax Infections : Prevalence and Determining Factors. *PLoS Negl Trop Dis*. 2015;9(1):1–12.

166. Massebo F, Balkew M, Gebre-michael T, Lindtjørn B. Entomologic Inoculation Rates of Anopheles arabiensis in Southwestern Ethiopia. *Am J Trop Med Hyg*. 2013;89:466–73.

167. Mohammed A, Ndaro A, Kalinga A, Manjurano A, Mosha JF, Mosha DF, et al. Trends in chloroquine resistance marker, Pfcrt-K76T mutation ten years after chloroquine withdrawal in Tanzania. *Malaria J* 2013;12(1):415.

168. 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. *Malaria J* 2014;13(244):1–9.

169. Kaslow DC, Biernaux S. RTS , S : Toward a first landmark on the Malaria Vaccine Technology Roadmap ☪. *Vaccine*. 2015;33(52):7425–32.

170. Beeson JG, Crabb BS. Towards a Vaccine against Plasmodium vivax Malaria. *PLoS Med*. 2007;4(12):1862–4.

171. Longley RJ, Sattabongkot J, Mueller IVO. Insights into the naturally acquired immune response to Plasmodium vivax malaria. *Parasitology.* 2016;143:154–70.

172. Patrick E Duffy , Tejram Sahu , Adovi Akue, Neta Milman CA. Pre-erythrocytic malaria vaccines : identifying the targets. *Expert Rev Vaccines.* 2013;11(10):1261–80.

173. Reyes-sandoval A. Plasmodium vivax malaria vaccines. *Hum Vaccines Immunother*. 2013;9(12):2558–65.

174. Richards JS, Beeson JG. The future for blood-stage vaccines against malaria *Nature* . 2009;87(5):377–90.

175. Stanisic DI, Good MF. Whole organism blood stage vaccines against malaria. *Vaccine* .2015;33(52):7469–75.

176. Hoffman SL, Vekemans J, Richie TL, Duffy PE. The march toward malaria vaccines. *Vaccine* . 2015;33:D13–23.

177. Sepúlveda N, Morais CG, Mourão LC, Freire MF, Fontes CJF, Lacerda MVG, et al. Allele - specific antibodies to Plasmodium vivax merozoite surface protein - 1 : prevalence and inverse relationship to haemoglobin levels during infection. *Malaria J*. 2016;15(559):1–11.

178. Sauerwein RW, Bousema T. Transmission blocking malaria vaccines : Assays and candidates in clinical development ☪. *Vaccine* .2015;33(52):7476–82.

179. Vallejo AF, Martinez NL, Tobon A, Alger J, Lacerda M V, Kajava A V. Global genetic diversity of the Plasmodium vivax transmission - blocking vaccine candidate Pvs48 / 45. *Malaria J.* 2016;15(202):1–9.

180. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing Patterns of Selection and Geospatial Genetic Diversity within Two Leading Plasmodium vivax Candidate Vaccine Antigens. *PLoS Negl Trop Dis.* 2014;8(4):1–17.

181. Tham W, Beeson JG, Rayner JC. Plasmodium vivax vaccine research – we ' ve only just begun. *Int J Parasitol*. 2017;47(2-3):111–8.

182. World Health Organization.*World Malaria Report 2012*. WHO Press; 2012. 182p.

183. Takala SL, Plowe C V. Genetic diversity and malaria vaccine design , testing and efficacy : preventing and overcoming “ vaccine resistant malaria .” *Parasite Immunol*. 2009;31(9):560–73.

184. Takala, Shannon L and Christopher V P. Genetic diversity and maalria vaccine design, testing, and efficacy : Preventing and overcoming “ vaccine resistant malaria .” *Parasite Immunol*. 2010;31(9):560–73.

185. Smith T, Beck H, Irion A, Betuela I, Alpers MP, Anders R, et al. Strain-Specific Humoral Response to a Polymorphic Malaria Vaccine. *Infection and Immunity*. 2004;72(11):6300–5.

186. Polley SD, Tetteh KKA, Lloyd JM, Akpogheneta OJ, Greenwood BM, Bojang KA, et al. Plasmodium falciparum Merozoite Surface Protein 3 Is a Target of Allele-Specific Immunity and Alleles Are Maintained by Natural Selection. *J Infect Dis*. 2007;195:279–87.

187. Barry AE, Arnott A. Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Front Immunol*. 2014;5(July):1–16.

188. Dutta S, Lee SY, Batchelor AH, Lanar DE. Structural basis of antigenic escape of a malaria vaccine candidate. *PNAS.* 2007;104(30):1–6.

189. Jiang J, Barnwell JW, Meyer EVS, Galinski MR. Plasmodium vivax Merozoite Surface Protein-3 (PvMSP3): Expression of an 11 Member Multigene Family in Blood-Stage Parasites. *PLoS One*. 2013;8(5):1–14.

190. Galinski MR, Corredor-medina C, Povoa M. Plasmodium vivax merozoite surface protein-3 contains coiled-coil motifs in an alanine-rich central domain. *Mol Biochem Parasitol*. 1999;101:131–47.

191. Beeson JG, Drew DR, Boyle MJ, Feng G, Fowkes FJI, Richards JS. Merozoite surface proteins in red blood cell invasion , immunity and vaccines against malaria. *FEMS Microbiol Rviews*. 2016;40(3):343–72.

192. Chen H, Yan G, Sattabongkot J. Genetic diversity and multiple infections of Plasmodium vivax malaria in western Thailand.
Am J Trop Med Hyg. 2003;68(5):613–9.

193. Arnott A, Wapling J, Mueller I, Ramsland PA, Siba PM, Reeder JC, et al. Distinct patterns of diversity , population structure and evolution in the AMA1 genes of sympatric Plasmodium falciparum and Plasmodium vivax populations of Papua New Guinea from an area of similarly high transmission.
Malaria J. 2014;13(233):1–16.

194. Soares LA, Evangelista J, Orlandi PP, Almeida ME, Sousa LP De, Chaves Y, et al. Genetic Diversity of MSP1 Block 2 of Plasmodium vivax Isolates from Manaus (Central Brazilian Amazon). *J Immuology Res*. 2014;2014:14–6.

195. Poirier P, Lang CD, Atchade PS, Lemoine JP, Louise M, De C, et al. The hide and seek of Plasmodium vivax in West Africa : report from a large - scale study in Beninese asymptomatic subjects. *Malaria J*.2016;15(570):1–9.

196. Zhou G, Yan G. Genetic diversity and multiple infections of Plasmodium vivax malaria in Western Thailand. *Am J Trop Med Hyg*. 2003;68(5):613–9.

197. Prajapati SK, Joshi H, Valecha N. Plasmodium vivax merozoite surface protein- 3 α : A high-resolution marker for genetic diversity studies marker for genetic diversity studies. *J Vector Borne Dis*. 2014;47(June):85–90.

198. Rice BL, Acosta MM, Pacheco MA, Escalante AA. Merozoite surface protein-3 alpha as a genetic marker for epidemiologic studies in Plasmodium vivax : a cautionary note. *Malaria J*. 2013;12(288):1–13.

199. Protein- MS, Bitencourt AR, Vicentin EC, Jimenez MC, Ricci R, Leite JA, et al. Antigenicity and Immunogenicity of Plasmodium vivax Merozoite Surface Protein-3. *PLoS One*. 2013;8(2):1–10.

200. Lima-junior JC, Jiang J, Rodrigues-da-silva RN, Banic DM, Tran TM, Ribeiro RY, et al. B cell epitope mapping and characterization of naturally acquired antibodies to the Plasmodium vivax Merozoite Surface Protein-3α (PvMSP-3 α) in malaria exposed individuals from Brazilian Amazon. *Vaccine*. 2011;29(9):1801–11.

201. Tetteh KKA, Stewart LB, Ochola LI, Amambua-ngwa A, Thomas AW, Marsh K, et al. Prospective Identification of Malaria Parasite Genes under Balancing Selection. *PLoS One*. 2009;4(5):1–9.

202. Ochola LI, Tetteh KKA, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele Frequency – Based and Polymorphism-Versus- Divergence Indices of Balancing Selection in a New Filtered Set of Polymorphic Genes in Plasmodium falciparum Research article. *Mol Biol Evol*. 2010;27(10):2344–51.

203. Central Statistic Agency. *Ethiopia Demographic and Health Survey*. Addis Ababa, Ethiopia.2012.

204. Baidjoe a, Stone W, Ploemen I, Shagari S, Grignard L, Osoti V, et al. Combined DNA extraction and antibody elution from filter papers for the assessment of malaria transmission intensity in epidemiological studies. *Malaria J*.2013;12(1475-2875)

205. Singh B, Bobogare A, Cox-singh J, Snounou G, Abdullah MS, Rahman HA. A Genus-and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. *Am J Trop Med Hyg*. 1999;60(4):687–92.

