POPULATION GENETIC ANALYSIS OF *PLASMODIUM FALCIPARUM* CIRCUMSPOROZOITE PROTEIN (CSP) IN TWO DISTINCT ECOLOGICAL REGIONS IN GHANA

ELIKPLIM ADZO AMEGASHIE

MASTER OF SCIENCE

(Bioinformatics and Molecular Biology)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2021

Population Genetic Analysis of *Plasmodium falciparum* Circumsporozoite Protein (CSP) in Two Distinct Ecological Regions in Ghana

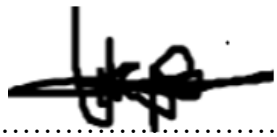
Elikplim Adzo Amegashie

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

2021

DECLARATION

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

Signature:  ……..Date:……...……

Elikplim Adzo Amegashie

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

Signature:……..Date:……...……

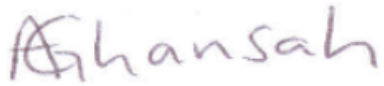
Dr. Peter Ogoti Mose, PhD

JKUAT, Kenya

Signature:……..Date:……...……

Dr. Kevin Mbogo, PhD,

JKUAT, Kenya

Signature:  ……..Date:……...……

Dr. Anita Ghansah, PhD

UG, Ghana

DEDICATION

I dedicate this work to my family: Mr. Gideon Amegashie, Mrs. Josephine Amegashie, Edem Amegashie, Mrs. Mona Abass, Mrs. Mabel Agbo and Derek Fiifi Diaba

ACKNOWLEDGEMENT

I have finally reached the thousandth mile of the journey I undertook to obtain my Master's degree. It has been such a rollercoaster. I could not have done it all by myself, others joined me on this ride. First of all, my utmost gratitude goes to God Almighty for His banner over me throughout my research.

I would like to express my deepest gratitude to my external Supervisor, Dr. Anita Ghansah for the immense support, opportunity, guidance and motivation throughout the research work. It was such a great honor and privilege to work under her supervision. Special thanks go to the Developing Excellence in Leadership and Genetic Training for Malaria Elimination (DELGEME) (Wellcome Trust Grant) for the funding, immense academic and moral support and the additional training provided during DELGEME summer courses. I would also like to thank the DELGEME team for their input, corrections, inciteful comments and mentorship throughout the entire research.

I am grateful to Dr. Steven Ger for his guidance during my stay in JKUAT as a foreign student. Special thanks go to my JKUAT supervisors, Dr. Kevin Mbogo and Dr. Peter Ogoti for their input, inciteful comments and support towards the completion of my thesis. My heartfelt appreciation of all JKUAT DELGEME fellows, especially Kevin Esoh, Jennifer Mutisya and Deriba Abera. I would forever cherish the companionship, support, motivation, the encouragement, and the exchange of ideas regarding analysis. I am extremely grateful to Mr. Courage Adobor of the Noguchi Memorial Institute of Medical Research (NMIMR) for the

tremendous support regarding this research, I could not have gone this far without your intense tutoring and guidance.

Finally, I am very grateful for my family: Mr. Gideon Amegashie, Mrs. Josephine Amegashie, Edem Amegashie, Mrs. Mona Abass, Mrs. Mabel Agbo and the love of my life Derek Fiifi Diaba. I could not have done this without all of them. I am grateful for their prayers, love, spiritual and emotional support as well as their encouragement. I say God richly bless everyone for making this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AEIR	Annual Entomological Inoculation Rate
ACT	Artemisinin Combination Therapy
AMA1	Apical Membrane Antigen 1
CSP	Circumsporozoite Protein
CRR	Central Repeat Region
CNV	Copy Number Variations
EBA	Erythrocyte Binding Antigen
EER	Exo Erythrocytic Forms
EIR	Entomological Inoculation Rate
EHH	Extended Haplotype Homozygosity
GATK	Genome Analysis Tool Kit
iHS	Integrated Haplotype Score
ITN	Insecticide Treated net
MAF	Minor Allele Frequency
MalariaGEN	Malaria Genomic Epidemiology Network
MOI	Multiplicity Of Infection
MSP	Merozoite Surface Protein
MVIP	Malaria Vaccine Implementation Program
NGS	Next Generation Sequencing
NMCP	National Malaria Control Program
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction

PDNA	Plasmodium Network Diversity Africa
PfCSP:	*Plasmodium falciparum* Circumsporozoite Protein
RDT	Rapid Diagnostic Test
SMC	Seasonal Malaria Chemoprevention
SNP	Single Nucleotide Polymorphism
TH2R	$CD4^+$ T-helper Receptor epitope
TH3R	$CD8^+$ T-helper Receptor epitope
VCF	Variant Call Format
WHO	World Health Organization

ABSTRACT

Malaria caused by *Plasmodium falciparum* parasite continues to be of major public health concern in Ghana. RTS,S/AS01 is the most advanced malaria vaccine candidate designed based on the 3D7 laboratory *Plasmodium falciparum* Circumsporozoite protein (PfCSP). Extensive diversity of the gene encoding the PfCSP (*Pfcsp*) is a major contributing factor to the moderate efficacy of the RTS,S/AS01 vaccine. The intensity of malaria transmission influences the extent of genetic variation within and between malaria parasite populations. However, the extent and dynamics of genetic diversity of *Pfcsp* in Ghana is poorly understood. Therefore, this study sought to investigate the extent of genetic diversity of *Pfcsp* parasite populations within and between two eco-epidemiologically distinct regions in Ghana and ascertain how forces of evolution such as selection shapes this diversity. Illumina generated paired-end short-read DNA sequences of *P. falciparum* parasites were retrieved from 101 and 131 blood samples taken from children between the age range of 6-59 months presenting with Clinical malaria at Cape Coast (located on the coastal belt) and Navrongo (in the Northern savannah region) respectively. The sequences were mapped to the genome of the 3D7 reference strain yielding high-quality genome-wide coding sequence data. After which data filtering and quality checks were done to remove missing data. There remained a total of 220 isolates which were analysed for allele frequency spectrum, genetic diversity both within host and between the populations and signatures of selection. *Pfcsp* was extensively diverse at the two sites with the higher transmission site, Navrongo, recording both higher within host and population level diversity as seen in higher malaria transmission regions. The vaccine strain C-terminal epitope of *Pfcsp* was found in only 5.9% and 45.7% of the Navrongo and Cape Coast sequences respectively indicating that the RTS,S vaccine might be more efficacious in Cape Coast than Navrongo. Amino acid variations ranging between 1 and 6 were observed specifically in the TH2R ($CD4^+$) and TH3R ($CD8^+$) immunodominant epitope regions of the PfCSP which could potentially reduce vaccine efficacy. Tajima's D was negatively skewed mainly Cape Coast isolates as expected from historical population expansion. On the contrary, positive Tajima's D was observed for the Navrongo *P. falciparum* population, consistent with balancing selection acting on the immunodominant TH2R and TH3R vaccine epitopes. These results have implications on the efficacy of the vaccine in Ghana and will inform the choice of alleles to include in future multivalent or chimeric vaccines.

CHAPTER ONE

INTRODUCTION

1.1 Background

Malaria is an acute febrile illness caused by *Plasmodium* parasites transmitted by the female Anopheles mosquito (vector). According to the World Health Organization (WHO), malaria cases reported worldwide were about 228 million in the year 2018 accounting for 400,000 deaths (WHO, 2019). An estimated 213 million of the malaria cases were reported in the African region accounting for 93% of the global malaria burden. Malaria caused by *Plasmodium falciparum* is the most prevalent in Africa causing about 91% of the world's malaria mortality. This continuously poses a threat to the health of individuals especially children under five years and pregnant women (WHO, 2019). Malaria control interventions include insecticide-treated bed nets (ITNs) and indoor residual spraying (IRS) however there have been several reports of insecticide and antimalarial drugs resistance of the vector and the parasite in several countries including Ghana. Of the 10 highest burden countries in Africa, Ghana is reported to be one of the countries with the highest absolute increases in cases of malaria in 2018 compared with 2017 (WHO, 2019) .

In Ghana, malaria transmission persist perennially but with marked seasonal peaks that vary with local ecology and overall transmission intensity (Koram et al., 2003). For the purpose of malaria control, transmission has been categorized eco-epidemiologically into three main zones across Ghana namely; forest ecology with perennial but increased malaria transmission during the rainy season (May-August and October-November), northern/Guinea savannah with seasonal and intense transmission during the rainy season (June-October), but with very low periods of transmission during dry season, and coastal savannah with low to moderate perennial transmission and a marked seasonal effect during the rainy season (Abuaku et al., 2019).

An addition of a broadly effective vaccine to compliment the panel of available malaria control tools is very necessary for malaria elimination and potential eradication. In this regard, a major goal of WHO is to develop safe and effective

licensed vaccines against *Plasmodium falciparum* with a protective efficacy of at least 75% by the year 2030. During the various stages of the life cycle of *P. falciparum* parasite, proteins expressed on its surface are being developed as vaccine candidates (Richards & Beeson, 2009).

The RTS,S/AS01 vaccine is the only vaccine that has successfully completed the phase III clinical trials and was approved by the European Medicines Agency in 2015 to undergo a Phase IV pilot clinical trial (European Medicines Agency, 2015). Efficacy data from a Phase III clinical trial conducted across 11 sites in 7 African countries in children (aged 5 to 17 months) and infants (aged 6-12 weeks) revealed that the vaccine conferred moderate protective efficacy against clinical disease and severe malaria which waned over time (Clinical Trial Partnership, 2015). The vaccine conferred only 36.3% protection against clinical malaria and 32.2% against severe malaria in children aged 5-17 months who received 3 primary doses of RTS,S with a booster in the 20th month (Clinical Trial Partnership, 2015). Currently, the phase IV trial is being implemented in three countries including Ghana, Kenya and Malawi via The Malaria Vaccine Implementation Programme (MVIP) led by WHO (WHO, 2019).

The RTS,S/AS01 vaccine was designed based on a fragment of Circumsporozoite protein (CSP) obtained from the *Plasmodium falciparum* 3D7 laboratory strain coupled with the Hepatitis B surface antigen, which serves as a carrier molecule (Stoute et al., 1997). The CSP is the main protein expressed on the outer membrane of the sporozoites during the pre-erythrocytic stage of *P. falciparum* (Mohapatra, 2013). Antigens expressed during this stage are considered as ideal vaccines due to their ability to produce antibodies and induce T-cell mediated immunity which prevent infection by inhibiting sporozoite motility and preventing hepatocyte invasion (Ouattara & Laurens, 2015). For cell-mediated immunity, RTS,S includes a fragment of the central NANP-NVDP repeat polymorphic B-cell epitope region and a highly polymorphic C-terminal non-repeat epitope region of PfCSP, which covers $CD4^+$ and $CD8^+$ T-cell epitopes denoted as TH2R and TH3R, respectively (Casares et al., 2010).

Several studies have reported high levels of polymorphism in the T-cell epitopes within the C-terminal region of PfCSP in natural parasite populations (Barry et al., 2009; Gandhi et al., 2014; Jalloh et al., 2009). A recent study by Pringle et al., (2018) has shown that the extent of genetic diversity of the *Plasmodium falciparum* CSP (PfCSP) may have led to the RTS,S/AS01 vaccine being less efficacious thus performing poorly against mismatch strains (Pringle et al., 2018). Subsequently, an ancillary next-generation deep sequencing analysis of Phase III trial samples in 2015 showed that the vaccine indeed conferred partial protection against clinical malaria for strain-specific vaccine alleles (50.3%) and poor protection against mismatched strains (33.3%) (Neafsey et al., 2015). Therefore, characterisation of the extent of genetic diversity of *Pfcsp* is important, especially in Countries such as Ghana, where the vaccine is undergoing the Phase IV implementation trial. Additionally, recent studies of the population structure of *Pfcsp* suggest that geographically variable levels of diversity may have an impact on the efficacy of *Pfcsp* based malaria vaccines in specific geographic regions (Barry et al., 2009; Pringle et al., 2018). In light of this, spatial variation of *Pfcsp* capturing two different eco-epidemiological regions involving Cape Coast and Navrongo in Ghana was assessed. Evolutionary factors such as selection operating on parasites differ locally owing to varying transmission patterns, ecology and degrees of acquired immunity in humans (Duffy et al., 2015). Therefore, a scan for signatures of selection on *Pfcsp* was assessed to determine how evolution shapes the diversity of *Pfcsp* in Ghana. Overall, this study provides insights into how well the RTS,S/AS01 vaccine might perform if implemented in Ghana.

1.2 Problem statement

Malaria control and elimination measures including the use of bed nets, insecticides and antimalarial drugs is being hindered by drug and insecticide resistance of the parasite and the mosquito. An ideal malaria control measure will be the introduction of a broadly potent vaccine. Many malaria vaccines under development have undergone various clinical trials but the complete potency of these vaccines have been hindered by factors such as extreme diversity of leading candidate vaccine antigens (Cortes et al., 2003; Pringle et al., 2018).

The RTS,S/AS01 being the most advanced malaria vaccine have been designed based on only the 3D7 laboratory strain of *Plasmodium falciparum* CSP without considering the broad range of evolving natural parasite populations. Various studies have reported the polymorphic nature of gene encoding the Circumsporozoite protein (*Pfcsp*) coupled with evidence of allele-specific immunity in natural parasite populations. Ghana has been selected as one of the countries for implementation of RTS,S Phase IV clinical trials. However, the extent of genetic diversity of *Pfcsp* taking into account the geographical variations, differences in malaria transmission intensities, as well evolutionary factors shaping the diversity in this malaria endemic country is poorly understood.

1.3 Justification

Phase III clinical trials and ancillary studies of genetic diversity of *Pfcsp* in Ghana have only been conducted in two sites located in the forest zones in the country. To provide new insights into how well RTS,S/AS01 may perform if implemented on a large scale in different regions and transmission setting, there is the need to determine the extent and dynamics of genetic diversity in other regions of the country capturing two eco-epidemiologically contrasting regions in Ghana specifically Cape Coast (Coastal zone) and Navrongo (Northern savannah zone) and to assess the intra-host diversity of this gene as well. Findings from this study is extremely critical in informing the outcome of the ongoing Phase IV pilot rollout of RTS,S in Ghana and future malaria vaccine trials for both Circumsporozoite and non-Circumsporozoite based vaccines. This should provide broader assessment of the extent to which the local natural diversity of *Pfcsp* could impact its efficacy and wider implementation in Ghana. In addition, this study provides baseline information on the extent of diversity of *Pfcsp* and the evolutionary forces driving these variations within Ghanaian natural parasite population

1.4 Hypothesis

There is no genetic and spatial variation in the gene encoding *Plasmodium falciparum* Circumsporozoite protein (*Pfcsp*) within selected regions in the Ghanaian malaria parasite populations

1.5 Objectives

1.5.1 General Objective

To analyse the population genetics of *Plasmodium falciparum* Circumsporozoite protein in two distinct ecological regions in Ghana

1.5.2 Specific Objectives

1. To estimate the extent of genetic variation of *Pfcsp* within Cape Coast and Navrongo parasite populations.
2. To determine the spatial variations of *Pfcsp* between Cape Coast and Navrongo parasite populations.
3. To evaluate the evidence of selection of *Pfcsp* within Cape Coast and Navrongo parasite populations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Malaria

Malaria is an acute febrile illness caused by Plasmodium parasites transmitted by the female Anopheles mosquito (vector). There are five different Plasmodium species that affect humans namely *Plasmodium falciparum*, *Plasmodium ovale*, *Plasmodium malariae*, *Plasmodium vivax* and *Plasmodium knowlesi* (WHO, 2019). In the year 2017, Malaria cases reported in 87 countries worldwide were about 219 million accounting for 435,000 deaths (WHO, 2018). However there has been an increase in number of cases in 2018 where 228 million cases were reported (WHO,2019). An estimated 213 million of the malaria cases were reported in the African region accounting for 93% of the global malaria burden, followed by South-East Asia, with 3.4% of the cases and the Eastern Mediterranean Region with 2.1%. In 2018, 405 000 deaths individuals died from malaria globally, compared with 416 000 estimated deaths in 2017, and 585 000 in 2010. In Africa, malaria caused by *Plasmodium falciparum* is the most prevalent causing about 91% of the world's malaria mortality. This continuously poses a threat to the health of individuals especially children under five years and pregnant women (WHO, 2019). Of the 10 highest burden countries in Africa, Ghana and Nigeria reported the highest absolute increases in cases of malaria in 2018 compared with 2017. The burden in 2018 was similar to that of 2017 in all other countries, apart from in Uganda and India, where there were reported reductions of 1.5 and 2.6 million malaria cases, respectively, in 2018 compared with 2017. In Africa, malaria caused by *Plasmodium falciparum* is the most prevalent causing about 91% of the world's malaria mortality. *Plasmodium falciparum* malaria is transmitted through the bite of the female Anopheles mosquito serving as a vector. Most individuals at risk are children under five years and pregnant women. In 2018, about 24 million children were estimated to be infected with *P. falciparum* in sub-Saharan Africa (WHO, 2019).

The *P. falciparum* is a protozoan parasite and has a life cycle which involves two host, the female Anopheles mosquito and human host **(Figure 2.1).** In the human host, there are various stages of infection. At each stage, the parasite expresses

proteins which serve as targets to induce protective immune response and therefore regarded as malaria vaccine candidates (Beeson et al., 2016).

The first stage is the pre-erythrocytic stage or the exo-erythrocytic stage. The human does not show any symptoms at this stage thus referred to as the clinically silent stage (Hafalla, 2011). On the onset, the female Anopheles probes into the skin, injects sporozoites from its salivary gland into the dermis of the human host. The sporozoites then move actively through the dermis penetrating dermal blood vessels, this is characterized as blood meal (Amino et al., 2006). The sporozoites then move from the bloodstream into the liver and invades hepatocytes, the entire period of movement from dermis to liver occurs within 15 to a few hours after initial inoculation (Yamauchi et al., 2007). When the sporozoites enter the liver cells (hepatocytes), they divide via mitosis and form schizonts which then rupture. A single schizont replicates asexually into several merozoites within a period of about 5-15 days after which they are released from the liver into the bloodstream (Amino *et al*., 2006). Since the pre-erythrocytic stage is a non-disease-causing stage in the human host, *P. falciparum* infection does not induce complete immunity, therefore makes it an ideal stage to develop vaccines aimed at infection prevention (Crompton et al., 2010).