206. Snounou G, Snounou G, Beck H. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment . *Parasitol Today*. 1998;14(11):462–9.

207. Verma A, Joshi H, Singh V, Anvikar A, Valecha N. polymorphisms : analysis in the Indian subcontinent. *Malaria J* 2016;15(492):1–13.

208. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673–80.

209. Librado P, Rozas J. DnaSP v5 : a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25(11):1451–2.

210. Wright S. The Genetical Structure of Populations. *Ann Eugen*. 1954;323–54.

211. Nei M, Gojoborit T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions ’. *Mol Biol Evol*. 1986;3(5):418–26.

212. Li W. Statistical Tests of Neutrality of Mutations. *Genetics*. 1993;133:693–709.

213. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, et al. A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. *Science*. 1999; 286(5443):1351–1353.

214. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*. 1985;111:147–64.

215. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3 : a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010;26(19):2462–3.

216. Smith JM. Analyzing the Mosaic Structure of Genes. *J Mol Evol*. 1992;34:126–9.

217. Padidam M, Sawyer S, Fauquet CM. Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*. 1999;265:218–25.

218. Ã SLKP, Posada D, Gravenor MB, Woelk CH, Frost SDW. BIOINFORMATICS APPLICATIONS NOTE Sequence analysis GARD : a genetic algorithm for recombination detection. *Oxford J*. 2006;22(24):3096–8.

219. Auton A, Mcvean G. Recombination rate estimation in the presence of hotspots. *Cold Spring Harb Perspect Med*. 2007;(17):1219–27.

220. Forster P, Ro A. Median-Joining Networks for Inferring Intraspecific Phylogenies. *Mol Biol Evol*. 1994;16(1):37–48.

221. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6 : Molecular Evolutionary Genetics Analysis Version 6 . *Oxford J*. 2013; 30(12):2725–9.

222. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Oxford J*. 1992;8(3):275–82.

223. Federal Ministry of Health. *Diagnosis and Treatment Guidelines for Health Workers in Ethiopia* .2nd Edition.. Federal Democratic Republic of Ethiopia Ministry of health; 2004.

224. World Health Organization. *World Malaria Report 2015 Summary*.WHO press 2015;32p

225. Getachew S, To S, Trimarsanto H, Thriemer K. Variation in Complexity of Infection and Transmission Stability between Neighbouring Populations of Plasmodium vivax in Southern Ethiopia. *PLoS One*. 2015;10(10).

226. Golassa L, White MT. Population - level estimates of the proportion of Plasmodium vivax blood - stage infections attributable to relapses among febrile patients attending Adama Malaria Diagnostic. *Malaria J*2017;16(301):1–11.

227. Mueller I, Galinski MR, Tsuboi T, Collins WE, King CL. Natural Acquisition of Immunity to Plasmodium vivax : Epidemiological Observations and Potential Targets. In: *Advances in Parasitology* .2013. p. 77–131.

228. Baird JK. Evidence and Implications of Mortality Associated with Acute Plasmodium vivax Malaria. *ASM*. 2013;26(1):36–57.

229. Golassa L, Enweji N, Erko B, Aseffa A, Swedberg G. Detection of a substantial number of sub- microscopic Plasmodium falciparum infections by polymerase chain reaction : a potential threat to malaria control and diagnosis in Ethiopia. *Malaria J*. 2013;12(352):1–10.

230. Schousboe ML, Rajakaruna RS, Amerasinghe PH, Konradsen F, Ord R, Pearce R, et al. Analysis of Polymorphisms in the Merozoite Surface Protein -3 α Gene and Two Microsatellite Loci in Sri Lankan Plasmodium vivax : Evidence of Population Substructure in Sri Lanka. *Am J Trop Med Hyg*. 2011;85(6):994–1001.

231. Ribeiro RS, Ladeira L, Rezende AM, Jesus C, Fontes F, Carvalho LH, et al. Analysis of the genetic variability of PvMSP- 3 α among Plasmodium vivax in Brazilian field isolates. *Mem Inst Oswaldo Cruz*. 2011;106:27–33.

232. Lo C, Forero-rodrı J, Patarroyo MA. Genetic Diversity and Selection in Three Plasmodium vivax Merozoite Surface Protein 7 (Pvmsp-7) Genes in a Colombian Population.*PloS One*. 2012;7(9):1–10.

233. Wickramarachchi T, Premaratne PH, Dias S, Handunnetti SM. Genetic complexity of Plasmodium vivax infections in Sri Lanka , as reflected at the merozoite-surface- protein-3 α locus .*Ann Trop Medicine Parasito*l. 2013;104(2):95–108.

234. M.C. Bruce , M.R. Galinski , J.W.Barnwell, C.A.Donnelly, M.Walmsley , M.P.Alpers, D.Walliker KPD. Genetic diversity and dynamics of Plasmodium falciparum and P . vivax populations in multiply infected children with asymptomatic malaria infections in Papua New Guinea. *Parasitology*. 1999;121(2000):257–272.

235. Delgado-ratto C, Gamboa D, Soto-calle VE, Den P Van, Torres E, Sánchez-martínez L, et al. Population Genetics of Plasmodium vivax in the Peruvian Amazon. *PLoS Negl Trop Dis*. 2016;10(1):1–23.

236. Roode JC De, Pansini R, Cheesman SJ, Helinski MEH, Huijben S, Wargo AR, et al. Virulence and competitive ability in genetically malaria diverse infections. *PNAS*. 2015;102(21):7624–8.

237. Edgar RC. MUSCLE : a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5(113):1–19.

238. Alene GD, Bennett S. Chloroquine resistance of Plasmodium fakiparum malaria in Ethiopia and Eritrea. *Trop Med Int Heal*. 1996;I(6):810–5.

239. Mula P, Fernández-Martínez A, De Lucio A, Ramos JM, Reyes F, González V, et al. Detection of high levels of mutations involved in anti-malarial drug resistance in Plasmodium falciparum and Plasmodium vivax at a rural hospital in southern Ethiopia. *Malaria J*. 2011;10(214):1–7.

240. Golassa L, Erko B, Baliraine FN, Aseffa A, Swedberg G. Polymorphisms in chloroquine resistance- associated genes in Plasmodium vivax in Polymorphisms in chloroquine resistance- associated genes in Plasmodium vivax in Ethiopia. *Malaria J*. 2015.14(164):1–5.

241. Chenet SM, Tapia LL, Escalante AA, Durand S, Lucas C, Bacon DJ. Genetic diversity and population structure of genes encoding vaccine candidate antigens of Plasmodium vivax. *Malaria J*. 2012;11(68).

243. Krishnarjuna B, Andrew D, Macraild CA, Morales RA V, Beeson JG, Anders RF, et al. Strain-transcending immune response generated by chimeras of the malaria vaccine candidate merozoite surface protein 2. *Nature* 2016;(August 2015):1–12.

244. Arnott A, Mueller I, Ramsland P a., Siba PM, Reeder JC, Barry AE. Global Population Structure of the Genes Encoding the Malaria Vaccine Candidate, Plasmodium vivax Apical Membrane Antigen 1 (PvAMA1). *PLoS Negl Trop Dis*. 2013;7(10).

10. Appendices

Appendix A. Standard operating procedure for Chelex Saponin extraction

Standard Operating Procedure	SOP No: HIRE_1/1 Version: **1** Effective Date: 19/02/16
Title: Combined High Throughput Saponin/ Chelex Extraction of DNA and Antibodies from Blood Spotted What No. 3 Filterpaper	

1.0 PURPOSE / INTRODUCTION: This SOP describes the detailed process of extracting DNA and Antibodies from blood spotted Whatman No.3 filter using Saponin for lyses of cell material and elution of Antibodies and Chelex for binding and removal of oxidative elements. These samples were previously collected on filter paper and stored in the freezer at least at -20°C. The protocol is optimized to extract material from 6mm diameter punch, yielding higher amounts of DNA and Antibodies which then can be used in downstream PCR and ELISA based applications for the detection of both human and parasite DNA and parasite directed Antibodies. This protocol is fast but more prone to contamination. Therefore we recommend using the same consumables as we did.

2.0 SHORT SUMMARY OF PROCEDURE

6mm diameter dried blood spot of blood is cut with puncher and taken and added into a 2mL Eppendorf tube incubated in 200µL of a 0.5% Saponin solution at RT overnight on a shaker. Cells will lyse and the lysate will be released into the supernatant. DNA will remain attached to a now near to clear white filterpaper disc. The saponin solution will contain serum components including antibodies. 200uL of the saponin is then stored in a new 96 deep well plate for ELISA. After a single washing step with PBS the disc(s) will be incubated in a 6% Chelex solution and finally heated at 95°C for up to 30 minutes, while intermediately being spanned down. Chelex will bind positively charged oxidative elements leaving DNA in solution after denaturation at 95°C. After incubation, a final high speed centrifuge step follows after which the supernatant (now containing the eluted DNA) is stored in a new 96 deep well plate with preferably no Chelex contamination.