The next stage is the asexual erythrocyte stage which is the disease-causing stage and this is where most clinical symptoms of malaria are manifested. At this stage, the asexually-multiplied merozoites in the bloodstream attaches themselves to erythrocytes, invade them, replicate, and rupture to attack several other red blood cells (Beeson et al., 2016). The parasite migrates into a red cell and takes up residence in the cell and utilize its hemoglobin to build up amino acids. By so doing, the parasite expresses its own proteins and transfers these proteins to the surface of the red cell, after which it is released and invades several other red cells the same way (Beeson et al., 2016). In malaria endemic areas, most individuals repeatedly infected acquire natural immunity against the disease but not necessarily against the infection at this stage (Crompton et al., 2010).

The merozoites then form the ring-shaped immature trophozoites which grow into matured trophozoites. The matured trophozoites form schizonts and rapture to release merozoites once again, this continues the cycle of invasion, replication, and rupture of red blood cells repeatedly at approximately every two days (Crompton et al., 2010).

Some small percentage of immature trophozoites however develop into microgametocytes (male) and macrogametocytes (female) characterizing the gametocyte stage which is also known as the sexual erythrocyte stage (Crompton *et al*., 2010). At this stage, gametocytes are taken up by the female Anopheles mosquito host during a blood meal. There is then fertilization of microgametes and macrogametes in the midgut of the mosquito to form a zygote. The zygote in turn become motile and elongated ookinetes which migrates by invading the epithelial cell of the midgut where they move through the cells to the hemolymph. In the hemolymph they develop into an oocyst which further ruptures to release thousands of sporozoites (Crompton *et al*., 2010). After 10-18 days, the sporozoites are then propelled into the salivary gland of the mosquito. The entire cycle continues whenever the female anopheles mosquito takes a blood meal from a human host.

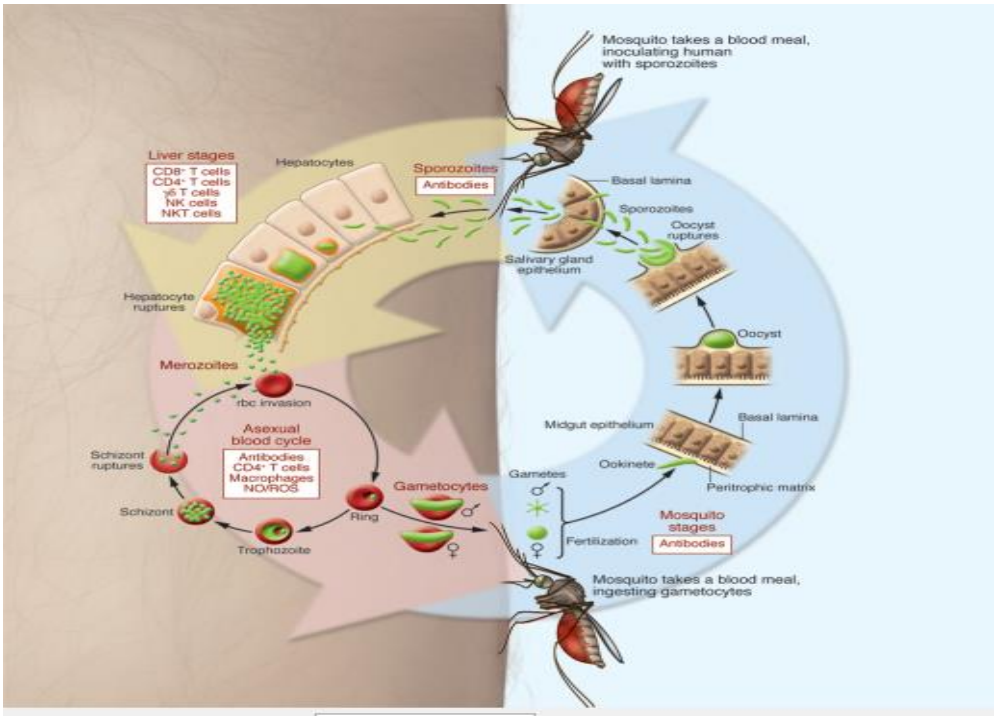


Figure 2.1: *Plasmodium falciparum* life cycle and the immune responses involved. Retrieved from: Crompton *et al*., 2010

2.2 *Plasmodium falciparum* Malaria transmission and its impact in Ghana

There were about 155 million malaria cases in eleven high burden to high impact countries in 2018, compared with 177 million in 2010 (WHO, 2019). Ghana is captured to be among the 10 highest burden malaria countries in Africa. Ghana is located in the sub-region of West Africa **(Figure 2.2**). Ghana together with Nigeria reported the highest absolute increases in cases of malaria in 2018 compared with 2017 (WHO, 2019). However, since 2003, Ghana has received many grants from the Global fund to support the fight against AIDS, Tuberculosis as well as malaria (Global Fund, 2019).

The eco-epidemiological *P. falciparum* malaria transmission zones are categorized into three main zones which consist of the forest zones which has perennial malaria transmission with high transmission during the rainy season (May to August) (Abuaku et al., 2019), the northern savannah zones with seasonal malaria transmission, with intense malaria transmission during the rainy season which begins from June to October (Koram *et al*., 2003), but with periods of relatively low

transmission during dry season (the rest of the year). The coastal savannah with low to moderate perennial transmission with a marked seasonal effect during the major rainy season (May to August) as well (Abuaku *et al*., 2019).

The intensity of malaria transmission is also characterized based on annual entomological inoculation rate (AEIR). AEIR is defined as the number of *P. falciparum* infective bites received by an individual per year. The AEIR is a measure of exposure to infectious mosquitoes. The forest zone is reported to have an annual entomological inoculation rate (AEIR) of about 866 infectious bites per person per year (Appawu et al., 2004; Kasasa et al., 2013; Nkrumah et al., 2014). In the northern savannah zones, the AEIR is generally reported to be up to 157 infective bites per person per year (Abuaku et al., 2019; Kasasa et al., 2013). The Coastal savannah area have an AEIR of 50 infective bites per person which is relatively low compared to other zones (Abuaku et al., 2019; Mensah-Brown et al., 2017).

In Ghana, the *Anopheles gambiae* and *Anopheles funestus* are the most prevalent vector species which spread the *Plasmodium falciparum* parasit*e* in the Forest zone (Abonuusum et al., 2011). However in the northern savannah, the most common vector is *Anopheles arabiensis* whilst *Anopheles melas* is the most common vector in the coastal savannah zones (de Souza *et al*., 2010). However, other parasites such as *Plasmodium malariae* and *Plasmodium ovale* is seen mostly as mixed infections with *P. falciparum*. In addition, *Plasmodium vivax* are yet to be reported in the country (WHO, 2019). Two very distinct sites of interest considered in this study are Cape Coast and Navrongo. The distance between Cape Coast and Navrongo is approximately 784.8 km **(Figure 2.3).**

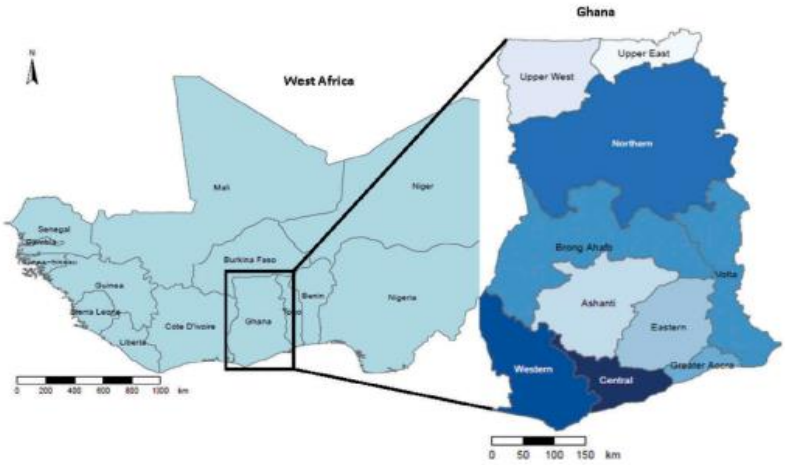


Figure 2.2: A map showing Ghana's location in West Africa: Retrieved from MapCruzin.com

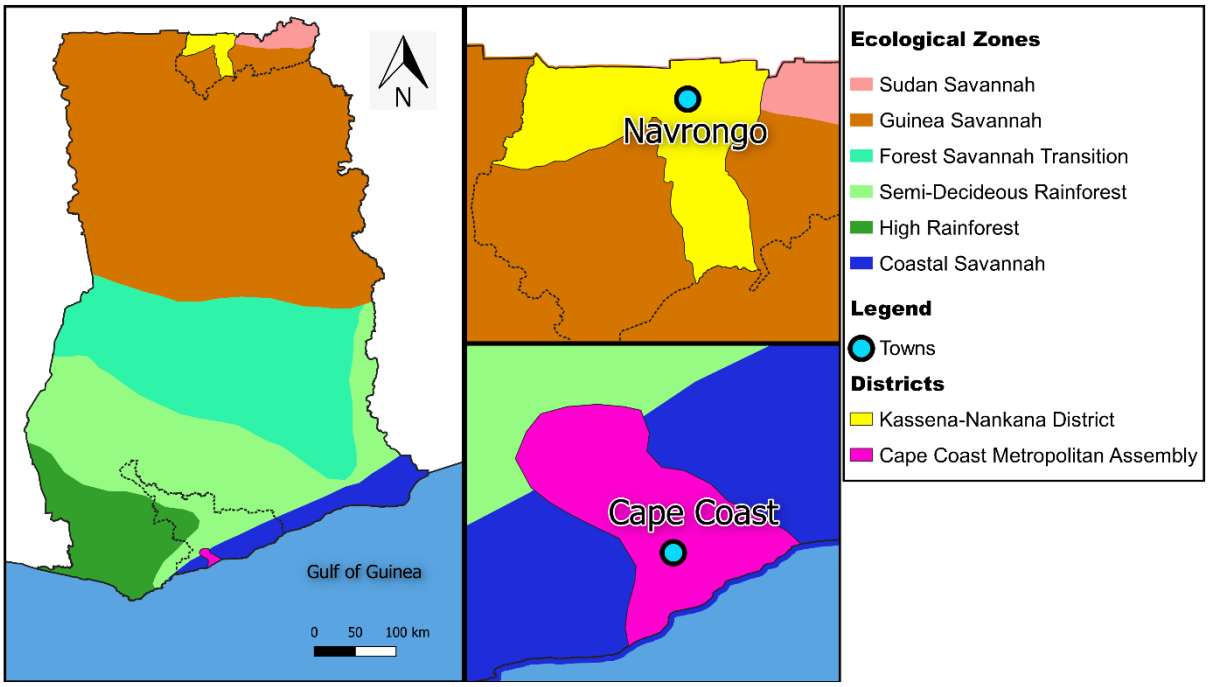


Figure 2.3: Location of the Navrongo and Cape Coast in Ghana

(Source: designed with QGIS software).

2.2.1 Malaria transmission in Cape Coast

Cape Coast is situated at the southern coastal savannah located in the central region of the country. Cape Coast is a coastal savannah area situated 165 km west of Accra (capital of Ghana) on the Gulf of Guinea. The Metropolis experiences a temperature

range of 21-36°C throughout the year, with a double-maxima rainfall between 750 mm and 1,000 mm. Cape-Coast is a humid area with mean monthly relative humidity varying between 85% and 99%. The major rainy season is between May to August and the minor rainy season between October and January. The metropolis has low to moderate perennial malaria transmission with peak season during the major rainy season (Abuaku et al., 2019)

2.2.2 Malaria Transmission in Navrongo

Navrongo is situated in the Kassena-Nankena (KND) district which is located in the Upper East region of Northern Ghana. Navrongo is situated in the northern savannah zones. The main vectors that transmit malaria in this district are *Anopheles gambiae* and *Anopheles funestus* which constitute about 94.3% of the vector population in the district (Oduro et al., 2012). The rainy season begins in the months of June to October with an average of 850mm of rain, however the rest of the months are relatively dry. Malaria transmission is fairly stable, however there is an increase in transmission intensity during the rainy season.

2.3 Malaria Control Interventions

Malaria control interventions include insecticide-treated bed nets (ITNs) and indoor residual spraying (IRS) for vector control. In 2018, 50% of the people living in Africa were protected sleeping under ITN, this saw an increase from 29% in 2010 (WHO, 2019). Furthermore, the percentage of the population with access to an ITN increased from 33% in 2010 to 57% in 2018. Households with at least one ITN for every two people increased to 72% in 2018, from 47% in 2010. However, this figure represents only a modest increase over the past 3 years, and remains far from the target of universal coverage (WHO, 2019).

The employment of indoor residual spraying (IRS) is also highly encouraged as a malaria prevention method. This involves spraying the inside walls of homes with insecticides. Globally, IRS protection declined from a peak of 5% in 2010 to 2% in 2018 across all WHO regions. The declines in IRS coverage may be due to the switch from pyrethroids (common commercial and household insecticides) to more expensive insecticides in response to increasing pyrethroid resistance (WHO, 2019).

Control of malaria also involves chemoprevention via the use of seasonal malaria chemotherapy (SMC). Seasonal malaria chemoprevention (SMC) in children is defined as the intermittent administration of full treatment courses of an antimalarial medicine during the malaria season to prevent malarial illness with the objective of maintaining therapeutic antimalarial drug concentrations in the blood throughout the period of greatest malarial risk (WHO, 2015). In pregnant women, intermittent preventive treatment with sulfadoxine-pyrimethamine (SP) is pre-requisite. Treatment of malaria involves the use of the following Artemisinin Combination therapy (ACT) regimen for a duration of three days: artemether + lumefantrine, artesunate + amodiaquine, artesunate + mefloquine, dihydroartemisinin + piperaquine, artesunate + sulfadoxine-pyrimethamine (SP) (WHO, 2015). Not all individuals are required to take these drugs, modifications in treatment regimens are done for pregnant women and HIV/AIDs patients. Pregnant women during the first trimester are recommended to take quinine + clindamycin for 7 days. Artesunate + SP is not recommended is not given to HIV/AIDS patients who are on co-trimoxazole treatment (WHO, 2015). Treatment of malaria with these is highly encouraged after confirmation of laboratory testing via microscopy and rapid diagnostic tests due to risk of antimalarial drug resistance.

Meanwhile, there has been several reports of insecticide resistance emerging and recurrence of antimalarial drugs resistance in several countries including Ghana (WHO, 2019). An ideal control measure will be the introduction of a broadly potent vaccine in the panel of available tools. Ghana together with Kenya and Malawi have been identified as one of the countries conducting RTS,S phase IV clinical trials via The Malaria Vaccine Implementation Program (MVIP) led by WHO (WHO, 2019). Currently, the Phase IV implementation trial of RTS,S/ASO1 vaccine is being conducted in Ghana, Kenya and Malawi (WHO, 2019). In Ghana, the trial is ongoing in three regions namely, Brong-Ahafo and Volta regions in the forest ecology and Central region on the coastal belt which is captured in this study.

2.5 Vaccine antigens

For many decades, antigens expressed during the various developmental stages of the *P. falciparum* life cycle have been targeted as vaccine candidates and have

undergone various developmental, pre-clinical and clinical trials. Vaccine antigens can be categorized into three groups. These include the pre- erythrocytic stage vaccine antigens, the blood stage antigens and the transmission blocking antigens. The classification is based on the target proteins expressed at the various stages of the parasite life cycle that are capable of inducing protective immune response (Richards & Beeson, 2009).

2.6 The Pre-erythrocyte stage vaccine antigens

They are termed pre-erythrocytic because they are based on antigens expressed on the surface of sporozoites. Antigens expressed during this stage are considered as ideal vaccines due to their ability to produce antibodies which prevent infection by inhibiting sporozoite motility and preventing hepatocyte invasion (Ouattara & Laurens, 2015) therefore escaping the infection and disease stage. Several experiments have been conducted using humans, non- primates and rodents where sterile immunity has been induced through immunization with whole radiation-attenuated sporozoites in multiple doses (Hoffman et al., 2002). The first study using the radiated- attenuated sporozoites in mice was conducted by Ruth Nussenzweig in 1967 (Kumar *et al*., 2006). Realistically, using whole sporozoite vaccines cannot be practical on a large scale due to issues of logistics and cost involving the harvesting of the sporozoites from the salivary glands of aseptic mosquitoes, purification, vialing, and cryopreservation in liquid nitrogen vapor (Hoffman et al., 2002). Other pre-erythrocytic vaccine antigens include Circumsporozoite protein fused to hepatitis B surface antigen (RTS,S), thrombospondin-related adhesion protein (TRAP), and *P. falciparum* cell-traversal protein for ookinetes and sporozoites (PfCelTOS) (Ouattara & Laurens, 2015).

2.7. The blood stage vaccines

These vaccines are termed disease blocking vaccines. This is due to their ability to control parasitemia through various mechanisms which include inhibiting sporozoite motility, prevention of hepatocyte invasion, targeting the merozoite ligand to inhibit erythrocyte invasion, prevention of cyto-adhesion of intra-erythrocyte targets thus avoiding pathogenesis (Ouattara & Laurens, 2015). These vaccines are based on antigens mainly expressed on the surface of the blood stage parasites. These parasites

include merozoite, intra-erythrocytic stages, or whole blood stage parasites (Richards & Beeson, 2009). The vaccine antigens expressed here include Erythrocyte-binding antigen 175 (EBA 175), Apical membrane antigen 1 (AMA 1), Reticulocyte binding homologue 5 (RH5), Recombinant Lactococcus lactis hybrid glutamate-rich protein (GMZ2), Merozoite surface protein 3 (MSP 3), Merozoite surface protein 1 (MSP 1), and Merozoite surface protein 2(MSP 2), glutamate-rich protein (GLURP). The challenges to developing anti-merozoite vaccines include the following; short time (seconds) when merozoites pass between erythrocytes and are accessible to antibodies, antigenic polymorphism, redundant invasion pathways, and the large number of parasites that need to be targeted (Duffy et al., 2020).

2.8 Transmission blocking vaccines

Vaccines produced here target antigens on the surface of the gametocyte expressed within the human host or anopheles mosquito vector. In a nutshell, they inhibit ookinete development in the mosquito midgut (Ouattara & Laurens, 2015). They are termed transmission blocking vaccines because of their ability to induce antibodies within the human host which are able to target gametocytes or induce antibodies to target gametocytes taken up by the Anopheles vector, thereby also targeting parasite proteins within the midgut of the mosquito (Beeson et al., 2012). These vaccines include *P. falciparum* surface protein 25 (Pfs25), *P. falciparum* surface protein 48 (Pfs48), *P. falciparum* surface protein 45 (Pfs45), *P. falciparum* surface protein 230 (Pfs230). Major challenges of these vaccines include achieving sufficient adaptive responses that maintain high levels of antibodies over time, as well as widespread coverage to accomplish herd immunity (Duffy et al., 2020).

2.9 The RTS,S and CSP

The RTS,S/AS01 is currently the most advanced and promising vaccine candidate with the AS01 serving as an adjuvant. This vaccine was designed based on a fragment of Circumsporozoite surface protein (CSP) from the *P. falciparum* 3D7 laboratory strain (Stoute *et al.,* 1998). The RTS,S vaccine includes a portion of central repeat region (16 NANP-NVDP) which is a B-cell immunodominant epitope and a highly polymorphic C-terminal region of CSP which engulfs $CD4^+$ and $CD8^+$

T-cell epitopes. The $CD4^+$ and $CD8^+$ T-cell epitopes are denoted as TH2R and TH3R respectively (Casares et al., 2010) (**Figure 2.4**). Hepatitis B antigen (HBsAg) is linked with CSP segment and this serves as the carrier matrix (Plassmeyer *et al*., 2009) (**Figure 2.4**).

CSP is expressed on the outer sporozoite membrane during the infective sporozoite stage of *P.falciparum*. CSP is located on Chromosome 3 of *Plasmodium falciparum*, it is made up of 397 amino acids (**Figure 2.5**). This antigen is made up of a less polymorphic N-terminal region, a less polymorphic central repeat region and highly polymorphic C-terminal region (Egan *et al*., 1993). The N-terminal region binds heparin sulfate proteoglycans (RI). The central repeat region (CRR) of CSP consists of tandem repeats of NANP as well as a small number of NVDP repeats ranging from 37 to 44 repeat units (Figure 2.5). This region is made up of immunodominant B cell sporozoite epitopes which are recognized by antibodies that block sporozoite infection of liver cell (Egan *et al*, 1993; Plassmeyer *et al*., 2009). The C-terminal region is GPI-anchored and contains a thrombospondin-like domain (RII). This region consists of two sub-regions namely TH2R and TH3R, they encode epitopes that elicit CD4+ and CD8+ T cell responses which inhibits hepatocyte invasion by sporozoites (Good *et al*., 1988). Previous studies have shown diversity occurring in the TH2R and TH3R region of CSP in the form of non-synonymous SNPs (Weedall et al., 2007). It has also been observed that the diversity observed increased as malaria transmission increased in distinct geographic locations with Africa showing the highest form of diversity (Weedall et al., 2007; Chenet et al., 2008; Ghandi et al., 2014).

The RTS,S/AS01 vaccine is currently in its Phase IV implementation trial stage in Ghana, Kenya and Malawi having been approved by the European Medicine Agency (EMA) in 2015 after the Phase III trial was completed. The EMA declared that the benefits of the vaccine outweighed the risks (European Medicine Agency, 2015).

Several studies have reported extreme genetic diversity especially at the T-cell epitopes located within the C-terminal region of *Pfcsp* (Barry et al., 2009; Gandhi et al., 2014; Jalloh et al., 2009; Zeeshan et al., 2012). Previous ancillary studies have

been conducted following up on Phase II clinical trials to assess *Pfcsp* haplotypes isolated from individuals receiving either the vaccine or the placebo to determine whether immune protection is as a result of the *Pfcsp* matching the 3D7 strain (Alloueche *et al*., 2003; Bojang *et al*., 2001; Enosse *et al*., 2006; Waitumbi *et al*., 2009). The results of such studies conducted in three sites including The Gambia, Kenya and Mozambique revealed that immune protection of the RTS,S/AS01 vaccine was not associated with having strain matching the *Pfcsp* 3D7 genotype (Alloueche *et al*., 2003; Bojang *et al*., 2001; Enosse *et al*., 2006; Waitumbi *et al*., 2009). However, such studies depended on older sequencing technologies such as Sanger sequencing and methodologies such as oligonucleotide probing available back then for genotyping of *Pfcsp* isolates. These technologies may not have been robust enough to generate enough read depth for novel or rare variant detection unlike using Next generation sequencing technologies which have the power to detect these novel variants. In addition, these studies were based on only a few hundred isolates and were not statistically powered to detect moderate effects such as the strain specific immune response of the vaccine. Meanwhile, results from an ancillary study conducted in the phase 3 trial samples in 2015 using next generation sequencing technologies such as Illumina Miseq and PacBio platforms revealed a rather strain specific immune protection.

The ancillary study conducted during the phase III clinical trials from the year 2009 to 2014 across 11 sites in 7 African countries (Burkina Faso, Gabon, Ghana, Kenya, Malawi, Mozambique, and Tanzania) involved 15,460 infants and children. The study revealed that the vaccine conferred partial efficacy against clinical and severe malaria disease. The results of the study conducted showed that only 50% immunity was conferred against vaccine strain-specific alleles and 33% protection against mismatch strains (Neafsey *et al.,* 2015).

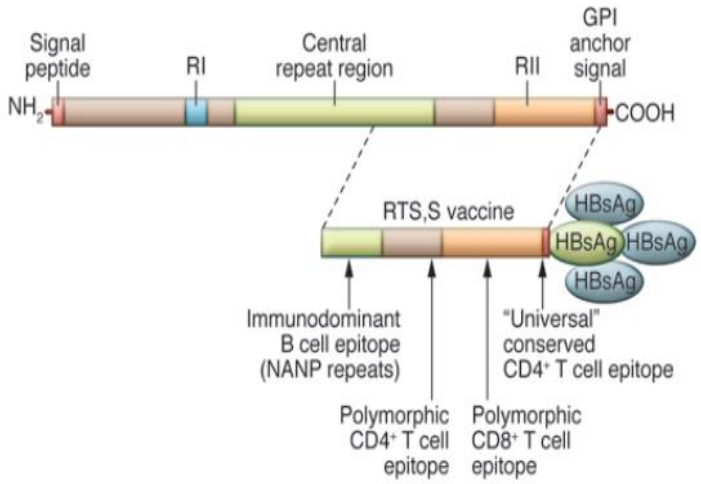


Figure 2.4: CSP gene and the RTS,S vaccine construct: Retrieved from Crompton *et al*., (2010)

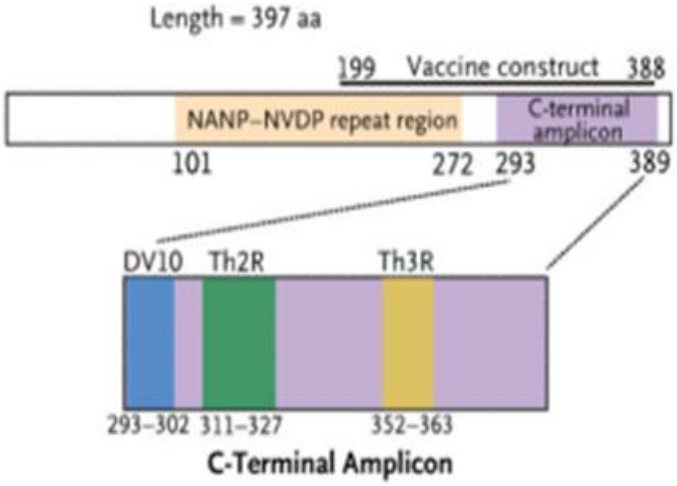


Figure 2.5: The central repeat region (CRR) and the C-terminal region of the CSP antigen: Retrieved from Neafsey *et al*., 2015

2.10 Population genetics to inform the outcome of vaccine efficacy

To better understand genes encoding protein immune targets of malaria parasites, a population genetic analysis approach is required (Amambua-Ngwa et al., 2012; Weedall & Conway, 2010). The *P. falciparum* 3D7 reference genome is a 23-megabase (MB) genome that encodes 5300 protein-coding genes and consist of 14

chromosomes (Gardner et al., 2002). The CSP gene is located on the sub-telomeric end of chromosome 3. Most genes found in the sub-telomeric ends of the chromosomes are known to be associated with antigenic variations (Gardner et al., 2002). Also, most of the genes of the *P. falciparum* 3D7 clone are reported to be highly involved in immune evasion (Gardner et al., 2002).

Different statistical methods can be used to measure the extent of diversity of malaria vaccine candidates at a particular locus (Barry *et al*., 2012). Some of these include estimation of alleles or haplotypes frequencies, haplotype diversity, multiplicity of infection, and identification of distinct population structures based on geographic location. A haplotype in this context refers to a set of alleles (SNPs) that occur together. Beneficially, population genetic analysis approach is important in detecting signatures of selection in populations (Weedall & Conway, 2010).

Natural selection allows for the adaptation of an organism in its environment. Variations in the genotype that makes an organism survive and reproduce are preserved and are maintained from generation to generation at the expense of less advantageous ones thus evolution often occurs as a consequence of this process. Selection operating on parasites differ locally owing to variables such as transmission patterns, ecology and degrees of acquired immunity in humans (Duffy *et al*., 2015). Selection could either be positive, balancing or negative. Positive directional selection is where advantageous alleles increase in frequency in the population and are maintained over time, negative selection comes as a result of the selective removal of rare alleles that are deleterious. Balancing selection is however defined as the type of selection where multiple alleles are maintained in the population (Biswas & Akey, 2006) and can be predicted using the Tajima D test (Tajima, 1989).

The Tajima's D test is a statistical test that measures the difference between the observed number of polymorphic segregating sites (S) and nucleotide diversity (π), it identifies departures from neutrality in allele frequency distributions (Tajima, 1989). Meanwhile, a positive Tajima's score >0 depicts balancing selection while a negative Tajima's D <0 indicates directional selection or clonal expansion. Previous studies

have proven that genes encoding protein immune targets of *Plasmodium falciparum* have shown signatures of balancing selection thus different alleles are being maintained within populations (Amambua-Ngwa et al., 2012; Weedall & Conway, 2010; Zhai et al., 2009). However, most alleles at antigenic loci tend to be under recent positive directional selection as well. Thus, to ascertain positive directional selection, an integrated haplotype score (iHS) using the rehh package in R is appropriate. The |iHS| measures the amount of extended haplotype homozygosity (EHH) at a given SNP along the ancestral allele relative to the derived allele (Voight et al., 2006). It is very important to determine selection at the haplotype level since alleles rising in frequency in a given population tend to be in linkage disequilibrium with adjacent alleles thus forming extended haplotypes which are maintained over time (Tariq et al., 2003).

Further analysis to determine differences in allele frequencies in terms of SNPs at a given locus between populations should be done using the Wrights FST (F-statistic test). The value of FST range from 0-1, a low Wrights FST value will show that different populations have similar allele frequencies and polymorphic regions (Barry et al., 2012). A principal component analysis (PCA) allows for the inference of population structure which lays emphasis on results of Fst. PCA is done to assess the genetic distance between isolates using genetic data between populations. The various outcomes of these statistical test will aid in assessing whether there is gene flow between Cape Coast and Navrongo.

Estimating the pattern of haplotype uniquely present in each population will enable the prediction of effectiveness of the vaccine in each population in the sense that if haplotypes in the field vary from the haplotype-based vaccine(3D7), there will be minimal protection against non-vaccine haplotypes (Barry *et al*., 2012).

An Fws (also known as inbreeding co-efficient) analysis is done to determine if the human host harbours multiple infections of *P. falciparum*. In this study referring to multiplicity of infection (MOI) within the human host. An individual can be infected by different strains of the parasite (polyclonal infections) or a single strain of the parasite (monoclonal). Studies have shown that the people living in higher malaria

transmission setting have higher polyclonal infections than people living in low transmission areas (Adjah *et al*., 2018; Peyerl-Hoffmann *et al*., 2001).

Network analysis is critical in understanding and exploring the genealogical relationships among sequences by linking haplotypes that are identical at a predefined proportion of polymorphic sites. Each node represents each haplotype which may have many edges representing the multiple connections. This analysis helps to define distinct clusters of highly related sequences but also the relationships among them (Barry *et al*., 2012).

A broadly potent malaria vaccine remains one of the most sought-after malaria-fighting intervention, but the issue of antigenic diversity continues to be a major hindrance. Natural populations of parasites of malaria are complex and dynamic, but modern molecular and analytical technologies including Next generation sequencing technologies are providing a better understanding of this parasite. It is therefore important that researchers consider using population genetics to inform malaria vaccine development from target discovery through to vaccine design and to assess the effects of clinical trials.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Research design

This study is a secondary analysis of *Plasmodium falciparum* sequenced data obtained during a cross-sectional study of *P. falciparum* malaria in Ghana specifically in Cape Coast and Navrongo sites to elucidate the population genetics of parasites in these eco-epidemiologically distinct regions.

3.2 Study sites

Cape Coast is located in the southern coastal savannah ecological zones with low to moderate perennial malaria transmission but with marked seasonal effect during the major rainy season (May to August). The estimated annual EIR in Cape Coast is fewer than 50 infective bites per person per year (Abuaku *et al*., 2019; Mensah-Brown *et al*., 2017). On the other hand, Navrongo is located at the northern Guinea savannah region with high seasonal malaria transmission during the rainy season (June to October), however there is minimal transmission in the rest of the year which are relatively dry months. The estimated annual entomological inoculation rate (EIR) of up to 157 infective bites per person per year (Kasasa *et al*., 2013; Oduro *et al*., 2012).

3.3 Study population

The study population consist of children aged 6-59 months presenting with *P. falciparum* Clinical malaria in Ghana. Clinical malaria in both sites was associated with participants infected with *P. falciparum* asexual parasitemia detected by RDT and microscopy, presence of axillary temperature ≥ 37.5℃ or history of fever during the past 24 hours. Informed consent was obtained from parent/guardian willingness to comply with the study protocol for the duration of the study.

3.4 Sampling technique

A stratified sampling technique was employed, where Cape Coast and Navrongo represents each stratum. In Cape Coast, *P. falciparum* parasites were isolated from 101 children with clinical *P. falciparum* malaria attending the Cape Coast District Hospital and aged 6-59 months. Samples were collected during the major rainy

season (May-August) in 2013. In Navrongo, *P. falciparum* parasites were isolated from 131 children with age range between 12-59 months also with clinical *P. falciparum* malaria living in the Kassena-Nankana Districts and attending the Navrongo War Memorial Hospital in the years 2010 (January to October), 2011 (January to February) and 2013 (August to October) during both dry and wet seasons.

3.5 Sample collection, DNA extraction and sequencing

3.5.1 Cape Coast

Venous blood (2-5mL) was collected from *P. falciparum* infected patients who gave consent following ethical approval of the local institutional ethical review committees. Each whole blood sample was depleted of leukocytes within 6 hours after collection and filtered through the Whatman CF11 cellulose powder filtration columns (GE Healthcare Bio-Sciences, Piscataway, NJ). DNA was extracted using the QiaAmp DNA prep kit (Qiagen, Valencia, CA) following manufacturer's protocol and confirmation of *P. falciparum* diagnosis was done by amplification using nested PCR with specific primers (Snounou *et al*., 1993).

Plasmodium falciparum whole genome sequencing of samples was done on Illumina's HiSeq 2500 platform at the Wellcome Trust Sanger Institute Hinxton, UK. Illumina sequencing libraries (200bp insert) were aligned to the reference *P. falciparum* 3D7 reference genome followed by variant calling using the GATK best practices pipeline. Each sample was genotyped for polymorphic coding SNPs across the genome ensuring a minimum of 5x paired end coverage across each variant per sample. The dominant allele was retained in the genotype file at loci with mixed reads. The data was genotyped denoting variable nucleotides as reference (0) or non-reference (1) allele. Polymorphic sites with low call rates and those in hypervariable, telomeric and repetitive sequence regions were excluded. The Whole genome sequences were deposited in the MalariaGen database.

3.5.2 Navrongo

DNA was extracted from Dried blood spot (DBS) using QIAamp DNA Investigator Kit (Qiagen, Valencia, California, United States). Approximately 1.5 cm (0.6 inch) diameter DBS circles from each filter paper were cut out into small pieces of 3 mm

diameter using a single-hole paper punch. Punched pieces from each sample were placed into 2 ml micro-centrifuge tubes from which DNA was extracted following the manufacturer's instructions except for the reagent volumes and incubation times, which were doubled to accommodate the increased amount of DBS used per sample. An average of 80 ng of DNA was obtained from the DBS extracts out of which 5 ng was used as template for selective whole genome amplification (sWGA) reaction.

sWGA products (~1 µg total DNA) were cleaned using Agencourt Ampure XP beads (Beckman Coulter) following manufacturer's instructions. Briefly, 1.8 volumes of beads per 1 volume of sample were mixed and incubated for 5 min at room temperature. After incubation, the tube containing bead/DNA mixture was placed on a magnetic rack to capture the DNA-bound beads and discarding the unbound solution. Beads were washed twice with 200 µl of 80% ethanol and the bound DNA eluted with 60 µl of EB buffer. Cleaned amplified DNA products (~ 0.5 – 1 µg DNA) were used to prepare a PCR-free Illumina library for high throughput sequencing. Standard whole genome amplified (WGA) products of the test samples were also sequenced as control to determine the extent of enrichment. Illumina paired-end sequencing libraries were constructed using the NEBNext DNA sample preparation kit (New England Biolabs) following the Illumina multiplex sample preparation protocol.

DNA libraries were also sequenced at the Wellcome Trust Sanger Institute using Illumina HiSeq 2500 instruments. Samples were sequenced using Illumina V.3 chemistry; paired-end sequencing was done with 100-base reads and an 8-base index read. 24-multiplex sample libraries were loaded to target at least 20 million reads per sample. The Illumina sequencing libraries were aligned to the reference *P. falciparum* 3D7 reference genome followed by variant calling using the GATK best practices pipeline. The dominant allele was retained in the genotype file at loci with mixed reads. The data was genotyped denoting variable nucleotides as reference (0) or non-reference (1) allele. Polymorphic sites with low call rates and those in hypervariable, telomeric and repetitive sequence regions were excluded. Whole genome sequences were then deposited in MalariaGen database.

3.6 Ethical statement

The study protocol was reviewed and approved by the Institutional Review Board of the Noguchi Memorial Institute for Medical Research, University of Ghana (056/12-13) and the Institutional Review Board of the Navrongo Health Research Centre (NHRCIRB203). Written informed consent was obtained from individuals, parents or guardians of all children before enrollment.