3.0 Study Title: Genetic Diversity and Natural Selection of a Malaria Vaccine Candidate in the Ethiopian Population; *Plasmodium vivax* Merozoite Surface Protein 3α

Study PI : Alebachew Messele

SOP prepared By: Fitsum GirmaTadesse (Msc, PhD Fellow), Institute of Biotechnology and Armauer Hansen Research Institute (fitsum.girma@aau.edu.et)

4.0 SCOPE / RESPONSIBILITY:

This SOP applies to laboratory personnel involved in the analyses of the samples

5.0 EQUIPMENT & CONSUMABLES

- 2 x Tweezers (per person)
- 2x Scissors (per person)
- Bunsen Burner
- 70% Ethanol
- Small bucket to hold Ethanol (with lid)
- Dry heat block/waterbath/incubator at 37◦C
- Dry heat block or waterbath at 95◦C
- Racks for 10mL & 50mL tubes
- 0-200 multichannel pipette
- 200-1mL multichannel pipette
- 200uL and 1mL pipette and pipette tips
- 2mL Eppendorf tubes
- 96 well plates (wells of 2mL+ volume) and according lids (axygen, see supply file)
- 96 well plates 1.2mL + lids
- Axygen sealing lids
- 1L/500mL Bottles
- Forceps
- Gloves
- Thermometer
- Vacuum system for aspiration
- Benchtop centrifuge

6.0 REAGENTS

a) 100% ethanol
b) 0,5% high grade saponin in PBS prepared fresh or stored at 4°C or -20°C (Sigma-Aldrich, Product Number S4521), best to have Saponin solution is at RT before use. (100mL per plate, for 1L mix add 5 grams of Saponin)
c) 1X Phosphate Buffered Saline (PBS) Calcium and Magnesium free pH 7.4, stored at 4°C (100mL per plate)
d) 6% Chelex-100 (Fisher-Scientific Catalog Number NC9708062) in DNA'ase/RNA'ase free water stored at room temperature (15mL per plate)
e) Distilled water DNA/RNA'ase free
f) PBS in DNA/RNA'ase free water

METHODOLOGY

Cutting

1. Select a reasonable number of samples from the -20 fridge. We want to limit the amount of time filter papers have at room temperature.
2. Clean down lab bench with 70% ethanol, and lay out paper (M-tork) and foil for disc cutting and tool sterilizing as is appropriate.

6.0 Label extraction tubes appropriately.

3. Use the 6mm diameter puncher to cut a single blood disc from blood covered areas of any blood spot. Add the discs to the extraction tubes with forceps.
4. Sterilize the puncher and forceps by dipping briefly in ethanol, and flaming using the Bunsen burner. Ensure no alcohol remains on the tool, but also that the tool is not too hot when cutting occurs. By ensuring prompt flaming, overheating can be avoided.
5. Continue with the next sample

Extraction

1. With a 200uL (1000uL) pipette, add 820µL (this will elute your serum to 1/200) of a 0.5% Saponin solution to every well. Normal elution dilutions (table below) achieve 1/100 blood dilutions (1/200 serum dilutions).
   The table for filter paper disc punch diameter and blood volume is as follows:

Diameter of punch (mm)	**Blood (µl)**	**Reconstitution volume (µl)**
1.0	0.2	20
1.5	0.5	50
2.0	0.9	90
2.5	1.4	140
3.0	2.1	200
3.5	2.8	280
4.0	3.7	360
4.5	4.7	460
5.0	5.8	570
5.5	7.0	690
6.0	**8.3**	**820**

2. Make sure that all the tubes are well tightly closed
3. Put the tubes (in a rack) on a shaker, use an intermediate shaking speed, make sure all the discs are freely moving in the tubes and that the rack is properly stabilized. **Leave it shaking overnight**.

4. Remove tubes from shaker and photograph, if not difficult, to making sure that filter paper elution success (colour) and sample ID are visible. Sharply shake tubes to remove discs stuck to the lid(Figure 21).
5. Centrifuge tubes **for 1 minute at high speed**. **Confirm that your filterpaper discs are white.** If not this protocol might not yield the best quality of DNA. This is usually a result of bad storage conditions. Commercial kits might offer a solution.
6. Open the tubes carefully to avoid serum transfer from one to another.
7. Transfer 200uL of the now reddish saponin to a new plate (0.5mL deep well plate) coded: **SERUM-[CODE]**. This plate can either be transferred to the ELISA room for direct processing or can be stored in a freezer at -20/-80◦C for subsequent use.
8. Aspirate the rest of the saponin from the wells using a vacuum system. Try to avoid touching the filter paper discs. Preferably change tip after every well, alternatively rinse tip in PBS, if filter paper has been 'touched', by default replace tip.
9. Add 1 mL of cooled (+4°c) PBS to each tube (no saponin), tightly close the lids and shake for 30 minutes, again assuring that punches move freely in the tubes. Incubate at 4°C for 30 minutes.
10. After sitting for 30 mins, shake sharply to get discs to well bottoms. Remove all PBS using a new tip for each tube. Discard PBS in contaminated waste. Discs MUST now be at the well bottoms. Use tips to move them if they are not.
11. Set the water bath to 96^{o}C, ensuring that the 'overheat temperature' is appropriately high (97^{o}C). Check the temperature with the thermometer.
12. Transfer 150uL of 6% Chelex solution into each well using a multichannel pipette. Make sure that the Chelex is properly distributed after every transfer round. Chelex settles quickly, this step ensures that not just water is being transferred.
13. Extract the parasite DNA by incubating plates for 4x7 minutes at 95-96^{o}C. Start the timer once 95^{o}C is reached. Stabilize tube racks with floats, and check once in each cycle that tubes have not overturned. When removing after each incubation, move to an ice tray and shake the tubes slightly to move the discs around in the chelex solution (without removing discs from well bottoms). Some evaporation appears to be inevitable, but by ensuring lids are kept closed, this can be kept to a minimum. To release pressure build up under the lids, either punch holes in the lid, or use tight sealing aluminum foil.
14. After incubation, centrifuge the tubes at maximum speed for 5 minutes. Prepare a new plate for storage of the DNA. Label this plate with a barcode **DNA-[CODE]**
15. Transfer 80uL of the eluted DNA solution from the spun tubes. Make sure that the Chelex pellet is not disturbed.
16. Spin plates for 10 minutes at high speed and store at -20 or lower degrees Celsius until downstream use.

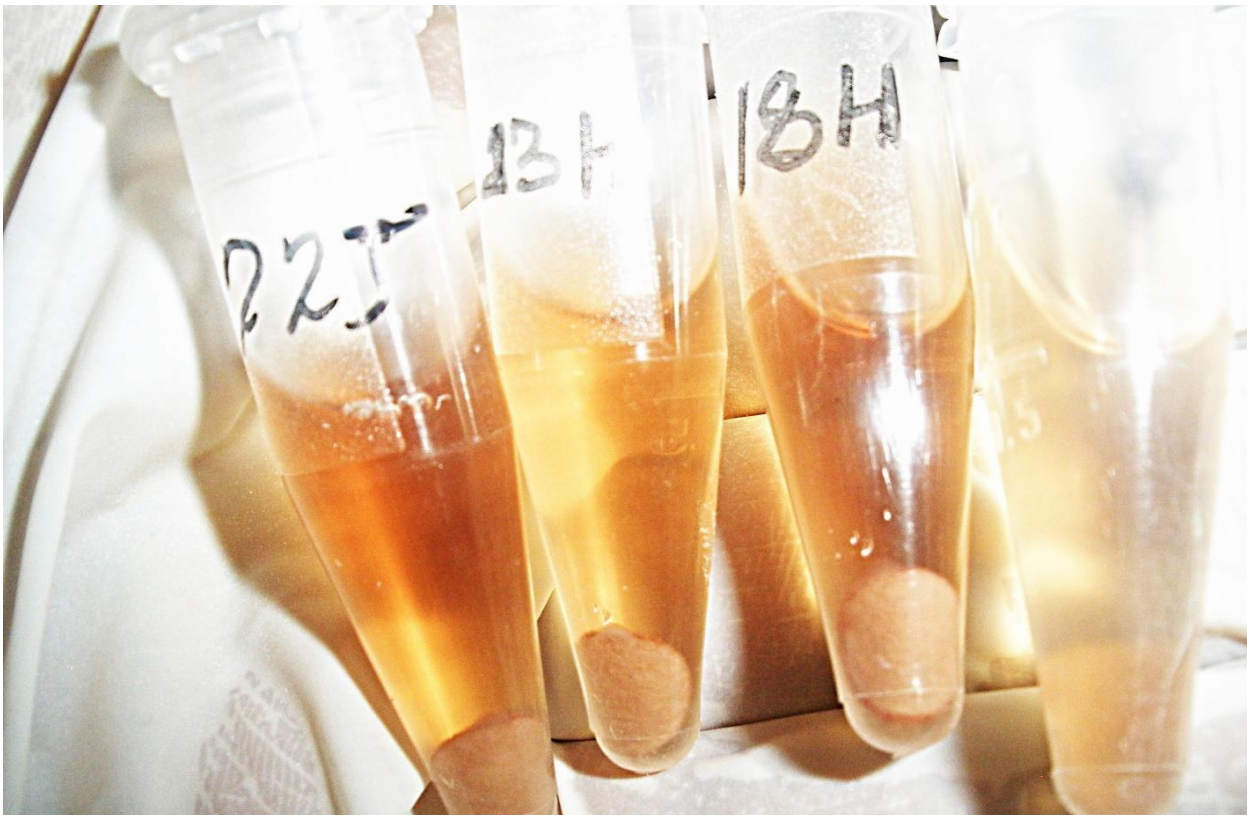
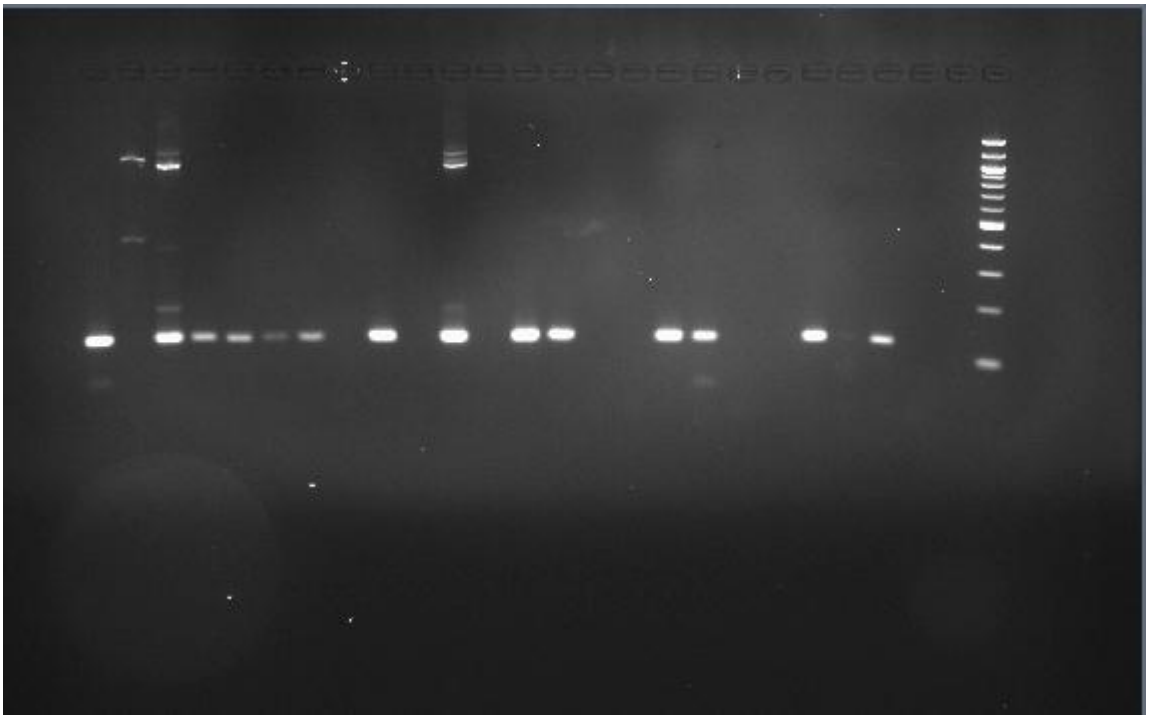
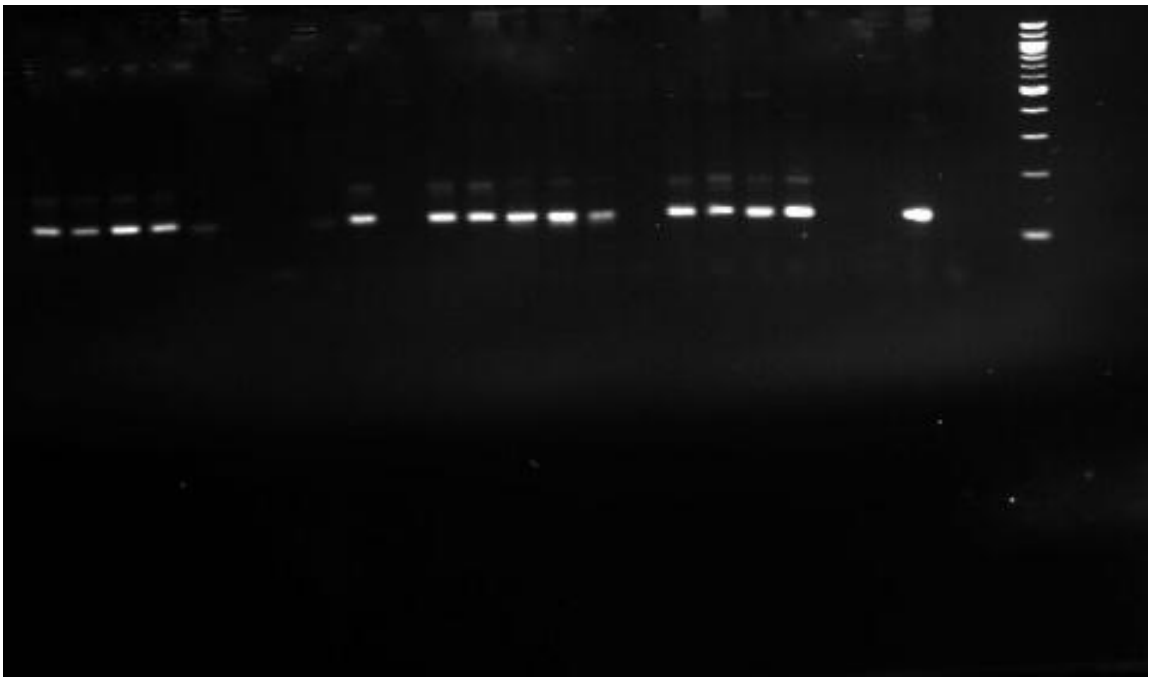


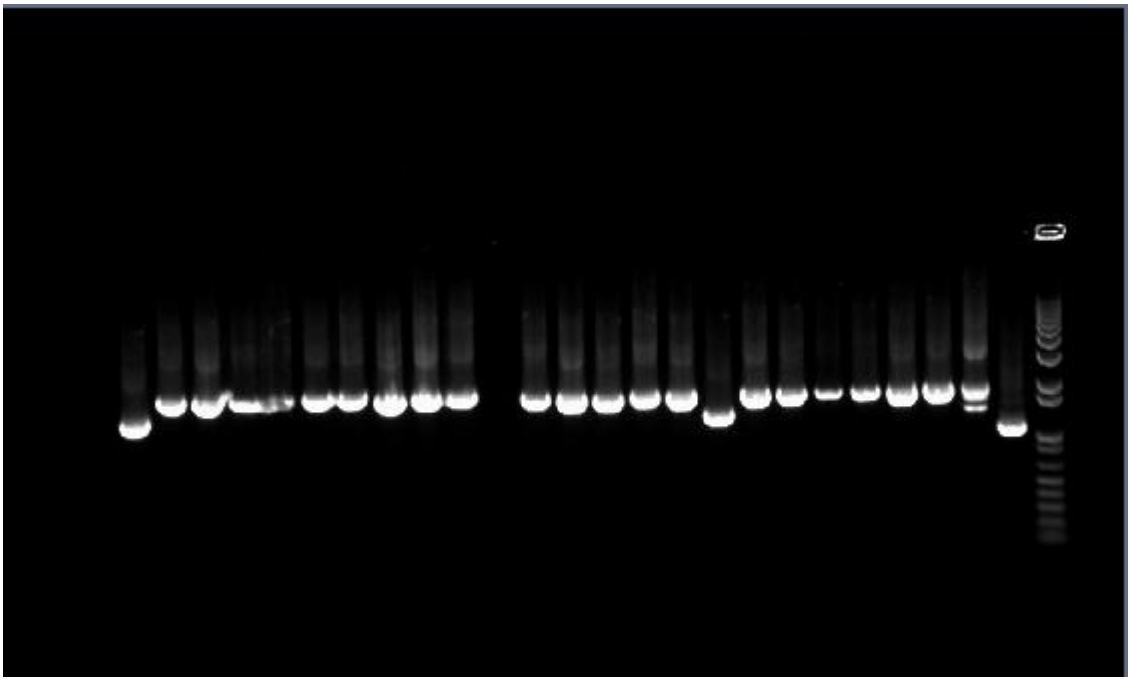
Figure 21 DBS filter paper elution after shaking overnight