3.7 Sequence acquisition and pre -processing

Whole genome sequences from Navrongo and Cape Coast were mined from the MalariaGEN Pf3k Project release 5.1 (http://www.malariagen.net/projects/parasite/pf) database. Data was obtained in variant call format (VCF). Specifically, genetic variants on chromosome 3 were retrieved for both Navrongo and Cape Coast. For the VCF file of Cape Coast, all genotypes at each SNP position were mono-allelic; biallelic genotypes were then modelled using a customized Python script**.** This was based on the approach by the MalariaGEN *Plasmodium falciparum* Community Project, 2016 where genotypes were modeled using the read depth and allelic depth. Read depth refers to the total number of sequence data per position in a given sequenced genome region while allelic depth refers to the depth of coverage of each allele per individual sample.

For read depth of two alleles at a SNP position in a sample, at positions where the read depth <5, the genotype was undetermined. At all other positions where the read depth ≥ 5, the sample was determined to be heterozygous if the allelic depth of both alleles were greater or equal to 2. All other samples were homozygous for the allele seen in the majority of reads. They were either genotyped as the homozygote reference allele or homozygote alternative allele.

Data for both populations were filtered to obtain only biallelic SNPs using bcftools v1.9.

Data was quality checked through the following criteria: SNPs that had passed all VCF filters were retained. Isolates with >10% missing SNPs and SNP positions with >5% of missing data were removed using PLINK v1.9. Further, heterozygosity was calculated and 8 isolates with outlier heterozygosities seen within the Cape Coast population were excluded. After which, SNPs with a minor allele frequency (MAF) less than 1% were removed. Missing SNPs were imputed and phasing was

done using Beagle v5.1. After the steps of quality control, the Cape Coast dataset remained with 2504 SNPs out of 26156 in the chromosome 3 region and 92 out of 101 samples, while the Navrongo dataset retained 1954 out of 43199 SNPs in the chromosome 3 region and 128 out of 131 samples. The CSP gene was then extracted from chromosome 3 (221323 – 222516), SNPs retained at the CSP loci were 13 and 22 for Cape Coast and Navrongo respectively.

3.8 Population genetics analysis

To achieve the objectives outlined in this study, the following downstream analysis were done. Measures employed to estimate the extent of genetic diversity within parasite populations included the following: Estimation of minor allele frequency distribution (MAF) of *Pfcsp* alleles prior to removal of missing SNPs, estimation of within host *Pfcsp* diversity, determination of diversity indices, *Pfcsp* C-terminal network analysis, and exploration of TH2R and TH3R amino acid frequency and diversity. Secondly, to estimate the extent of genetic variation of *Pfcsp* between Cape Coast and Navrongo parasite population, a population differentiation, admixture and structure analysis was carried out. Lastly, to scan for signatures of selection of *Pfcsp* variants, a Tajima's D as well as integrated haplotype score (iHS) and Extended haplotype Homozygosity were estimated.

3.8.1 Minor allele frequency distribution

To determine the distribution of variants in both populations, the minor allele frequency distribution (MAF) for all putative SNPs (n = 90) within the *Pfcsp* loci for both Cape Coast (n = 35 SNPs) and Navrongo (n = 55 SNPs) was determined using Plink1.9. This was done prior to removal of rare alleles (MAF ≤0.01). MAF is the frequency at which the second most common allele occurs at a given SNP position in a population. The MAF is derived from total read counts for the two alleles across all samples in the population (Manske et al., 2012). The MAF was calculated and SNPs were binned into ten equally-sized MAF intervals ([0-0.01], [0.01-0.05], [0.05-0.10], [0.10-0.15], [0.15-0.20], [0.20-0.25], [0.25-0.30], [0.30-0.35], [0.40-0.45], [0.45-0.5]).

3.8.2 Within host parasite diversity estimation

The genetic diversity of parasites infecting individuals was assessed by estimating the F_{WS} which is similar to Wright's inbreeding co-efficient). The primary interest for this analysis was the within host diversity of *Pfcsp* which refers to the number of different *Pfcsp* strains contained within an individual infection. The retained variants (13 and 22 SNPs) from the 92 Cape Coast samples and 128 Navrongo samples were used for this analysis.

Fws metric estimates the heterozygosity of parasites (H_W) within the individual relative to the heterozygosity within the parasite population (H_S) using the allele counts from dataset. The F_{ws} metric calculation for each sample was done using the equation:

$$F_{ws} = 1 - H_W/H_S$$

where H_W refers to the allele frequency of each unique allele found at specific locus of the parasite sequences within the individual and H_S refers to the corresponding allele frequencies of those unique alleles within population (Auburn et al., 2012; Manske et al., 2013). Fws ranges from 0 to 1 whereby a lower range indicates lower inbreeding rates within the parasite population thus high within host diversity relative to the population. The specific threshold for Fws indices is ≥0.95 which indicates samples with clonal (single strain) infections while samples with Fws <0.95 are considered highly to have mixed strain infections accounting for within host diversity. Fws was calculated using moimix package which is an R package (Lee, 2016). The Pearson Chi square test was used to measure the statistical significance of any differences observed in the within host diversity estimates (Fws analysis) between the population pair. The test was done using R software with P values of <0.05 considered statistically significance.

3.8.3 Genetic diversity within parasite populations

The diversity of the *Pfcsp* in each population was examined by exploring the variants in the C-terminal region of the gene (909 – 1140bp). In total, 184 *Pfcsp* FASTA DNA sequences were re-constructed with the retained variants (13 SNPs) from the 92 Cape Coast isolates and 256 DNA sequences (22 SNPs) from the 128 Navrongo isolates

using a customized Python script. Individual *Pfcsp* C-terminal haplotype sequences was determined using DnaSP software (version 6.10.01) (Rozas et al., 2017).

The following metrics were then used to assess diversity of *Pfcsp* C-terminal within each parasite population using the DnaSP software (version 6.10.01): number of sequences (n), number of haplotypes (h), segregating sites (S), average number of pairwise nucleotide differences (K), nucleotide diversity (π) and haplotype diversity (Hd) (Nei & Li, 1979).

The *Pfcsp* haplotype sequences were then used to construct network based on the method described by Templeton, Crandall, and Sing (TCS) using PopArt (Leigh & Bryant, 2015). This analysis was to assess genealogical relationships between *Pfcsp* haplotypes found in Navrongo and Cape Coast (Clement et al., 2000; Templeton et al., 1992).

Further, the amino acid haplotypes within each population were explored. The *Pfcsp* DNA sequences were translated to amino acid sequences and compared to the 3D7 0304600.1 (PlasmoDB) reference strain using a customized Python script.

The prevalence of haplotypes specifically within TH2R (311-327 amino acid positions) and TH3R (352- 363 amino acid positions) epitope regions in each parasite population were determined also using a customized Python script.

Patterns of amino acid sequence at the C-terminal region between the two sites was explored by generating sequence logos online with Weblogo (Crooks et al., 2004). The weblogos generated shows which amino acid positions are conserved or polymorphic within the C-terminal regions of both parasite populations. The amino acids are denoted in the weblogo as a stack of letters and measured in bits with the height of each letter proportional to the observed frequency of the corresponding amino acid at each position and the overall height of each stack is proportional to the sequence conservation (Crooks et al., 2004).

3.8.4 Population differentiation and structure analysis

Weir and Cockerham's FST per SNP was calculated using Vcftools v0.1.5 to determine differentiation at SNP loci between Cape Coast and Navrongo parasite populations. Fst is used to determine if there are differences in allele frequencies at specific loci between the population pair. In addition, Principal component analysis (PCA) was done to assess the genetic distance between isolates using smartpca

(Cambrigde, MA, USA) in EIGENSOFT package v6.1.3 (Patterson et al., 2006). This allows for the inference of population structure between the population pair with the genetic data. To reduce bias in Fst analysis and PCA, SNPs (the 2504 Cape Coast chromosome 3 retained SNPs and the 1954 Navrongo chromosome 3 retained SNPs) with a pairwise linkage disequilibrium (LD) value of r2 >0.5, within a window size of 100bp in the entire chromosome 3 dataset using a step size of 10 were pruned out. The remaining SNPs set at Chromosome 3 shared between both populations after pruning was 516 of which 10 were *Pfcsp* SNPs. These were then used to estimate Fst, PCA and glpca analysis. PCs were computed with the number of outlier removal iterations set at 10 while maintaining other parameters. Further, to determine admixture between the population pair, glpca analysis pipeline in R was done. This allows the inference of ancestral relationship or transmission connectivity between the two populations.

3.8.5 Signatures of selection

To test the hypothesis proposed by Tajima (Tajima., 1989) that all mutations are neutrally selected, a Tajima's D statistical test was done in sliding windows of size 100bp and step size of 10 with *Pfcsp* monoclonal samples (66 in Cape Coast and 65 in Navrongo) using Vcftools v0.1.5. Tajima's D test compares the average pairwise differences (pi) and the total number (S) of segregating sites. Negative values indicate directional or purifying selection while positive values indicate balancing selection.

Further, to determine loci likely to be under recent positive selection in Cape Coast and Navrongo populations, the standardised integrated haplotype score (|iHS|) for each SNP with a minor allele frequency of >0.05 in chromosome 3 (358 out of the 2504 and 608 out of the 1954 remaining SNPs from Cape Coast and Navrongo respectively) was calculated (Voight *et al*, 2006). Again, for the purpose of this analysis, the Fws metric was used to estimate these monoclonal chromosome 3 isolates with the retained variants within the Chromosome 3 region (2504 in Cape Coast and 1954 SNPs in Navrongo). The |iHS| measures the amount of extended haplotype homozygosity (EHH) at a given SNP along the ancestral allele relative to the derived allele (Voight *et al*., 2006). The reference and alternate alleles were

depicted as the ancestral and derived alleles respectively. This was done in R using the rehh package v2.0.4 (Gautier & Vitalis, 2012). Genomic regions under positive selection were identified as those with multiple SNPs having |iHS| values >3. For significant core alleles at focal SNPs which were positively selected, the extended haplotype homozygosity (EHH) for both the reference and alternate alleles were calculated and plotted to determine the decay of EHH at increasing distances from the focal SNP loci (Sabeti *et al*., 2002). EHH is a used to identify extended haplotypes which comes as a result of a new beneficial variant rising rapidly in high frequencies in recent generations and in linkage disequilibrium (LD) with surrounding variants where recombination events or novel mutations have not caused LD to decay (Vitti et al., 2013). Bifurcation plots were generated to visualize the decay of the EHH of a core allele of a focal SNP at increasing distances (Sabeti *et al*., 2002). These plots were generated using rehh package version 2.0.4 in R.

CHAPTER FOUR

RESULTS

4.1 Determination of the extent of genetic variation of *Pfcsp* within Cape Coast and Navrongo parasite populations.

4.1.1. Minor Allele Frequency Distribution

Prior to removal of rare alleles, which is MAF ≤0.01, a total of 90 SNPs within the *Pfcsp* were analyzed for minor allele frequency distribution. The allele frequency distribution of all putative SNPs within the *Pfcsp* loci ranged between 0.001-0.45 in Navrongo and 0.001-0.40 in Cape-Coast **(Table 4.1/Appendice I**). The Navrongo isolates had more variable loci (55 SNPs) than the Cape Coast (35 SNPs) isolates. The allele frequency spectrum was dominated by rare alleles (MAF = ≤0.01) in the two populations at 62.9% (22/35) and 61.8% (34/55) and for Cape Coast and Navrongo respectively signifying a recent population expansion. In addition, low-frequency variants 10% (6/55) and 7.2% (4/55) were observed in Navrongo whilst 20% (7/35) and 5.7% (2/35) were observed in Cape Coast ranging from 0.01-0.05 and 0.05-0.10. However, the remaining alleles were observed in moderate to high MAF in both populations (**Table 4.1**) implying some underlining evolutionary events. Table 4.1 provides the minor allele frequency distribution of a total of 90 SNPs set within the *Pfcsp* loci in samples from both Cape Coast (n= 35) and Navrongo (n =55). Each column shows the number of SNPs in each bin.

Table 4.1: Minor allele distribution

Population	Bin1	Bin2	Bin3	Bin4	Bin5	Bin6	Bin7	Bin 8	Bin9	Bin 10
	0.0-0.01	**0.01-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.45-0.45**
Cape Coast	22	7	2	1	2	0	0	0	1	0
Navrongo	34	6	4	6	1	1	0	0	2	1

4.1.2 Within host genetic diversity

In Cape Coast, 66 samples (n = 66/92) had an Fws value of ≥0.95 indicating monoclonal infections (single strain infections), this accounts for 71.7% of the population and 26 isolates (n = 26/92) had Fws value of <0.95 indicating within host diversity (mixed strain infections) also making 28.3% of the population (**Table 4.2, Figure 4.1**).

In Navrongo 65 samples (n = 65/128) had Fws value of ≥0.95 accounting for 50.8% of the population and 63 isolates (n = 63/128) had Fws <0.95 accounting for 49.2% of the Navrongo population (**Table 4.2**).

Navrongo isolates have significantly higher within host diversity compared to Cape Coast ($p < 0.0001$, Pearson's Chi square correlation co-efficient). No significant difference was observed within parasites with clonal infections between the two populations ($p = 0.9$, Pearson's Chi square correlation co-efficient).

Table 4.2: Number of monoclonal and polyclonal samples in Cape Coast and Navrongo

	Cape Coast	Navrongo
Monoclonal samples (Fws ≥0.95)	66	65
Polyclonal samples (Fws <0.95)	26	63

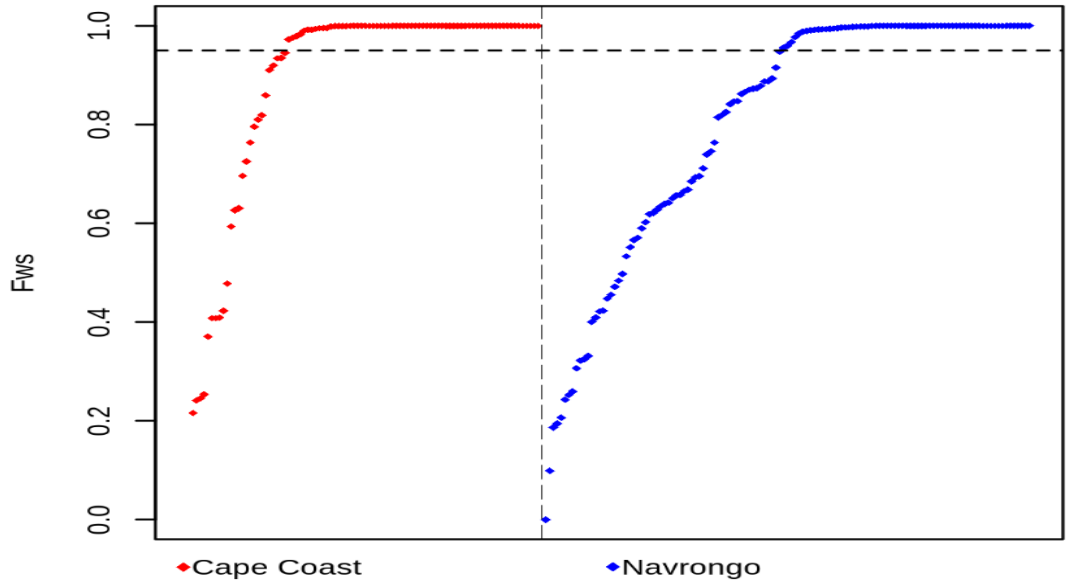


Figure 4.1: **The distribution of Fws estimates in samples from Cape Coast and Navrongo:** Cape Coast is highlighted in red and Navrongo is highlighted in blue.

4.1.3 Genetic diversity of *Pfcsp* C-terminal haplotypes

To assess the extent of genetic diversity and similarity within and between the two populations, the diversity in the C-terminal region of *Pfcsp* (231bp) from 440 DNA sequences from Cape Coast (n = 184) and Navrongo (n = 256) was explored and summarized in a Templeton, Crandall, and Sing (TCS) network **(Figure 4.2).**

In total, 66 haplotypes were observed from the 440 *Pfcsp* sequences obtained from both Navrongo and Cape Coast parasite populations **(Figure 4.2).** The haplotypes have been denoted as 3D7, Hap 2 up to Hap 66 in the network **(Figure 4.2, Appendice II).** Of these, 15 and 53 haplotypes were defined from the 184 Cape Coast and 256 Navrongo DNA sequences respectively. The 3D7 vaccine haplotype and 1 non-vaccine haplotype (denoted as "Hap 10") were shared by both populations **(Figure 4.2, Appendice II).**

The 3D7 haplotype represented only 5.9% (n= 15/256) haplotypes in Navrongo and 45.7% (n= 84/184) in Cape Coast respectively **(Appendice II).** Also Hap 10

represented 0.4% (n =1/256) in Navrongo and 6.0% of haplotypes in Cape Coast (n=11/184)**.** The 3D7 haplotype was found to be the most circulating CSP terminal haplotype (45.7%) in isolates from Cape Coast whilst the most frequent haplotype in the Navrongo isolates, "Hap 16" was found in 20.3% (52/256) of parasites **(Appendice II).**

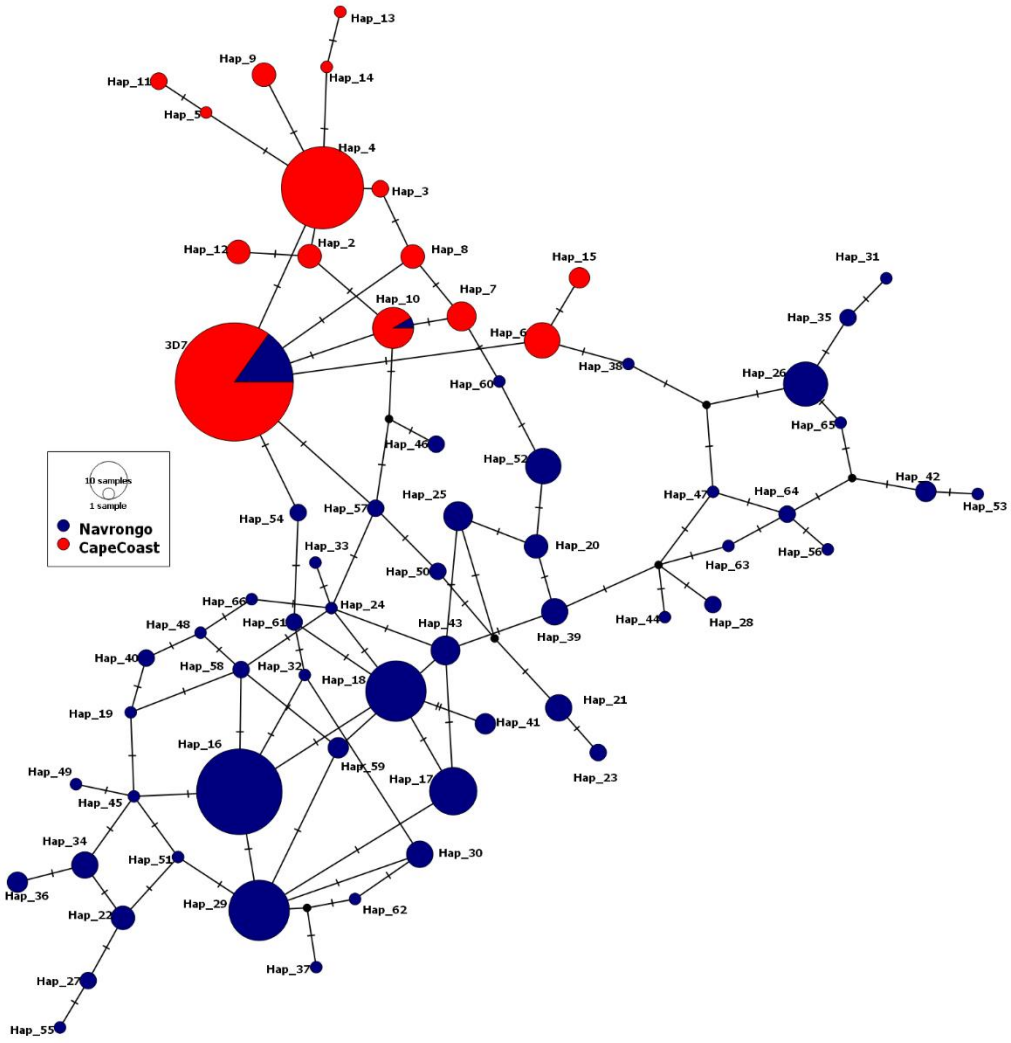


Figure 4.2: Templeton, Crandall, and Sing (TCS) network: Circles represent each *Pfcsp* C-terminal haplotype, and circles are scaled according to the frequency with which the haplotype was observed. The hatch marks on the edges shows the number of polymorphisms between haplotypes. Each haplotype is denoted as "Hap" with the Vaccine strain 3D7 (0304600.1, PlasmoDB) denoted as "3D7". The color blue

represents haplotypes found in Navrongo and the color red represents haplotypes found in Cape Coast.