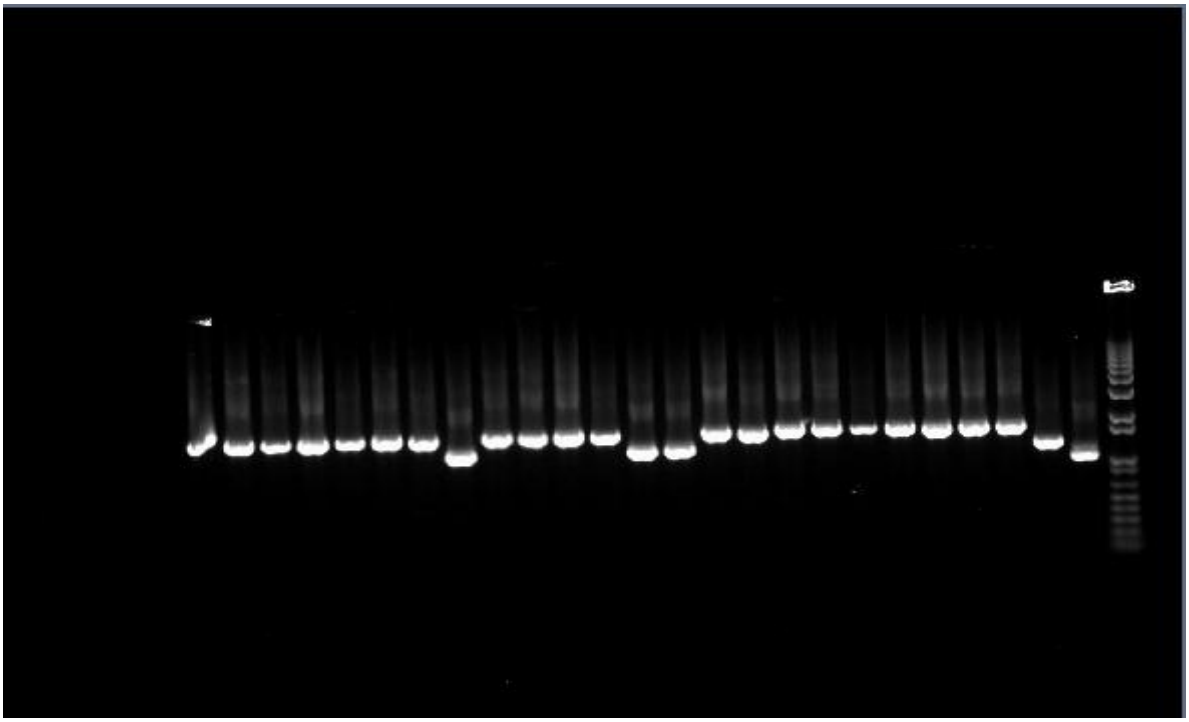
Appendix B. *Plasmodium vivax* positive samples by Nested PCR procedure,Gel images

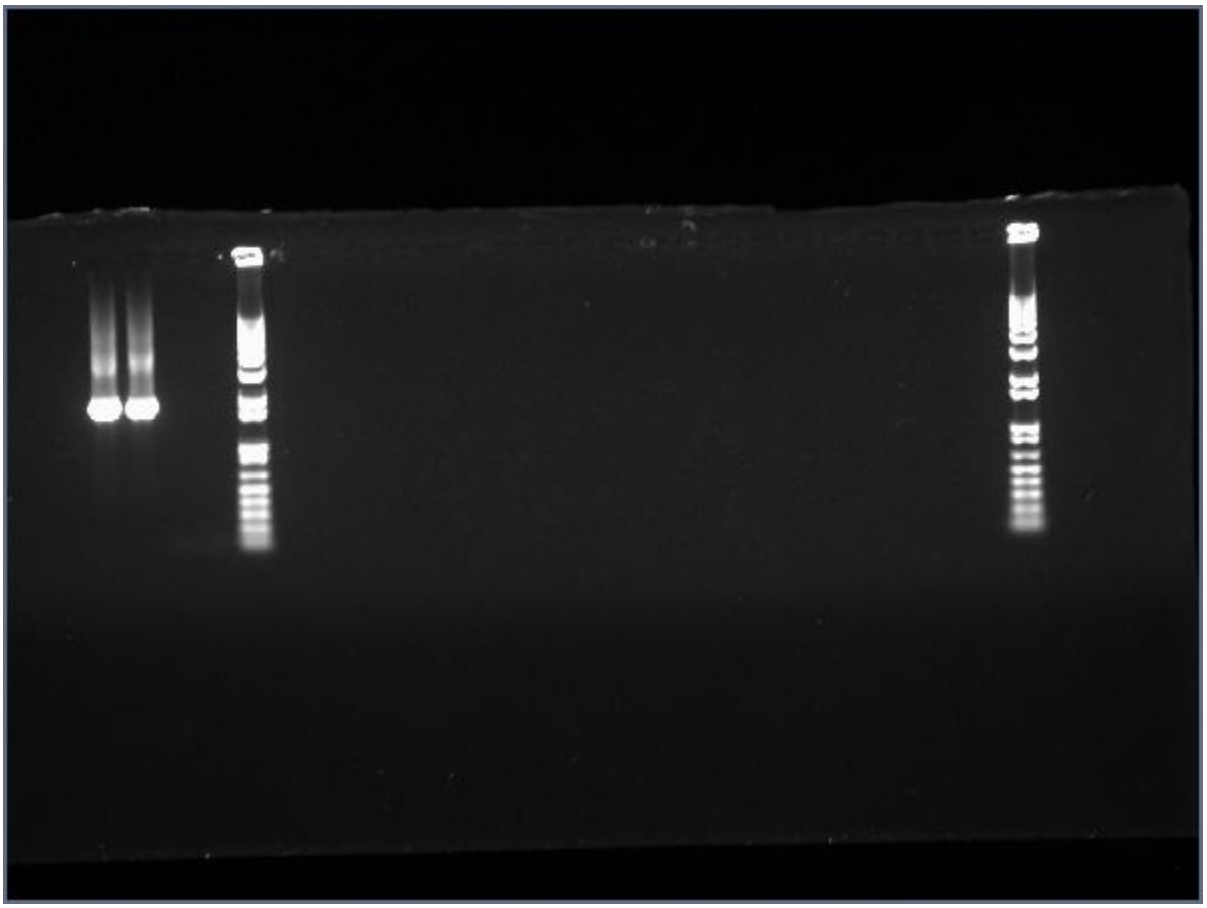




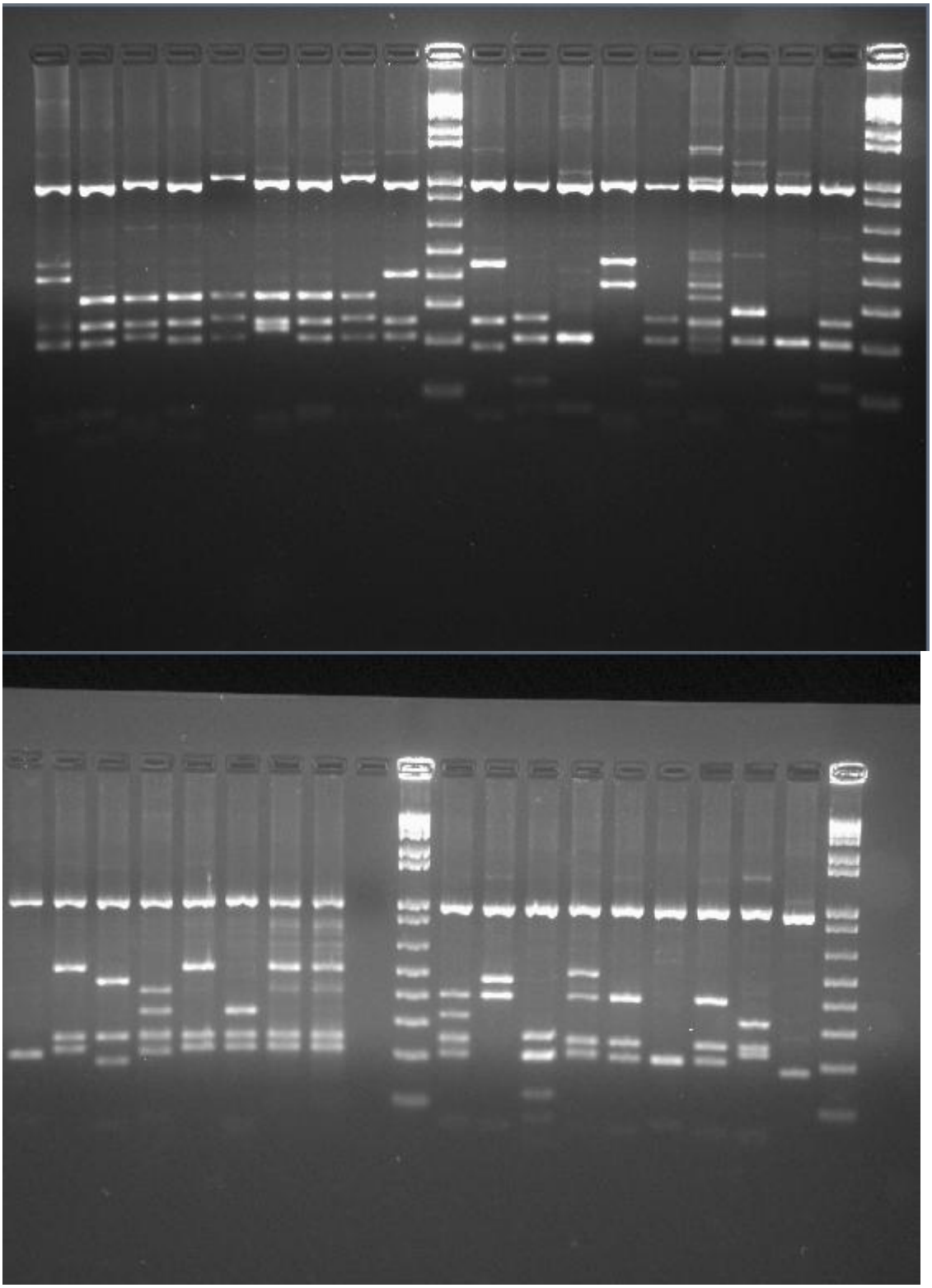
Appendix C. MSP 3α Nested Products, gel Images

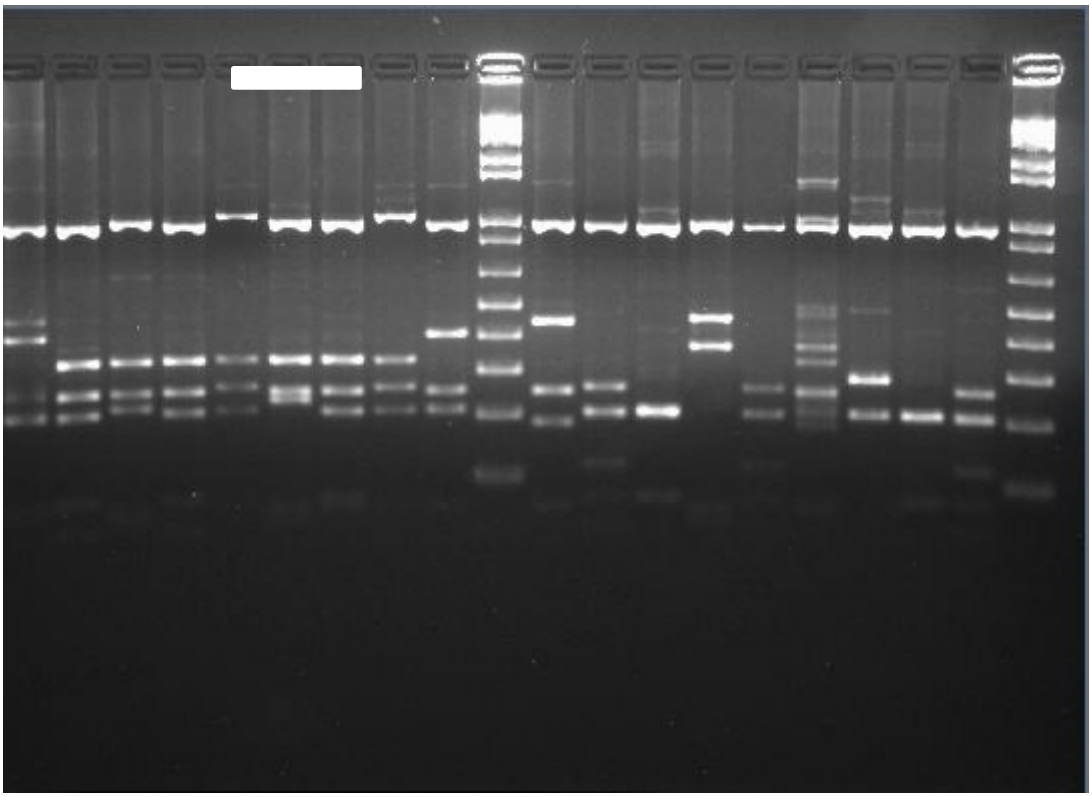






Appendix D Restriction fragment length polymorphism, *Hha I* digestion products





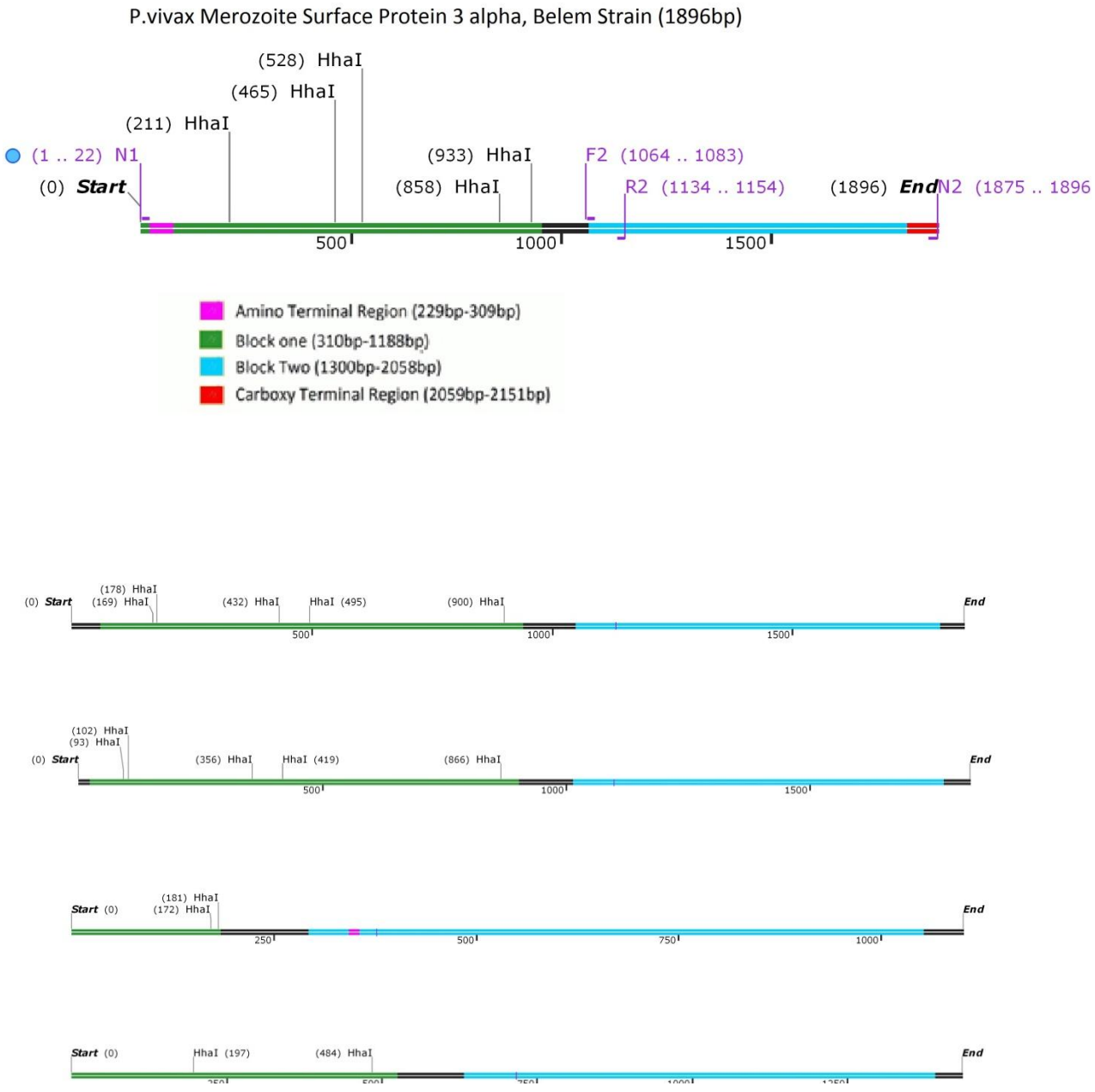
Appendix E. Reference sequence and primers used for nested PCR and Sanger sequencing