The diversity indices analyzed revealed DNA sequence polymorphisms in the *Pfcsp* C-terminal is generally higher in Navrongo than in Cape Coast (Table 4.3). Navrongo had a higher number of nucleotide polymorphisms (K= 3.761) and segregating sites (S =16) compared to Cape Coast (K= 1.148, S = 8). Analysis of π revealed that nucleotide diversity was significantly higher in the Navrongo isolates (π=0.016 ± 0.0007) than in isolates from Cape Coast (π = 0.005 ± 0.0004). Essentially, haplotype diversity was significantly higher in Navrongo (Hd = 0.925 ± 0.009) in comparison with Cape Coast (0.718 ± 0.026) parasite isolates.

Table 4.3: **Diversity indices of *Pfcsp* C-terminal region**

Population	n	h	S	K	π±S.D	Hd± S.D
Cape Coast	184	15	8	1.15	0.005 ± 0.0004	0.718 ± 0.026
Navrongo	256	53	16	3.76	0.016 ± 0.0007	0.925 ± 0.009

The following are what the indices indicate; n= number of sequences, h = number of unique haplotypes, S= number of segregating sites, K= average number of pairwise nucleotide differences, π = nucleotide diversity, Hd = haplotype diversity

4.1.4 TH2R and TH3R amino acid haplotype diversity

The TH2R (311-327aa) and TH3R (352-363aa) sites have been found to be very polymorphic in both populations than the remaining amino acid sequence in the C-terminal region. In general, non-synonymous mutations predominated all the isolates in both TH2R and TH3R epitope regions with implications for cross-protection.

Of the 92 (184 amino acid haplotypes) and 128 isolates (256 amino acid haplotypes) from Cape Coast and Navrongo, there were 8 and 27 non-vaccine TH2R amino acid haplotypes in Cape Coast and Navrongo respectively **(Appendice III).** There were 2 and 10 non-vaccine TH3R amino acid haplotypes in Cape Coast and Navrongo

respectively with 1 non-vaccine amino acid haplotype (NKPKDELNYAND) shared between the two populations **(Appendice IV).**

The frequency of 3D7 TH2R vaccine haplotype (PSDKHIKEYLNKIQNSL) within both populations were 56.8% and 7.8% in Cape Coast and Navrongo respectively **(Figure 4.3A).**

The frequency of 3D7 TH3R vaccine amino acid haplotype (NKPKDELDYAND) was 79.5% and 18.7% in Cape Coast and Navrongo respectively **(Figure 4.3B**). The amino acid differences observed between 3D7 reference 0304600.1 (PlasmoDB) and the Ghanaian isolates ranged between 1 - 6 in both epitope regions.

Sequence logos were generated to assess patterns and conservation of these amino acid polymorphisms within the CS epitopes **(Figure 4.3C and 4.3D).** The sequence conservation is measured in bits. Lower values represent polymorphism at that amino acid position while higher values represent conservation. The TH2R and TH3R sites appear to be very polymorphic in both populations meanwhile the rest of the amino acids in both populations appear to be more conserved. The regions outside the TH2R and TH3R epitopes remained highly conserved with only two polymorphic sites (298 and 301) in both populations. The sequence logos indicate that the two geographically distinct Ghanaian parasite populations had varying levels of polymorphisms at similar positions with the same types of amino acids that populate these sites appearing to be conserved between them. In Cape Coast, four sites within the TH2R epitope regions namely 314, 321, 324, 327 and two sites within the TH3R epitope regions were polymorphic namely 352 and 359. In Navrongo, five sites within the TH2R epitopes; 314, 317, 321, 324, 327 and five sites in TH3R epitopes; 352, 354, 356, 357 and 359 were also polymorphic. Between the two parasite populations, there were mutual polymorphic amino acid positions namely 314, 321, 324 and 327 in the TH2R epitope and 352 and 359 in the TH3R epitope region. Overall, there were similar non-3D7 alleles at particular amino acids positions in the TH2R and TH3R epitopes in both parasite populations namely K314Q, L327I, and D359N **(Figure 4.3C, 4.3D and Appendice V)** present at different frequencies. However, the frequency of K314Q was similar in both parasite populations **(Appendice V).**

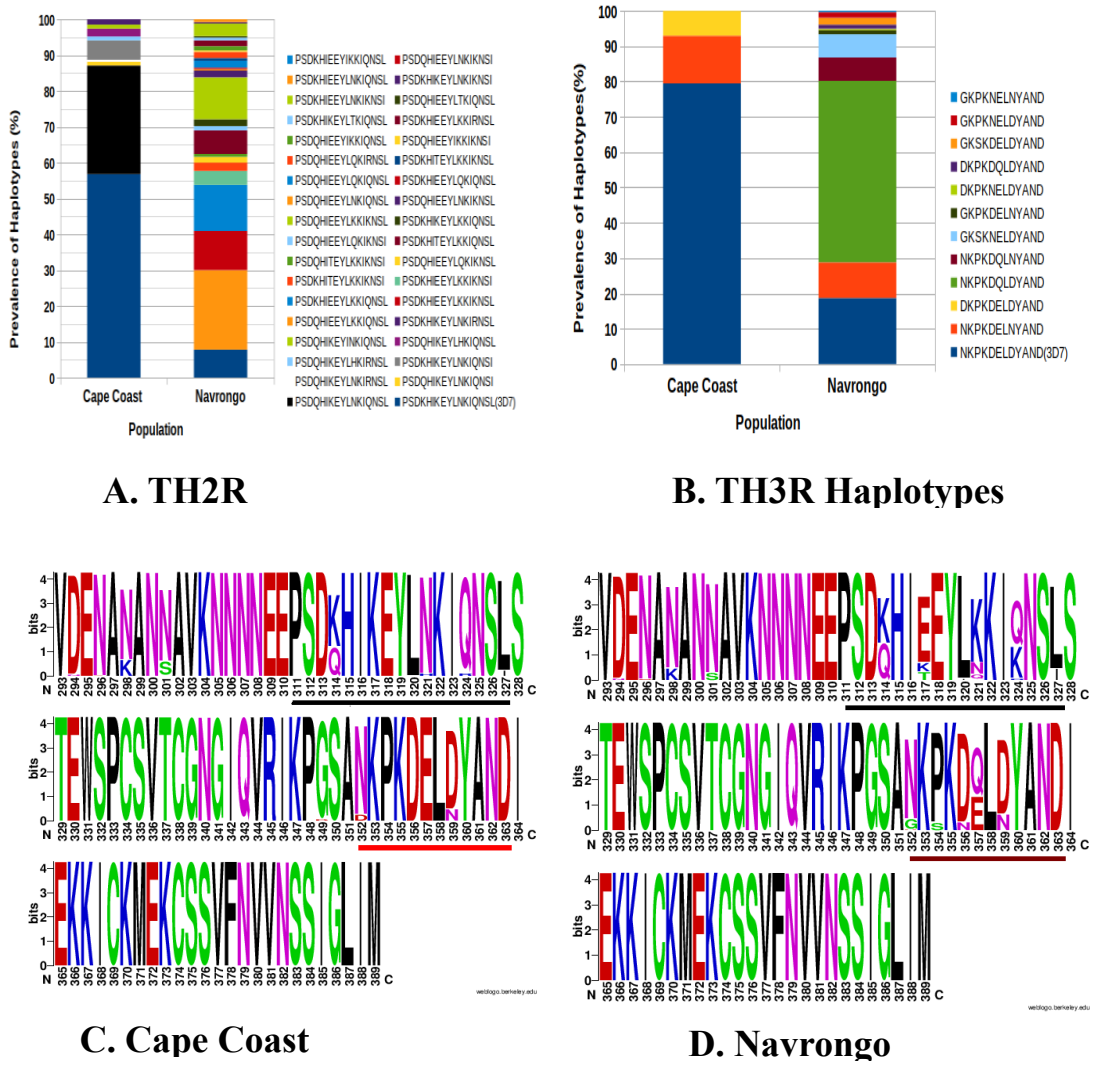


Figure 4.3: Amino acid sequence conservation, polymorphisms, and haplotypes within the C-terminal region of the Circumsporozoite Protein: (A and B) shows the percentage of isolates sharing specific amino acid haplotypes within the TH2R (311-327aa) and TH3R (352-363aa) epitope regions. The haplotypes are represented by colored columns in the bar graph. The proportion of samples in each population having the 3D7 haplotype are represented in the first blue colored column from the bottom. The proportion of samples having the non-vaccine haplotypes are shown in the rest of the colored column bars. (C and D) These are weblogos showing amino acid sequence conservation and polymorphisms of the C-terminal region of the Circumsporozoite Protein from Cape Coast (panel C) and Navrongo (panel D) respectively. The TH2R region and TH3R region are underlined in black and red in both Panels.

4.2 Determination of spatial variations of Pfcsp between Cape Coast and Navrongo parasite populations

4.2.1 Population differentiation and Structure

Weir and Cockerham's Fst calculated at SNP loci between Cape Coast and Navrongo populations showed an Fst <0.05 which signifies minimal population genetic differentiation which confirms the lack of population substructuring of Cape Coast and Navrongo. *Pfcsp* isolates observed through principal component analysis showing isolates from Cape Coast and Navrongo did not form distinct clusters. In all 10 PCs were computed with the 5 outlier samples removed from 92 in Cape Coast and 128 in Navrongo **(Figure 4.4).**

Similarly, glpca analysis revealed Navrongo *Plasmodium falciparum* isolates exhibit admixture with Cape Coast thus sharing a common genetic ancestry inferring gene flow between the two populations **(Figure 4.5).**

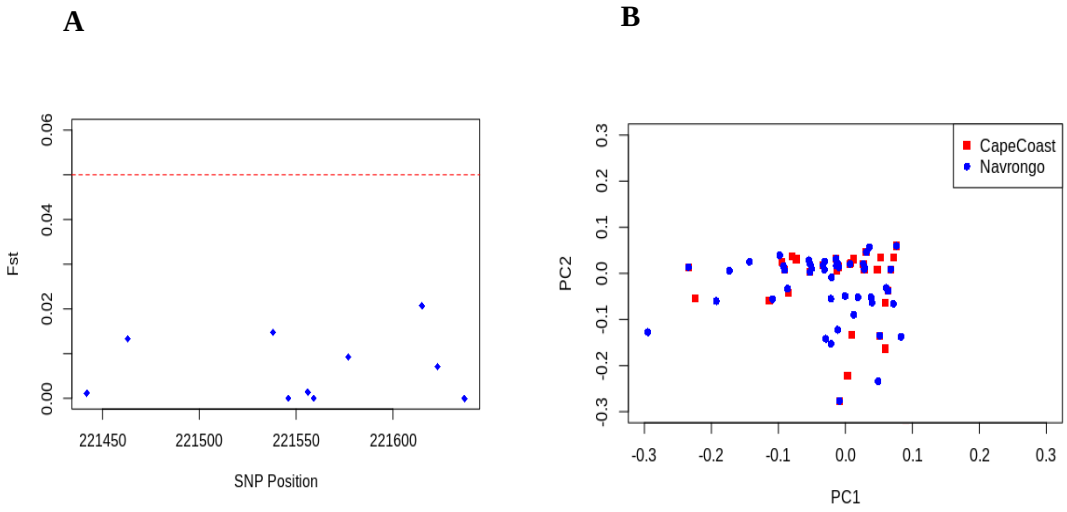


Figure 4.4: Weir and Cockerham's Fst (A) and (B) Principal Component Analysis plot*:* (A) shows Weir and Cockerham's Fst calculated at SNP loci between Cape Coast (n=92) and Navrongo (n=128) population samples for 10 SNPs. The red line shows the borderline of 0.05 which signifies moderate population differentiation. (B) shows first (PC1) and second principal component (PC2) of samples in both Cape Coast (n= 83) and Navrongo (n= 123) population after outlier samples were removed.

:

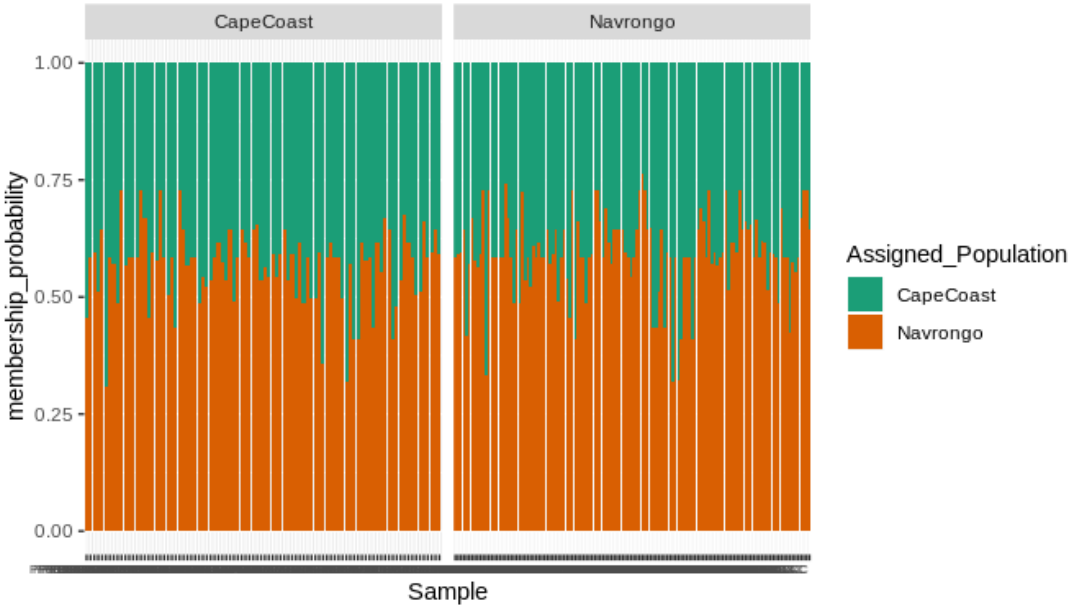


Figure 4.5: A composite barplot showing admixture between Cape Coast and Navrongo *Pfcsp* populations

4.3 Evaluation of evidence of selection of *Pfcsp* within Cape Coast and Navrongo parasite populations.

4.3.1 Evidence of Selection within populations

Tajima's D values were greater than zero in the TH2R and TH3R epitope regions of the C-terminal loci of *Pfcsp* (221,422-221,583) for the population of monoclonal *Pfcsp* isolates from Navrongo (65 monoclonal samples) **(Figure 4.6A)** suggesting balancing selection. However, a Tajima's D < 0 is seen in the Cape Coast (66 monoclonal samples) population at these loci revealing likely directional selection or clonal expansion in the population **(Figure 4.6A)**. Evidence of signatures of positive directional selection in chromosome 3 using the standardized integrated haplotype score (|iHS|) plotted as -log10 (P-value) for monoclonal isolates showed an |iHS| > 3 at SNP locus 221554 in chromosome 3 in Navrongo **(Figure 4.6C).** This SNP position is within the segment coding for the TH2R epitope suggesting that SNPs at this locus are under recent positive selection in the Navrongo population **(Figure 4.6C).** The extended haplotype homozygosity revealed extended haplotypes from this focal SNP (221554) at increasing distances from both sides of the focal SNP

however no long range haplotypes were formed **(Figure 4.7).** Evidence of positive directional selection was observed at *Pfcsp* loci only in Navrongo however there was no evidence of selection observed in the *Pfcsp* loci in Cape Coast **(Figure 4.6B & 4.6C)**. The EHH plot shows extended haplotypes from the focal SNP locus (221554) at increasing distances from both sides of the focal SNP however no long-range haplotypes were formed **(Figure 4.7A)**. Also, bifurcation diagrams in figure 4.9B shows the visualization of the breakdown of these extended haplotypes from increasing distances in Navrongo parasite population.

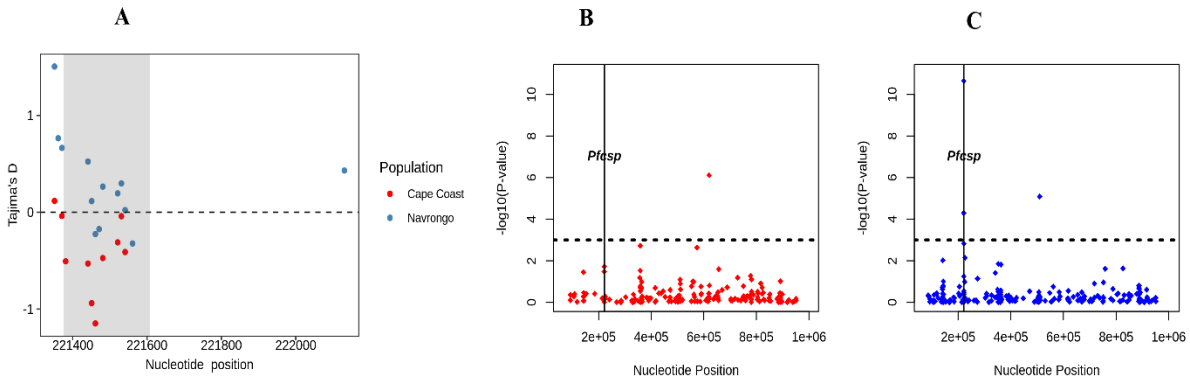


Figure 4.6: Tajima's D and standardized integrated haplotype score (|iHS) plot: (A) Tajima's D plot of monoclonal *Pfcsp* samples in Cape Coast and Navrongo. Region highlighted in grey represents SNPs located in the C-terminal region (221,422-221,583) of *Pfcsp*. (B&C) Evidence of signatures of positive directional selection on chromosome 3 using the standardized integrated haplotype score (|iHS|) plotted as -log10 (P-value) in both Cape Coast (B) and Navrongo (C) respectively. SNPs on chromosome 3 are identified in red in Cape Coast (B) and blue in Navrongo (C). Horizontal lines indicate the threshold for high-scoring SNPs with a standardised |iHS| >3. Vertical lines indicate the positions of *Pfcsp* in both Cape Coast and Navrongo, respectively.

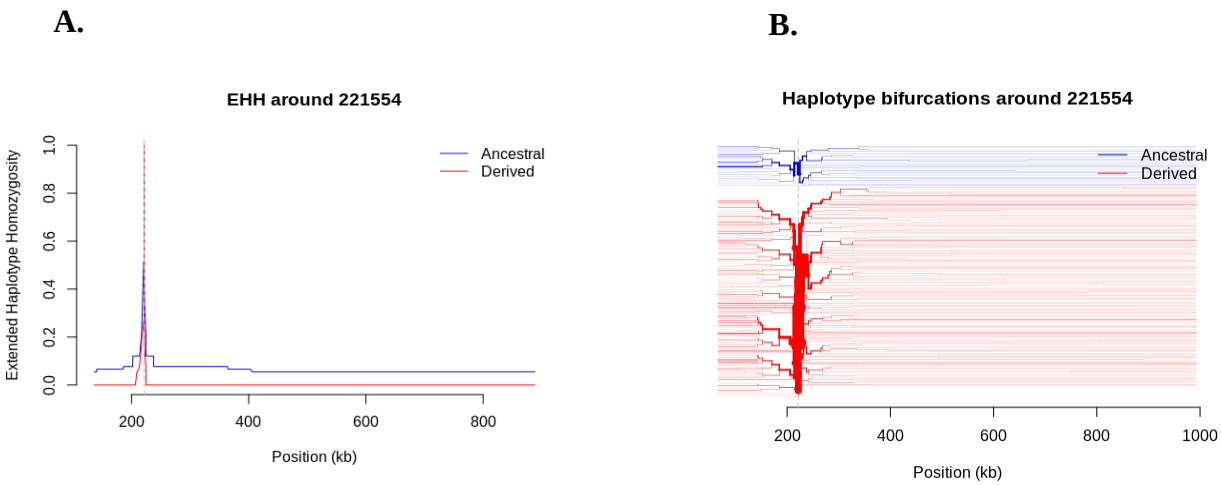


Figure 4.7: Extended haplotype homozygosity (EHH) plot and Bifurcation diagram. A) EHH plot showing extended haplotypes from a focal SNP locus (221,554) in Navrongo parasite population. B) Bifurcation diagrams showing the breakdown of these extended haplotypes from increasing distances.

CHAPTER FIVE

DISCUSSION

Given that the RTS,S/AS01 vaccine has been designed based on only the sequence of *Pfcsp* gene of the *P. falciparum* 3D7 parasite strain (Pringle et al., 2018) with the evidence of strain-specific immunity reported from a previous study (Neafsey et al., 2015), considering the extent of diversity in circulating parasites from different transmission settings in a country can provide us with insight into how well RTS,S/AS01 may perform if implemented on a large scale in that country. Further, adequate knowledge of population structure and within-host diversity of the malaria parasite in distinct geographic locations with varying malaria transmission patterns is critical in identifying region-specific malaria control measures (Adjah et al., 2018; Karunamoorthi, 2014).

Analysis of the *Pfcsp* sequence data generated from whole genome sequencing of parasite isolates from Cape Coast in the coastal savanna region and Navrongo in the Guinea savannah zone of Ghana revealed a significantly higher within host malaria *Pfcsp* diversity in Navrongo with 49.2% of infections having Fws <0.95 than in Cape Coast where only 28.3% of infections had Fws <0.95 Coast ($p < 0.0001$, Pearson's Chi square correlation co-efficient) **(Table 4.2/Figure 4.1).** This could be attributed to higher malaria transmission patterns and EIR (157) in Navrongo as compared to Cape Coast (EIR = 50). This is consistent with reports of increased genetic diversity in high transmission areas resulting in individuals harbouring polyclonal infections than people living in low transmission areas where infections are mostly monoclonal infections (Abukari et al., 2019). This marked difference in within-host diversity is essentially noteworthy for future region-specific vaccine designs and for local intervention strategies. The abundance of rare alleles shown in both Cape Coast and Navrongo contributes to the parasite population. In this light, potentially over time, rare alleles attracting less 'immune attention' would be expected to progressively proliferate while dominant alleles are cleared by the immunity they induce (Ochola-Oyier et al., 2016).

A previous study laid emphasis on the importance of assessing the diversity of C-terminal *Pfcsp* to better understand how genetic diversity at this region affects

RTS,S/AS01 efficacy (Pringle et al., 2018). Thus, to estimate the extent of diversity, this study explored the polymorphism in the C-terminal region. Analyses of genetic sequence revealed that C-terminal region of *Pfcsp* is very diverse **(Figure 4.2)** which is in agreement with findings from previous studies (Barry et al., 2009; Pringle et al., 2018). A lower frequency of the vaccine *Pfcsp* 3D7 haplotype in Navrongo (3.6%, n= 16/440) compared to Cape Coast (19.1%, n= 84/440) was also observed **(Figure 4.2, Appendice II).** Adaptive immunity from the RTS,S/AS01 vaccine is partly mediated by T cells epitopes (TH2R and TH3R) localized in the C-terminal region of CSP (Good et al., 1988; Riley et al., 1990). A high genetic diversity specifically in the TH2R and TH3R amino acid epitopes was observed in the two sites **(Figure 4.3).** This diversity in these epitopes is consistent with findings from previous studies (Bailey et al 2012; Escalante et al., 2002; Gandhi et al., 2014; Jalloh et al., 2009; Kumkhaek et al., 2004; Zeeshan *et al.,* 2012). Such polymorphisms at the T-cell epitopes have been suggested to be an immune evasion mechanism, in response to host T-cell immune responses (Zeeshan et al., 2012) or selection in the mosquito host during the malaria transmission cycle (Bailey et al., 2012). Another hypothesis drawn from previous study suggest that polymorphism at the T cell epitopes could also be driven by an evolutionary response to intermolecular interactions at the surface of CSP (Aragam et al., 2013). The high degree of location specific *Pfcsp* diversity observed in Ghana might result into differences in vaccine efficacy, potentially reducing RTS,S/AS01 vaccine effectiveness, particularly in Navrongo where 3D7 haplotype was rare. Monitoring differential vaccine efficacy by *Pfcsp* haplotype during the RTS,S/AS01 implementation programs will be valuable for such a high transmission area, where post-vaccination expansion of non-haplotype alleles in the population is likely to be observed leading to reduced vaccine efficacy.

The conformation of epitopes may therefore be altered by non-synonymous substitutions consequently compromising the host immunity. This analysis revealed high levels of non-synonymous substitutions in the immunogenic TH2R and TH3R epitopes in both sites **(Figure 4.3).** There were more amino acid substitutions observed in Navrongo parasite populations than in Cape Coast reflecting in the lower frequency of 3D7 vaccine haplotype observed in the network analysis in the

Navrongo parasite population. This might have implications for the vaccine efficacy during the implementation trials in comparable high malaria burden populations in Ghana.

An ancillary study conducted following the phase III RTS,S/AS01 vaccine clinical trials showed that the immunity of the RTS,S vaccine waned significantly for parasites not matching 3D7 in the *Pfcsp* C-terminal region and that the efficacy of the vaccine decreased as the number of amino acid differences increases (Neafsey et al., 2015). In this study, amino acid differences within the TH2R and TH3R epitope regions were seen to range from 1 to 6 at each epitope in both parasite populations (**Figure 4.3/appendice III & IV**). These amino acid differences observed is similar to amino acid haplotype differences observed in the C-terminal region in Zambian and DRC population, ranging from 2 to 10 (Pringle et al., 2018). Therefore, the Ghanaian isolates differing from 3D7 at six amino acids positions in the vaccine C-terminal epitopes is potentially significant to the immune response and thus the efficacy of the RTS,S vaccine in Ghana. In addition, there were more amino acid substitutions in the Navrongo parasite population than in the Cape Coast parasite population **(Figure 4.3),** which is consistent with the lower frequency of the vaccine haplotype observed in the network analysis for the Navrongo parasite population **(Figure 4.2).** This will have implications for vaccine efficacy in Navrongo whereby RTS,S vaccine could likely be less efficacious in Navrongo than in Cape Coast.

Despite this high level of genetic diversity resulting from non-synonymous nucleotide and amino acid substitutions observed, there remained a shared gene pool between the two sites that resulted in a largely homogeneous parasite population **(Figure 4.4/Figure 4.5).** Weir and Cockerham Fst with *Pfcsp* sequence data from both Cape Coast and Navrongo revealed minimal genetic differentiation (<0.05) indicating gene flow between these two populations (Duffy *et al.,* 2017). This was further confirmed with PCA revealing a lack of population structure and glpca analysis showing population admixture. Thus, these two populations are not genetically isolated. Previous studies have also indicated that human population mixing is likely to cause gene flow of *P. falciparum* parasites (Amambua-Ngwa et al., 2019; Henden et *al*., 2018). Despite the ecological and epidemiological diversity

between the two study sites**,** human movement from one site to the other could then account for *Pfcsp* gene flow since the two sites are within the same country and this could result in the spread of any emerging vaccine-resistant parasite. However, high levels of genetic recombination in high transmission areas may explain the observed differences in the haplotype diversity in Navrongo in comparison with Cape Coast (Yuan et al., 2013) despite the observed gene flow between the two sites (**Figure 4.2**).

Factors such as malaria transmission patterns, ecology, and degrees of acquired immunity in humans impact on the adaptation of *P. falciparum* in a population (Amambua-Ngwa *et al*., 2012; Duffy *et al.*, 2015; Mbengue *et al.*, 2015). Evaluation of signatures of selection within each parasite population showed evolutionary selective pressure on *Pfcsp*.

A negative Tajima's D was observed in the Cape Coast isolates **(Figure 4.6A).** This may imply that the population has become monoclonal and so there is more of a single haplotype circulating which could be as a result of recent population expansion of the 3D7 major haplotype in an area with moderate malaria transmission after over 15 years of enhanced nationwide malaria control interventions (chemotherapy and vector control). This result is in line with finding from Thiès, Senegal where increased deployment of malaria control interventions resulted in an increase in the frequency of clonal strains and decrease in the probability of multiple infections (Daniels et al., 2013).

Meanwhile, evidence of both recent positive and balancing selection was observed in the Navrongo parasite isolates **(Figure 4.6A and Figure 4.6C).** Majority of alleles present at the C-terminal region in the Navrongo parasite population had a positive Tajima's D score and were highly polymorphic, which is likely due to balancing selection in response to host immune pressure on this immunogenic epitope (Bailey et al*.,* 2012; Tetteh et al., 2009., Weedall & Conway, 2010). This is consistent with findings from previous studies conducted where evidence of balancing selection had been reported in CSP in Malawi (Bailey et al., 2012).

In addition, balancing selection has been reported in other vaccine antigen candidates such as in the domain I epitope of Pf38 gene (found on the merozoite surface) in

Papau New Guinea and the Gambia (Reeder et al., 2011) and also in the extracellular domains of AMA1, a target of allele-specific immune responses (Polley & Conway, 2001). However, genetic drift among loci attributable to sampling bias in Navrongo (sampling in the Navrongo population was not done in one season and in one year) may contribute to the balancing selection observed. Evidence of recent positive directional selection at the T-cell epitope loci in the Navrongo parasite population was observed (iHS >3). This could be due to the addition of new and useful alleles to the already existing repertoire of alleles which are being maintained by balancing selection in the population (Duffy *et al*., 2015). In addition, taking into account factor such as differences in the EIR and eco-epidemiological background of these two populations, the intensity of transmission at the two ecologically distinct sites could potentially account for differences in selection signals.

It is worth noting that, there are some limitations to this study. Samples analyzed in this study were non-randomly selected from the population therefore may have some limitations and bias the inferences that can be drawn from the isolates. Notably, the Navrongo and Cape Coast isolates were unequally deposited into the Pf3K database at different time periods, leading to a geographically biased set of sequences which could likely over-represent genotypes from a small number of geographic foci while under-representing large higher frequency SNPs. Further, as sample sizes and geographic distributions are continuously updated and expanded, conclusions drawn from sequences obtained from any given sequence repository are subject to change. In contrast to sequences from Cape Coast which were obtained from the same sampling periods, Navrongo sequences reflect different time periods, which may preclude samples from these two regions from being optimally comparable. However, despite these inherent limitations, the sequence analysis elaborated here is a powerful approach capable of elucidating global patterns in pathogen population genetic diversity and for monitoring the effect and efficacy of interventions. A wider geographic and temporal analysis will further reveal the full extent of the diversity of *Pfcsp* locally and across Africa. The study provides some important baseline data for comparison to parasites sampled after implementation of RTS,S. Therefore, assessing the diversity of C-terminal *Pfcsp* should be a major component of the RTS,S/AS01 Malaria Vaccine Implementation Programme.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusion

- The extent of genetic polymorphisms of *Pfcsp* observed in the study sites would likely implicate an allele-specific immune response in children during the phase IV implementation trials being held in Ghana.
- The high degree of location specific *Pfcsp* diversity observed in Ghana might result into differences in vaccine efficacy depending on the eco-epidemiological region, potentially reducing RTS,S/AS01 vaccine effectiveness.
    - Cape Coast could potentially have a higher RTS,S/AS01 vaccine efficacy than Navrongo because more of the 3D7 haplotypes were seen in Cape Coast whilst few 3D7 haplotypes were observed in Navrongo.
    - Polyclonal strains of *Pfcsp* observed suggest the need for polyvalent malaria vaccine to provide cross-protective vaccine immunity.
- The intensity of transmission at these two ecologically distinct sites could account for differences in selection signals observed in Cape Coast and Navrongo.
- Consequentially, vaccine efficacy during the ongoing Phase IV implementation trial in Ghana may be dependent on the extent of homology between the amino acid haplotypes circulating in the natural parasite population and the 3D7 vaccine haplotypes. This might result in directional selective advantage of unmatched CSP haplotypes because they might not be targeted by the vaccine (Dutta *et al*., 2007).

6.2 Recommendation

- The study emphasizes the importance of incorporating population genetic studies into future malaria vaccine design to predict malaria vaccine outcomes in natural parasite populations thus it is highly recommended that

this method is employed in future malaria vaccine trials to adequately evaluate diversity of vaccine antigens.

- The extensive genetic diversity observed lays emphasis on the need for polyvalent malaria vaccines whereby vaccine design should include peptides from several antigens, each one covering the major haplotypes would increase strain coverage and enhance immune recognition.
- It might become more feasible to consider designing vaccines covering specific local haplotypes observed in a region.
- This analysis should be carried out in the other countries where the vaccine is being tested. This will refine our understanding of how genetic diversity affects RTS,S/AS01.
- Further, yearly temporal studies should be conducted to determine the frequency of 3D7 vaccine parasite strain and other non-vaccine haplotypes circulating in the Ghanaian natural parasite populations since potentially vaccine efficacy may be dependent on the extent of homology between the amino acid haplotypes circulating in the natural parasite population and the 3D7 vaccine haplotypes.
- Temporal studies would also help to identify specific haplotypes that are conserved enough to be used in a vaccine either on their own or in a combination of haplotypes to trigger an efficient immune response.
- For optimal comparison of difference in diversity of various eco-epidemiological sites in a country, ideally data should be collected within similar time span in future analysis.
- Structural vaccinology approaches are needs to be applied to design improved CSP-based vaccines by defining epitopes of functional human monoclonal antibodies.

REFERENCES

Abonuusum, A., Owusu-Daako, K., Tannich, E., May, J., Garms, R., & Kruppa, T. (2011). Malaria transmission in two rural communities in the forest zone of Ghana. *Parasitology Research*, *108*(8),1465-1471. https://doi.org/10.1007/s00436-010-2195-1

Abuaku, B., Duah-Quashie, N. O., Quaye, L., Matrevi, S. A., Quashie, N., Gyasi, A., Owusu-Antwi, F., Malm, K., & Koram, K. (2019). Therapeutic efficacy of artesunate-amodiaquine and artemether-lumefantrine combinations for uncomplicated malaria in 10 sentinel sites across Ghana: 2015-2017. *Malaria Journal*, *15*(18),1–12. https://doi.org/10.1186/s12936-019-2848-1

Abukari, Z., Okonu, R., Nyarko, S. B., Lo, A. C., Dieng, C. C., Salifu, S. P., Gyan, B. A., Lo, E., & Amoah, L. E. (2019). The diversity, multiplicity of infection and population structure of *P. Falciparum* parasites circulating in asymptomatic carriers living in high and low malaria transmission settings of Ghana. *Genes*, *10*(6),1–14. https://doi.org/10.3390/genes10060434

Adjah, J., Fiadzoe, B., Ayanful-Torgby, R., & Amoah, L. E. (2018). Seasonal variations in *Plasmodium falciparum* genetic diversity and multiplicity of infection in asymptomatic children living in southern Ghana. *BMC Infectious Diseases*, *18*(1), 1–10. https://doi.org/10.1186/s12879-018-3350-z

Alloueche, A., Milligan, P., Conway, D. J., Pinder, M., Bojang, K., Doherty, T., Tornieporth, N., Cohen, J., & Greenwood, B. M. (2003). Protective efficacy of the RTS,S/AS02 *Plasmodium falciparum* malaria vaccine is not strain specific. *American Journal of Tropical Medicine and Hygiene*, *68*(1), 97–101.