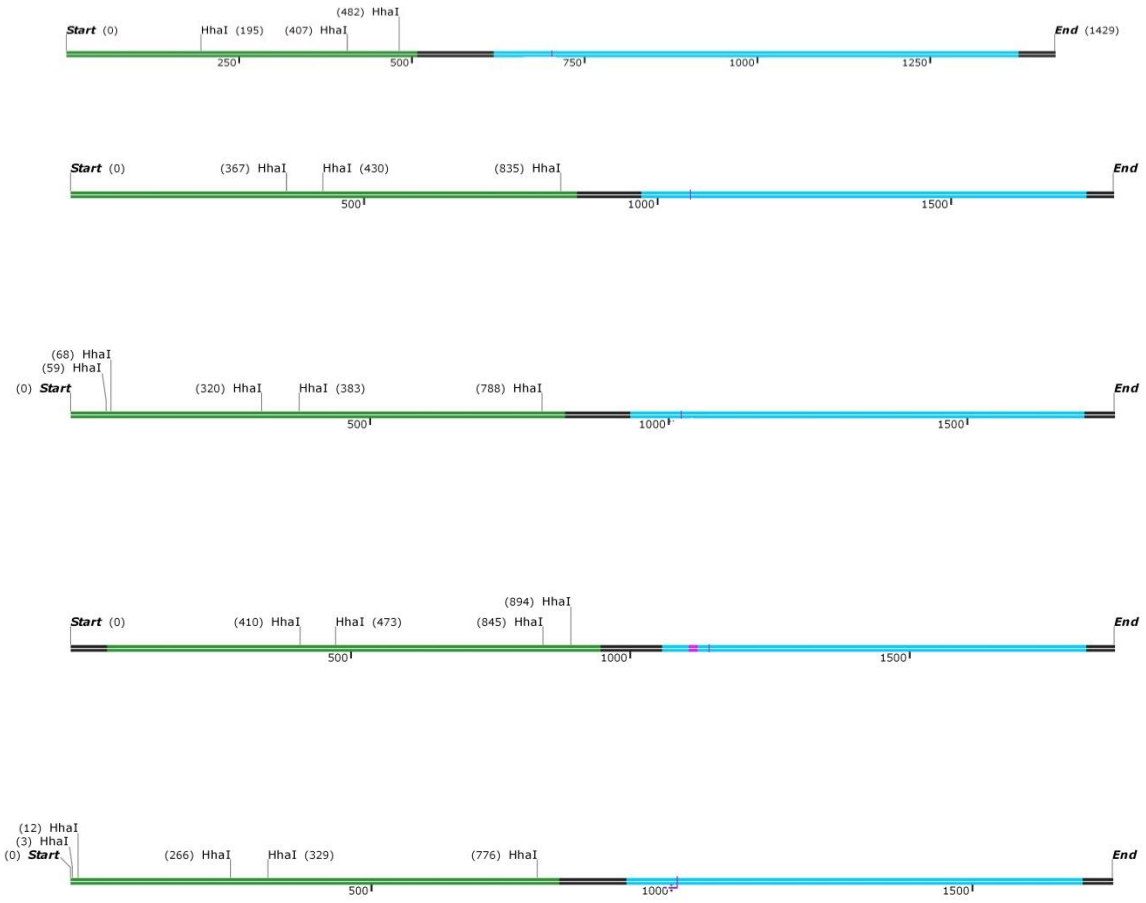
> AF093584.2 *Plasmodium vivax* merozoite surface protein-3 (Msp-3) gene,

GACCAGTGTGATACCATTAACCTAGATCAAGTGACCGAGAAGGAAAAAAAAACAATAGAAGAGGCAAG
CGTGCAAGCACAGGATGGTCCAAATGCTGAGCCAAATAATGCGGAACAGATACAAGCAGAATTGCAAAA
AGTCAAAACAGCAAAAGAAAATTCAGCAGCCTCAGCTACGGCTGCAGAAACTGCTAAAAACAACGCAGT
GAGCGCAGGAAAGGGATTAGATGCAGCGAAAACAGCCATAGAAAAGGCAAAAGCAGCAGCAGAGGAA
GCGAAAAAAGAAGCTGCTATAGCAGAAAAGGCGGAAAAGGACGCTGAGGCAGCTCAAAAGAAAGATA
ATTTGGAAGATGTAAAAAGTCAAGTACCAACTGCAGAAACAGCGGCAACCAATGCTGAGAAAAAAAAAA
CGAAGGCGGAAATAGCCGTAGAGATAGTCAAAGCAGTAGTGGCAAAGGAGGAAGCGCAAAAGGCGTC
AGATGAGGCCCAGAAGGCATGCGAGAAAGCCAAGAAGGCACACGCGAAAGCGCAAAAGGCTTCTGATA
CAACAAAAACGGTAGAAACGTTCAAAACAAACGCAGAAGCTGCGGCAAATAAGGCAAAGGATGAAGCG
GGCAAAGCCGAAACAGCGGCGACGGATGCTAAAAGTACTGAAGATTTAAGCGTAGCAAAAGACAAAGC
GAAAGCAGCAGAAACAGCCGCTATAGAAGCGAAAAAGCAACAGGTAAAAGCAGAAATAGCTGCCGAAG
TGGCGAAGGCAAAAGTTGCAAAAGAAGAAGCAGATGCAGCACAAAAAAAAGCAGAAGAGGCAAAGAA
AATCGTAGACAAAATAGCAAAAGATTCTGAAGTACCAGAGGCGCAAAAGGCAGCAGAATTCGCCACTGA
AACGGTTAAAAAGGCAACCACCGCGGCGACAGAAGCAGGCGAGAAAGCGCAAGAAGCTGAAAACTTAC
CAGCAGAAGCAGAAACGTCAGACGCGGTAAAAGAAAAGGCCGACGGTGCAGAGAAAGCCGCAGGAGA
AGCAAAAAAAGCATCGATCGAAACGGAAATAGCAGTAGAGGTAGCAAAGGCGGAAGTGCTGAACGCA
GAAGTAAAAAAGACAGCACAAGAAGCGGAAAATGACGCAACGGAGGCTAAGAAGCAGGCAGAAAAGG
CTAAGGCAGCAGCGGAAGAGGCAAAGACGCACGGAGAAAAAGCAGAAAAGGTAGGAGAGTCAACAAA
AGCACATTCAGATGAAGCACAGCAAGAAAATAAAAACGCAAAGGACGCCTCGGAAGAAGCAGAAAACA
GAGCGGTGGACGCATTGGAAGAAGCGTATGCAGTGGAAGCACATCTGGCAAGAACCAAAAATGCAGCA
GAGTCAGCCAAAAGTGCAACAGATTTGAGTAAATTAGAAGAGGCAAAGGAGGAAGCGATAGATGCAGC
AAATATAGCTCATCAAAAGTGGTTAAAAGCAACCCAGGCAGCTACCATTGCAAAAGAAAAGAAGGAAGC
TGCAAAAGTAGCCGCTGAAAAAGCACAAAAAGAAGCAACCGCTGCAAAATTAAAGGCAGCTAAGGCTG
AAGCGAAGAAGGCAGAAACAGAAGCGGTAAAGGCAGCGGTAGAAGCGAGAGCGGCGGCGGCGGAAG
CGAAACAAGAGGCAGCAAAAGTAGGGGCATCGAAAGAACCACAAGAAACTAAAAATAAAGCAAATGTG
GAAGCAGAAGCAACAGGAAATGAAGCAAAAAAGGCAGAAGATGCCGCGGAGGAAGCTAAAGAAGCAG
CCAAAAAAGCAAATGAAGCAACCGATGCAGATGTAGCAAGAGCAGAGGCAGACAAAACAATCGCCGCA
GCGAAGAAGGCAAAGAAGGCGAGAGAAAAAGCAGCGTACGGGCTCCTGAAGACGAAGAACCAGTAT

N1 (fw)	GACCAGTGTGATACCATTAACC	22-mer	45% GC	6703.4 Da
N2 (rv)	ATACTGGTTCTTCGTCTTCAGG	22-mer	45% GC	6707.4 Da
F2 (fw)	CAAAGGCGGAAGTGCTGAAC	20-mer	55% GC	6200.1 Da
R2 (rv)	GCCTTTTCTGCCTGCTTCTTA	20-mer	48% GC	6305.1 Da

Appendix F. Pictorial representation of the blocks and restriction enzyme recognition sites on the reference Belem strain and some of the isolates





*The color black represents the Intervening sequence predicted to act as helix breaking β turn(32).