Amambua-Ngwa, A., Amenga-Etego, L., Kamau, E., Amato, R., Ghansah, A., Golassa, L., Randrianarivelojosia, M., Ishengoma, D., Apinjoh, T., Maïga-Ascofaré, O., Andagalu, B., Yavo, W., Bouyou-Akotet, M., Kolapo, O., Mane, K., Worwui, A., Jeffries, D., Simpson, V., D'Alessandro, U., Kwiatkowski, D., & Djimde, A. A. (2019). Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa . *Science*, *365*(6455):813–816. https://doi.org/10.1126/science.aav5427

Amambua-Ngwa, A., Tetteh, K. K. A., Manske, M., Gomez-Escobar, N., Stewart, L. B., Deerhake, M. E., Cheeseman, I. H., Newbold, C. I., Holder, A. A., Knuepfer, E., Janha, O., Jallow, M., Campino, S., MacInnis, B., Kwiatkowski, D. P., & Conway, D. J. (2012). Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites. *PLoS Genetics*, *8*(11). https://doi.org/10.1371/journal.pgen.1002992

Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., & Ménard, R. (2006). Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nature Medicine*, *12*(2): 220–224. https://doi.org/10.1038/nm1350

Appawu, M., Owusu-Agyei, S., Dadzie, S., Asoala, V., Anto, F., Koram, K., Rogers, W., Nkrumah, F., Hoffman, S. L., & Fryauff, D. J. (2004). Malaria transmission dynamics at a site in northern Ghana proposed for testing malaria vaccines. *Tropical Medicine and International Health*, *9*(1):164–170. https://doi.org/10.1046/j.1365-3156.2003.01162.x

Aragam, N. R., Thayer, K. M., Nge, N., Hoffman, I., Martinson, F., Kamwendo, D., Lin, F. C., Sutherland, C., Bailey, J. A., & Juliano, J. J. (2013). Diversity of T Cell Epitopes in Plasmodium falciparum Circumsporozoite Protein Likely Due to Protein-Protein Interactions. *PLoS ONE*, *8*(5):1–13. https://doi.org/10.1371/journal.pone.0062427

Auburn, S., Campino, S., Miotto, O., Djimde, A. A., Zongo, I., Manske, M., Maslen, G., Mangano, V., Alcock, D., MacInnis, B., Rockett, K. A., Clark, T. G., Doumbo, O. K., Ouédraogo, J. B., & Kwiatkowski, D. P. (2012). Characterization of within-host plasmodium falciparum diversity using next-generation sequence data. *PLoS ONE*, *7(2)*, 3–9. https://doi.org/10.1371/journal.pone.0032891

Barry, A. E., Beeson, J., Reeder, J. C., Fowkes, F. J. I., & Arnott, A. (2012). Using Population Genetics to Guide Malaria Vaccine Design. *Malaria Parasites*, 350. https://doi.org/10.5772/34031

Barry, A. E., Schultz, L., Buckee, C. O., & Reeder, J. C. (2009). Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, *Plasmodium falciparum*. *PLoS ONE, 4*(12):e8497. https://doi.org/10.1371/journal.pone.0008497

Beeson, J. G., Drew, D. R., Boyle, M. J., Feng, G., Fowkes, F. J. I., & Richards, J. S. (2016). Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *Microbiology Reviews*, *40*(3), 343–372. https://doi.org/10.1093/femsre/fuw001

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

Bojang, K. A., Milligan, P. J. M., Pinder, M., Vigneron, L., Alloueche, A., Kester, K. E., Ballou, W. R., Conway, D. J., Reece, W. H. H., Gothard, P., Yamuah, L., Delchambre, M., Voss, G., Greenwood, B. M., Hill, A., McAdam, K. P. W. J., Tornieporth, N., Cohen, J. D., & Doherty, T. (2001). Efficacy of RTS,S/AS02 malaria vaccine against Plasmodium falciparum infection in semi-immune adult men in The Gambia: A randomised trial. *Lancet*, *358*(7):1927–1934. https://doi.org/10.1016/S0140-6736(01)06957-4

Casares, S., Brumeanu, T. D., & Richie, T. L. (2010). The RTS,S malaria vaccine. *Vaccine, 28*(31*)*, 4880–4894. https://doi.org/10.1016/j.vaccine.2010.05.033

Clement, M., Snell, Q., Walker, P., Posada, D., & Crandall, K. (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol*. 9:1657–9.

Clinical Trial Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa : final results of a phase 3, individually randomised, controlled trial. *Lancet*. 2015;386:31–45.

Cortés, A., Mellombo, M., Mueller, I., Benet, A., Reeder, J. C., & Anders, R. F. (2003). Geographical structure of diversity and differences between symptomatic and asymptomatic infections for Plasmodium falciparum vaccine candidate AMA1. *Infection and Immunity, 71*(3), 1416–142

Crompton, P. D., Pierce, S. K., Miller, L. H., Crompton, P. D., Pierce, S. K., &

Miller, L. H. (2010). Advances and challenges in malaria vaccine development. *The Journal of clinical investigation*, *120*(12), 4168–4178. https://doi.org/10.1172/JCI44423.4168

Daniels, R., Chang, H. H., Séne, P. D., Park, D. C., Neafsey, D. E., Schaffner, S. F., & Volkman, S. K. (2013). Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PloS one*, *8*(4): e60780.

de Souza, D., Kelly-Hope, L., Lawson, B., Wilson, M., & Boakye, D. (2010). Environmental Factors Associated with the Distribution of Anopheles gambiae s.s in Ghana; an Important Vector of Lymphatic Filariasis and Malaria. *PLoS ONE*, *5*(3), e9927. https://doi.org/10.1371/journal.pone.0009927

Duffy, C. W., Assefa, S. A., Abugri, J., Amoako, N., Owusu-Agyei, S., Anyorigiya, T., MacInnis, B., Kwiatkowski, D. P., Conway, D. J., & Awandare, G. A. (2015). Comparison of genomic signatures of selection on Plasmodium falciparum between different regions of a country with high malaria endemicity. *BMC Genomics*, *16*(1), 527. https://doi.org/10.1186/s12864-015-1746-3

Duffy, C. W., Ba, H., Assefa, S., Ahouidi, A. D., Deh, Y. B., Tandia, A., Kirsebom, F. C. M., Kwiatkowski, D. P., & Conway, D. J. (2017). Population genetic structure and adaptation of malaria parasites on the edge of endemic distribution. *Molecular Ecology*, *26*(11), 2880–2894. https://doi.org/10.1111/mec.14066

Duffy, P. E., & Patrick Gorres, J. (2020). Malaria vaccines since 2000: progress, priorities, products. In *npj Vaccines* *5*(1), 1–9. Nature Research. https://doi.org/10.1038/s41541-020-0196-3

Dutta, S., Seung, Y. L., Batchelor, A. H., & Lanar, D. E. (2007). Structural basis of antigenic escape of a malaria vaccine candidate. *Proceedings of the National Academy of Sciences of the United States of America*104:12488–12493. https://doi.org/10.1073/pnas.0701464104

Egan, J. E., Hoffman, S. L., Haynes, J. D., Sadoff, J. C., Schneider, I., Grau, G. E.,

Hollingdale, M. R., Ballou, W. R., & Gordon, D. M. (1993). Humoral immune responses in volunteers immunized with irradiated *Plasmodium falciparum* sporozoites. *The American Journal of Tropical Medicine and Hygiene*, *49*(2), 166–173. http://www.ncbi.nlm.nih.gov/pubmed/8357078

Enosse, S., Dobaño, C., Quelhas, D., Aponte, J. J., Lievens, M., Leach, A., Sacarlal, J., Greenwood, B., Milman, J., Dubovsky, F., Cohen, J., Thompson, R., Ballou, W. R., Alonso, P. L., Conway, D. J., & Sutherland, C. J. (2006). RTS,S/AS02A Malaria Vaccine Does Not Induce Parasite CSP T Cell Epitope Selection and Reduces Multiplicity of Infection. *PLoS Clinical Trials*, *1*(1), e5. https://doi.org/10.1371/journal.pctr.0010005

European Medicines Agency. (2015). First malaria vaccine receives positive scientific opinion from EMA. *Pharm J*, 44:30–2

Gandhi, K., Thera, M. A., Coulibaly, D., Traoré, K., Guindo, A. B., Ouattara, A., Takala-Harrison, S., Berry, A. A., Doumbo, O. K., & Plowe, C. V. (2014). Variation in the Circumsporozoite Protein of *Plasmodium falciparum*: Vaccine Development Implications. *PLoS ONE*, *9*(7), e101783. https://doi.org/10.1371/journal.pone.0101783

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., & Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, *419*(6906), 498–511. https://doi.org/10.1038/nature01097

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

Global Fund. (2019). https://data.theglobalfund.org/locations/GHA

Good, M. F., Pombo, D., Quakyi, I. A., Riley, E. M., Houghten, R. A., Menon, A., Alling, D. W., Berzofsky, J. A., & Miller, L. H. (1988). Human T-cell recognition of the circumsporozoite protein of Plasmodium falciparum: immunodominant T-cell domains map to the polymorphic regions of the molecule. *Proceedings of the National Academy of Sciences*, *85*(4,1199–1203. https://doi.org/10.1073/PNAS.85.4.1199

Hafalla, J. C. (2011). Cell biology and immunology of malaria. *240*(5),297–316.

Henden, L., Lee, S., Mueller, I., Barry, A., & Bahlo, M. (2018). Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. In *PLoS Genetics* 14:e1007279. https://doi.org/10.1371/journal.pgen.1007279

Hoffman, S. L., Goh, L. M. L., Luke, T. C., Schneider, I., Le, T. P., Doolan, D. L., Sacci, J., de la Vega, P., Dowler, M., Paul, C., Gordon, D. M., Stoute, J. A., Church, L. W. P., Sedegah, M., Heppner, D. G., Ballou, W. R., & Richie, T. L. (2002). Protection of Humans against Malaria by Immunization with Radiation-Attenuated *Plasmodium falciparum* Sporozoites. *The Journal of Infectious Diseases*, *185*(8), 1155–1164. https://doi.org/10.1086/339409

Jalloh, A., Jalloh, M., & Matsuoka, H. (2009). T-cell epitope polymorphisms of the Plasmodium falciparum circumsporozoite protein among field isolates from Sierra Leone: Age-dependent haplotype distribution? *Malaria Journal*, *8*(1),1–9. https://doi.org/10.1186/1475-2875-8-120

Karunamoorthi, K. (2014). Malaria vaccine: A future hope to curtail the global malaria burden. *International Journal of Preventive Medicine*, *5*(5), 529–538.

Kasasa, S., Asoala, V., Gosoniu, L., Anto, F., Adjuik, M., Tindana, C., Smith, T., Owusu-Agyei, S., & Vounatsou, P. (2013). Spatio-temporal malaria transmission patterns in Navrongo demographic surveillance site, northern Ghana. *Malaria Journal*, *12*(1), 1–10. https://doi.org/10.1186/1475-2875-12-63

Koram, K. A., Owusu-Agyei, S., Fryauff, D. J., Anto, F., Atuguba, F., Hodgson, A.,

Holfman, S. L., & Nkrumah, F. K. (2003). Seasonal profiles of malaria infection, anaemia, and bednet use among age groups and communities in northern Ghana. *Tropical Medicine and International Health*, *8*(9),793–802. https://doi.org/10.1046/j.1365-3156.2003.01092.x

Kumar, K. A., Sano, G., Boscardin, S., Nussenzweig, R. S., Nussenzweig, M. C., Zavala, F., & Nussenzweig, V. (2006). The circumsporozoite protein is an immunodominant protective antigen in irradiated sporozoites. *Nature*, *444*(7121), 6–9. https://doi.org/10.1038/nature05361

Lee Stuart. (2016). Bahlo M. moimix: an R package for assessing clonality in high-throughput sequencing data.

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9),1110–1116. https://doi.org/10.1111/2041-210X.12410

Manske, M., Miotto, O., Campino, S., Auburn, S., Zongo, I., Ouedraogo, J., Michon, P., Mueller, I., Su, X., Amaratunga, C., Fairhurst, R., Socheat, D., Imwong, M., White, N. J., Sanders, M., Anastasi, E., Rubio, V. R., Jyothi, D., & Amenga-etego, L. (2013). Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature, 487*(7407), 375-379 https://doi.org/10.1038/nature11174

Mbengue A, Bhattacharjee S, Pandharkar T, Liu H, Estiu G, Stahelin RV, Rizk SS, Njimoh DL, Ryan Y, Chotivanich K, Nguon C, Ghorbal M, Lopez-Rubio JJ, Pfrender M, Emrich S, Mohandas N, Dondorp AM, Wiest O, H. K. (2015). A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. *Nature*, *520*(7549), 683–687. https://doi.org/10.1038/nature14412.A

Mensah-Brown, H. E., Abugri, J., Asante, K. P., Dwomoh, D., Dosoo, D., Atuguba, F., Conway, D. J., & Awandare, G. A. (2017). Assessing the impact of differences in malaria transmission intensity on clinical and haematological indices in children with malaria. *Malaria Journal*, *16*(11), 1–11. https://doi.org/10.1186/s12936-017-1745-8

Mohapatra, S. (2013). Malaria vaccine: A myth or a reality. *Annals of Tropical Medicine and Public Health*, *6*(3), 274. https://doi.org/10.4103/1755-6783.120982

Neafsey, D. E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., Lievens, M., Abdulla, S., Adjei, S., Agbenyega, T., Agnandji, S. T., Aide, P., Anderson, S., Ansong, D., Aponte, J. J., Asante, K. P., Bejon, P., Birkett, A. J., Bruls, M., Connolly, K. M., D'Alessandro, U., Dobaño, C., Gesase, S., Greenwood, B., Grimsby, J., Tinto, H., Hamel, M. J., Hoffman, I., Kamthunzi, P., Kariuki, S., Kremsner, P. G., Leach, A., Lell, B., Lennon, N. J., Lusingu, J., Marsh, K., Martinson, F., Molel, J. T., Sacarlal, J., Sogoloff, B., Sorgho, H., Tanner, M., Theander, T., Valea, I., Volkman, S. K., Yu, Q., Lapierre, D., Birren, B. W.,…Wirth, D. F. (2015). Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *New England Journal of Medicine*, *373*(21), 2025–2037. https://doi.org/10.1056/NEJMoa1505819

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nkrumah, F., Ama, N., Klutse, B., Adukpo, D. C., Owusu, K., Quagraine, K. A., Owusu, A., Gutowski, W., Ama, N., & Akumenyi Quagraine, K. (2014). Rainfall Variability over Ghana: Model versus Rain Gauge Observation. *International Journal of Geosciences*, *5*(7), 673–683. https://doi.org/10.4236/ijg.2014.57060

Ochola-Oyier, L. I., Okombo, J., Wagatua, N., Ochieng, J., Tetteh, K. K., Fegan, G., ... & Marsh, K. (2016). Comparison of allele frequencies of *Plasmodium falciparum* merozoite antigens in malaria infections sampled in different years in a Kenyan population. Malaria journal, *15*(1), 1-11.

Oduro, A. R., Wak, G., Azongo, D., Debpuur, C., Wontuo, P., Kondayire, F., Welaga, P., Bawah, A., Nazzar, A., Williams, J., Hodgson, A., & Binka, F. (2012). Profile of the Navrongo health and demographic surveillance system.

International Journal of Epidemiology, *41*(4), 968–976. https://doi.org/10.1093/ije/dys111

Ouattara, A., & Laurens, M. B. (2015). Vaccines against malaria. *Clinical Infectious Diseases*, *60*(6), 930–936. https://doi.org/10.1093/cid/ciu954

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(3), 2074–2093. https://doi.org/10.1371/journal.pgen.0020190

Peyerl-Hoffmann, G., Jelinek, T., Kilian, A., Kabagambe, G., Metzger, W. G., & Von Sonnenburg, F. (2001). Genetic diversity of Plasmodium falciparum and its relationship to parasite density in an area with different malaria endemicities in West Uganda. *Tropical Medicine and International Health*, *6*(8), 607–613. https://doi.org/10.1046/j.1365-3156.2001.00761.x

Plassmeyer, M. L., Reiter, K., Shimp, R. L., Kotova, S., Smith, P. D., Hurt, D. E., House, B., Zou, X., Zhang, Y., Hickman, M., Uchime, O., Herrera, R., Nguyen, V., Glen, J., Lebowitz, J., Jin, A. J., Miller, L. H., MacDonald, N. J., Wu, Y., & Narum, D. L. (2009). Structure of the *Plasmodium falciparum* Circumsporozoite Protein, a Leading Malaria Vaccine Candidate. *Journal of Biological Chemistry*, *284*(39), 26951–26963. https://doi.org/10.1074/jbc.M109.013706

Pringle, J. C., Carpi, G., Almagro-Garcia, J., Zhu, S. J., Kobayashi, T., Mulenga, M., Bobanga, T., Chaponda, M., Moss, W. J., & Norris, D. E. (2018). RTS,S/AS01 malaria vaccine mismatch observed among *Plasmodium falciparum* isolates from southern and central Africa and globally. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-24585-8

Richards, J. S., & Beeson, J. G. (2009). The future for blood-stage vaccines against malaria. *Immunology and Cell Biology*, *87*(5), 377–390. https://doi.org/10.1038/icb.2009.27

Riley, E. M., Allen, S. J., Bennet, S., Thomas, P. J., Donnell, A. O., Lindsay, S. W., Good, M. F., & Greenwood, B. M. (1990). Recognition of dominant T cell-

stimulating circumsporozoite protein of *Plasmodium malaria* morbidity in Gambian children epitopes falciparum from the and relationship to malaria morbidity in Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *84*(5), 648–657. https://doi.org/10.1016/0035-9203(90)90133-Y

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832–837. https://doi.org/10.1038/nature01140

Snounou, G., Viriyakosol, S., Xin Ping Zhu, Jarra, W., Pinheiro, L., do Rosario, V. E., Thaithong, S., & Brown, K. N. (1993). High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Molecular and Biochemical Parasitology*, *61*(2), 315–320. https://doi.org/10.1016/0166-6851(93)90077-B

Stoute, J. A., Slaoui, M., Heppner, D. G., Momin, P., Kester, K. E., Desmons, P., Wellde, B. T., Garçon, N., Krzych, U., Marchand, M., Ballou, W. R., & Cohen, J. D. (1997). A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria. *New England Journal of Medicine*, *336*(3), 86–91. https://doi.org/10.1056/NEJM199701093360202

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3):585-95. https://doi.org/10.1093/genetics/123.3.585.

Takala, S. L., & Plowe, C. V. (2010). efficacy : Preventing and overcoming " vaccine resistant malaria." *Parasite Immunol*, *31*(9), 560–573. https://doi.org/10.1111/j.1365-3024.2009.01138.x

Tariq Ahmad, Matt Neville, Sara E. Marshall, Alessandro Armuzzi, Kim Mulcahy-Hawes, Jonathan Crawshaw, Hiroe Sato, Khoon-Lin Ling, Martin Barnardo, Sue Goldthorpe, Robert Walton, Mike Bunce, Derek P. Jewell, Ken I. Welsh, (2003). Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. Human molecular genetics, *12*(6), 647-656. https://doi.org/10.1093/hmg/ddg066

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, *132*(2): 619–633.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, 47: 97–120. https://doi.org/10.1146/annurev-genet-111212-133526

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

Waitumbi, J. N., Anyona, S. B., Hunja, C. W., Kifude, C. M., Polhemus, M. E., Walsh, D. S., Ockenhouse, C. F., Heppner, D. G., Leach, A., Lievens, M., Ballou, W. R., Cohen, J. D., & Sutherland, C. J. (2009). Impact of RTS,S/AS02A and RTS,S/AS01B on genotypes of *P. falciparum* in adults participating in a malaria vaccine clinical trial. *PLoS ONE*, *4*(11):1–8. https://doi.org/10.1371/journal.pone.0007849

Weedall, G. D., & Conway, D. J. (2010). Detecting signatures of balancing selection to identify targets of anti-parasite immunity. In *Trends in Parasitology*. *26*(7): 363-369. https://doi.org/10.1016/j.pt.2010.04.002

WHO. (2015). WHO Guidelines for the Treatment of Malaria. *Diagnosis of Malaria*, 27–30. https://doi.org/10.1016/0035-9203(91)90261-V

WHO. (2018). World Malaria Report 2020. World Health Organization. In *World Health*.

WHO. (2019). Malaria Vaccine Implementation Programme (MVIP): Proposed Framework For Policy Decision on RTS,S/AS01 Malaria Vaccine (Issue March)

WHO. (2019). World Malaria Report 2020. World Health Organization. In *World Health*.

Yamauchi, L. M., Coppi, A., Snounou, G., & Sinnis, P. (2007). Plasmodium sporozoites trickle out of the injection site. *Cellular microbiology*, *9*(5), 1215-1222. https://doi.org/10.1111/j.1462-5822.2006.00861.x

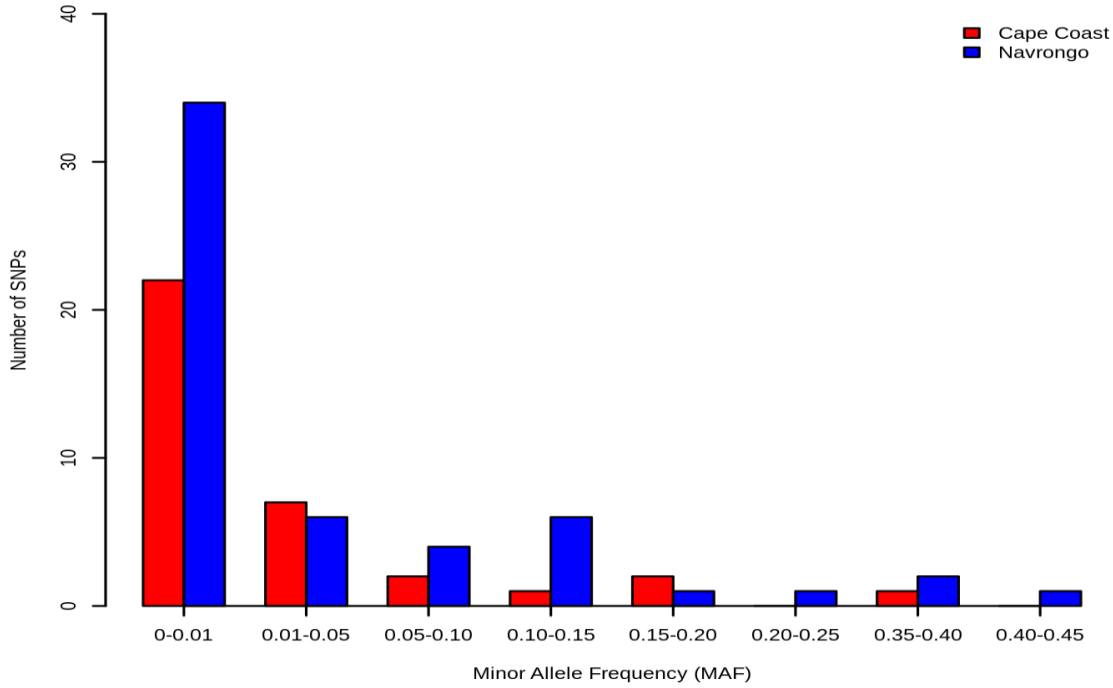
Yuan, L., Zhao, H., Wu, L., Li, X., Parker, D., Xu, S., Zhao, Y., Feng, G., Wang, Y., Yan, G., Fan, Q., Yang, Z., & Cui, L. (2013). *Plasmodium falciparum* populations from northeastern Myanmar display high levels of genetic diversity at multiple antigenic loci. *Acta Tropica*, *125*(1), 53–59. https://doi.org/10.1016/j.actatropica.2012.09.008

Zeeshan, M., Alam, M. T., Vinayak, S., Bora, H., Tyagi, R. K., Alam, M. S., Choudhary, V., Mittra, P., Lumb, V., Bharti, P. K., Udhayakumar, V., Singh, N., Jain, V., Singh, P. P., & Sharma, Y. D. (2012). Genetic variation in the plasmodium falciparum circumsporozoite protein in india and its relevance to RTS,S malaria vaccine. *PLoS ONE*, *7*(8), 1–10. https://doi.org/10.1371/journal.pone.0043430

Zhai, W., Nielsen, R., & Slatkin, M. (2009). An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data. *Molecular Biology and Evolution*, *26*(2), 273–283. https://doi.org/10.1093/molbev/msn231

APPENDICES

Appendix i: **Minor Allele Frequency distribution**

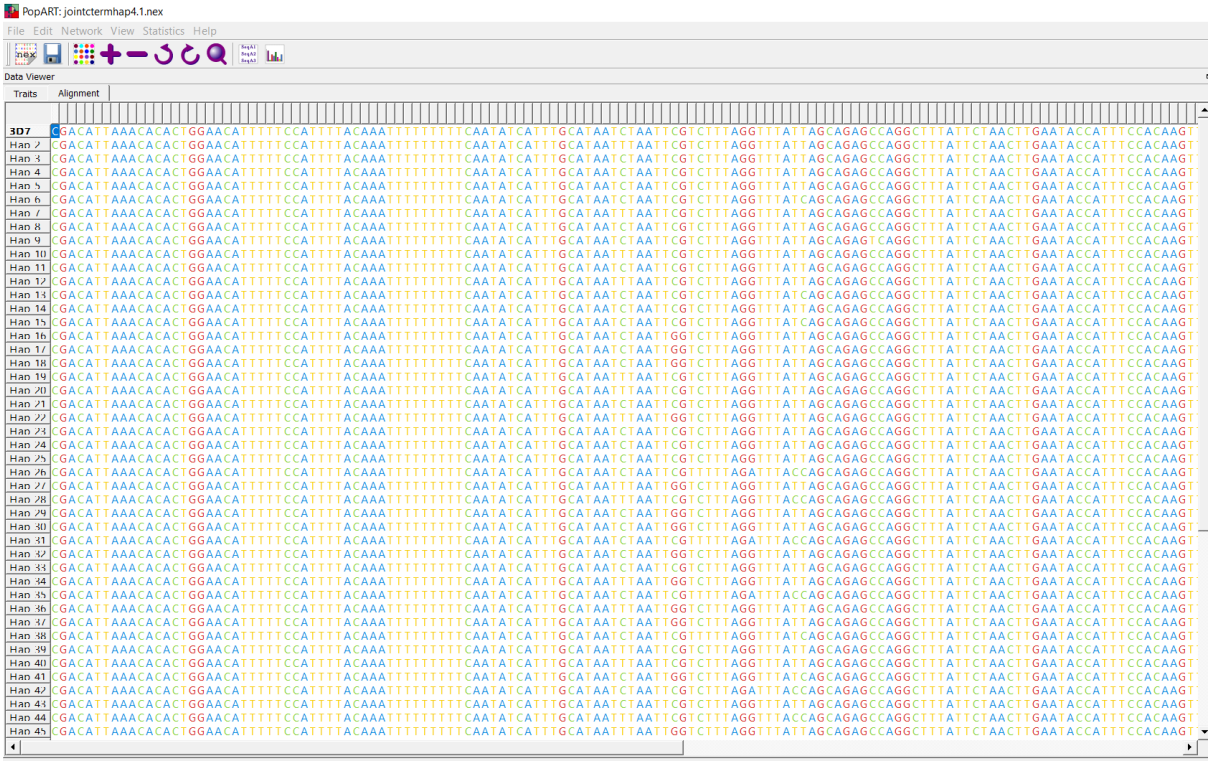


Plot shows the allele frequency distribution of a total of 90 SNPs set within the *Pfcsp* loci in samples from both Cape Coast (n= 35) and Navrongo (n= 55). Vertical axis represents the number of SNPs in each category of allele frequency and the horizontal axis shows the binned SNPs set representing the following MAFs ranges; **[0.0-0.01], [0.01-0.05], [0.05-0.10], [0.10-0.15], [0.15-0.20], [0.20-0.25], [0.35-0.40], [0.40-0.45])** respectively. There were no alleles found within the **[0.25-0.30]** and **[0.30-0.35]** in both populations.

Appendix ii: C-terminal Pfcsp haplotype sequences and frequencies

The screenshot and table below show the C-terminal Pfcsp haplotype sequences and frequencies included in the Templeton, Crandall, and Sing (TCS) network.

Screenshot of the C-terminal *Pfcsp* haplotype sequences



Pfcsp C-terminal haplotype frequency table

Haplotype	Navrongo	Cape Coast
3D7	15	84
Hap_2	0	4
Hap_3	0	2
Hap_4	0	48
Hap_5	0	1
Hap_6	0	9
Hap_7	0	6
Hap_8	0	4
Hap_9	0	4
Hap_10	1	11
Hap_11	0	2
Hap_12	0	4
Hap_13	0	1
Hap_14	0	1

Hap_15	0	3
Hap_16	52	0
Hap_17	16	0
Hap_18	26	0
Hap_19	1	0
Hap_20	4	0
Hap_21	5	0
Hap_22	4	0
Hap_23	2	0
Hap_24	1	0
Hap_25	6	0
Hap_26	14	0
Hap_27	2	0
Hap_28	2	0
Hap_29	26	0
Hap_30	5	0
Hap_31	1	0
Hap_32	1	0
Hap_33	1	0
Hap_34	5	0
Hap_35	2	0
Hap_36	3	0
Hap_37	1	0
Hap_38	1	0
Hap_39	5	0
Hap_40	2	0
Hap_41	3	0
Hap_42	3	0
Hap_43	6	0
Hap_44	1	0
Hap_45	1	0
Hap_46	2	0
Hap_47	1	0
Hap_48	1	0
Hap_49	1	0
Hap_50	2	0
Hap_51	1	0
Hap_52	9	0
Hap_53	1	0
Hap_54	2	0
Hap_55	1	0
Hap_56	1	0
Hap_57	2	0
Hap_58	2	0

Hap_59	3	0
Hap_60	1	0
Hap_61	2	0
Hap_62	1	0
Hap_63	1	0
Hap_64	2	0
Hap_65	1	0
Hap_66	1	0

Appendix iii: TH2R amino acid haplotype frequencies

TH2R Amino Acid Haplotypes	Cape Coast frequency	Navrongo frequency	Amino Acid differences
PSDKHIKEYLNKIQNSL(3D7)	104	19	0
PSDQHIKEYLNKIQNSL	56	0	1
PSDKHIKEYLNKIQNSI	10	0	1
PSDKHIKEYLNKIRNSL	3	0	1
PSDKHIKEYLKKIQNSL	0	5	1
PSDKHIKEYLTKIQNSL	0	2	1
PSDKHIEEYLNKIQNSL	0	2	1
PSDQHIKEYLNKIQNSI	2	0	2
PSDQHIKEYLNKIRNSL	1	0	2
PSDQHIKEYLHKIQNSL	4	0	2
PSDQHIKEYINKIQNSL	2	0	2
PSDKHIEEYLKKIQNSL	0	33	2
PSDKHITEYLKKIQNSL	0	17	2
PSDQHIEEYLNKIQNSL	0	1	2
PSDKHIEEYLQKIQNSL	0	1	2
PSDKHIKEYLNKIKNSI	0	1	2
PSDQHIKEYLHKIRNSL	2	0	3
PSDQHIEEYLKKIQNSL	0	57	3
PSDKHIEEYLKKIKNSL	0	28	3
PSDQHIEEYLNKIKNSL	0	5	3
PSDQHIEEYLQKIQNSL	0	5	3
PSDKHITEYLKKIKNSL	0	2	3
PSDKHIEEYLKKIRNSL	0	4	3
PSDQHIEEYLTKIQNSL	0	1	3
PSDKHIEEYLNKIKNSI	0	9	3
PSDKHIEEYIKKIQNSL	0	1	3

PSDKHIEEYLKKIKNSI	0	10	4
PSDKHITEYLKKIKNSI	0	6	4
PSDQHIEEYLQKIKNSL	0	4	4
PSDQHIEEYLKKIKNSL	0	30	4
PSDQHIEEYLQKIRNSL	0	3	4
PSDQHIEEYIKKIQNSL	0	3	4
PSDQHIEEYLNKIKNSI	0	1	4
PSDQHITEYLKKIKNSI	0	2	5
PSDQHIEEYLQKIKNSI	0	3	5
PSDQHIEEYIKKIKNSI	0	1	6

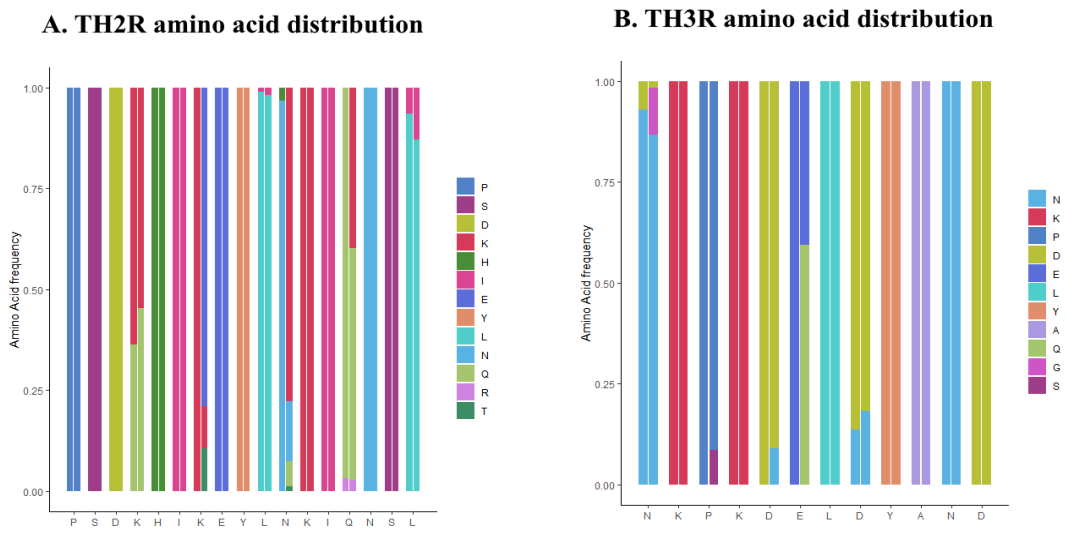
Data presented here shows TH2R (311-327 amino acid positions) amino acid haplotype frequencies observed in both Cape Coast and Navrongo populations with their corresponding amino acid differences which have been calculated against *Plasmodium falciparum* 3D7 reference (0304600.1, PlasmoDB)

Appendix iv: TH3R amino acid haplotype frequencies

TH3R Amino Acid Haplotypes	Cape Coast Frequency	Navrongo Frequency	Amino Acid Differences
NKPKDELDYAND(3D7)	146	47	0
NKPKDELNYAND*	25	26	1
DKPKDELDYAND	13	0	1
NKPKDQLDYAND	0	132	1
NKPKDQLNYAND	0	17	2
GKPKDELNYAND	0	3	2
DKPKNELDYAND	0	1	2
DKPKDQLDYAND	0	3	2
GKSKDELDYAND	0	5	2
GKPKNELDYAND	0	4	2
GKSKNELDYAND	0	17	3
GKPKNELNYAND	0	1	3

Data presented here shows TH3R (352-363 amino acid positions) amino acid haplotype frequencies observed in both Cape Coast and Navrongo populations with their corresponding amino acid differences which have been calculated against *Plasmodium falciparum* 3D7 reference (0304600.1, PlasmoDB).

Appendix v: TH2R and TH3R amino acid distribution



Plot shows the distribution of each amino acid in the TH2R and TH3R epitopes Cape Coast and Navrongo. The reference 3D7 amino acid sequences are shown on the X axis. The amino acid column represents the frequency of amino acids seen in isolates from Cape Coast (left half of the column) and Navrongo (right half of the column).

Appendix vi: Python Scripts

All customized python scripts written for the purpose of this analysis can be accessed on this github account: https://github.com/Eliameg