Appendix G. List of sequences used for analysis

genbank accession number	Country of Isolate	genbank accession Number	Country of Isolate
KR902511.1_	Thailand	AY833011.1_	Thailand
KR902512.1_	Thailand	AY833012.1_	Thailand
KR902513.1_	Thailand	AY833013.1_	Thailand
KR902514.1_	Thailand	AY833014.1_	Thailand
KR902515.1_	Thailand	AY833015.1_	Thailand
KR902516.1_	Thailand	AY833016.1_	Thailand
KR902519.1_	Thailand	AY833017.1_	Thailand
KR902520.1_	Thailand	AY833018.1_	Thailand
KR902521.1_	Thailand	AY833019.1_	Thailand
KR902522.1_	Thailand	AY833020.1	Thailand
KR902523.1_	Thailand	AY833021.1_	Thailand
KR902524.1_	Thailand	AY833022.1_	Thailand
KR902525.1_	Thailand	AY833023.1_	Thailand
KR902526.1_	Thailand	AY833024.1_	Thailand
KR902528.1_	Thailand	AY833025.1	Thailand
KR902529.1_	Thailand	AJ864967.1_	Venezuela
KR902530.1_	Thailand	AJ864941.1_	Venezuela
AF491945.1_	Brazil	AJ864942.1_	Venezuela
AF491946.1_	Brazil	AJ864944.1_.	Venezuela
AF491947.1_	Brazil	GU175285.1_	Sri Lanka
AF491948.1_	Brazil	GU175269.1_	Sri Lanka
AF491949.1_	Brazil	GU175270.1_.	Sri Lanka
AF491950.1_	0ng	GU175271.1_	Sri Lanka
AF491951.1_	Bangladesh	GU175272.1_	Sri Lanka
AF491952.1_	Ecuador	GU175277.1_	Sri Lanka
AF491955.1_	Papua New Guinea	GU175279.1_	Sri Lanka
AF491956.1_	Brazil	EU430600.1_	Myanmar
AF491957.1_	India	EU430576.1_	Myanmar
AF491958.1_	North Korea	EU430577.1_	Myanmar
AF491959.1_	Papua New Guinea	EU430578.1_.	Myanmar
AF491961.1_	Sri Lanka	EU430579.1_.	Myanmar
AF491962.1_	Thailand	EU430580.1_.	Myanmar
AY266091.1_.	Indonesia	EU430581.1_	Myanmar
AY118174.1_	Indonesia	EU430582.1_	Myanmar
AY266089.1_	Malaysia	EU430583.1.	Myanmar
AY266090.1_	Pakistan	EU430584.1_	Myanmar
AY266087.1_	South Korea	EU430585.1_	Myanmar
AY833026.1_	Thailand	EU430586.1_	Myanmar
AY833010.1_	Thailand	KC935447.1_	Panama

AY833011.1_	Thailand	KC935422.1_	Venezuela
AY833012.1_	Thailand	KC935423.1_	Venezuela
AY833013.1_	Thailand	KC935424.1_	Venezuela
AY833014.1_	Thailand	KC935426.1_	Venezuela
AY833015.1_	Thailand	KC935427.1_	Venezuela
AY833016.1_	Thailand	KC935428.1_	Venezuela
AY833017.1_	Thailand	KC935429.1_	Venezuela
AY833018.1_	Thailand	KC935430.1_	Venezuela
AY833019.1_	Thailand	KC935431.1_	Venezuela
AY833020.1	Thailand	KC935441.1_	Indonesia
AY833021.1_	Thailand	KC935442.1_	Thaliand
AY833022.1_	Thailand	KC935443_	Vietnam
AY833023.1_	Thailand	AY833010.1_	Thailand
AY833024.1_	Thailand	KC935445.1_	Mauritania
AY833025.1	Thailand	KC935446.1_	India
AJ864967.1_	Venezuela	KC935425.1_	Venezuela
AJ864941.1_	Venezuela	JQ317289.1_	South Korea
AJ864942.1_	Venezuela	JQ317283.1_	South Korea
AJ864944.1_.	Venezuela	JQ317284.1_	South Korea
GU175285.1_	Sri Lanka	JQ317285.1_	South Korea
GU175269.1_	Sri Lanka	JQ317286.1_	South Korea
GU175270.1_.	Sri Lanka	JQ317287.1_	South Korea
GU175271.1_	Sri Lanka	JQ317288.1_	South Korea
GU175272.1_	Sri Lanka	EF204163.1_	South Korea
GU175277.1_	Sri Lanka	EF204144.1_	South Korea
GU175279.1_	Sri Lanka	EF204145.1_	South Korea
EU430600.1_	Myanmar	EF204146.1_	South Korea
EU430576.1_	Myanmar	EF204147.1	South Korea
EU430577.1_	Myanmar	EF204148.1_	South Korea
EU430578.1_.	Myanmar	EF204149.1_	South Korea
EU430579.1_.	Myanmar	EF204150.1_	South Korea
EU430580.1_.	Myanmar	EF204151.1_	South Korea
EU430581.1_	Myanmar	EF204152.1_	South Korea
EU430582.1_	Myanmar	EF204153.1_	South Korea
EU430583.1.	Myanmar	EF204154.1_	South Korea
EU430584.1_	Myanmar	EF204155.1_	South Korea
EU430585.1_	Myanmar	EF204156.1_	South Korea
EU430586.1_	Myanmar	EF204157.1_	South Korea
KC935447.1_	Panama	EF204158.1_	South Korea
KC935422.1_	Venezuela	EF204159.1_	South Korea

Appendix H. Name and Sequences of the 39 haplotypes

Hap_1		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_2		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_3		A	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_4		A	A	C	C	G	A	T	T	A	T	A	G	A	A	A	G	C	C	A	A	A	A	G	G	A	A	A	A	C	C	C	A
Hap_5		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	A	G	G	A	A	A	A	C	C	C	A
Hap_6		G	A	C	C	G	A	C	T	A	T	T	A	G	A	G	G	A	T	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_7		G	A	C	C	G	A	C	T	A	T	A	G	A	T	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_8		G	A	C	C	G	A	C	T	A	T	T	A	G	A	G	G	A	T	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_9		A	A	C	C	G	A	T	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_10		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	T	C	T	C	C	T	T	A
Hap_11		G	A	C	T	G	A	C	T	A	T	T	A	G	A	G	G	C	C	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_12		G	A	C	C	G	A	C	T	A	T	T	A	G	A	G	G	A	T	A	A	A	A	G	A	A	A	A	C	C	C	G
Hap_13		G	A	C	C	G	A	T	T	A	T	T	A	G	A	G	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_14		A	A	A	A	G	T	T	T	A	T	T	A	G	A	G	G	A	C	A	A	G	A	A	A	A	A	A	C	C	C	A
Hap_15		G	A	C	C	G	A	C	T	T	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_16		G	A	C	C	G	A	C	T	A	T	T	A	G	A	G	G	C	C	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_17		G	A	C	C	G	A	T	T	A	T	A	G	A	A	A	G	C	C	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_18		G	A	C	C	G	A	C	T	A	T	A	G	A	A	G	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_19		G	A	C	C	G	A	C	T	A	T	A	G	G	A	A	G	C	C	A	A	G	G	G	A	A	T	C	A	T	T	A
Hap_20		G	A	C	C	G	A	C	T	A	T	A	G	A	G	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_21		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	G	A	G	G	G	A	C	T	C	A	T	T	A
Hap_22		A	A	C	C	G	A	T	T	A	C	T	A	G	A	G	G	A	T	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_23		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	G
Hap_24		A	A	C	C	G	A	T	T	A	T	T	A	G	A	G	G	A	T	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_25		A	A	C	C	G	A	T	T	T	T	A	G	A	A	A	G	C	C	A	A	A	A	G	T	A	A	A	C	C	C	A
Hap_26		G	A	C	C	G	A	A	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_27		A	A	C	C	G	A	T	T	A	T	A	G	A	A	A	G	C	C	G	C	A	A	C	C	G	A	T	A	G	C	G
Hap_28		A	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	G	A	G	G	G	A	C	T	C	A	T	T	A
Hap_29		A	A	C	C	G	A	C	C	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_30		A	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	G	G	G	G	A	C	T	C	A	T	T	A
Hap_31		G	A	C	C	G	A	C	C	A	T	A	G	A	A	A	G	C	C	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_32		G	A	C	C	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	C	T	T	A
Hap_33		A	A	A	A	G	T	T	T	A	T	T	A	G	A	G	G	A	C	A	A	A	A	A	A	A	A	A	C	C	C	A
Hap_34		A	A	C	A	A	A	C	T	A	T	T	A	G	A	G	G	A	T	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_35		A	A	C	C	G	A	T	T	G	C	T	A	G	A	G	G	A	T	A	A	A	A	G	A	A	A	A	C	C	C	A
Hap_36		G	A	C	C	G	A	T	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_37		G	A	C	T	G	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_38		A	A	C	C	G	A	C	T	A	T	A	G	A	A	A	C	C	C	A	A	G	G	G	A	C	T	C	A	T	T	A
Hap_39		G	G	C	A	A	A	C	T	A	T	A	G	A	A	A	G	C	C	A	A	A	A	G	A	A	A	A	C	C	C	A