Curriculum Vitae

Nadiatou Ouedarogo
M.S. Candidate
Cellular and Molecular Biomedical Sciences
University of Maryland School of Medicine

Date: June 24, 2022

Contact Information
E-mail: nadia.ouedraogo@som.ummaryland.edu

Education

2016	B.S.	State University of New York, New Platz Biology, French minor
July 2022	M.S.	University of Maryland School of Medicine, Baltimore, MD Cellular and Molecular Biomedical Sciences

Research Experience

2021-Present	M.S. Candidate Cellular and Molecular Biomedical Sciences Malaria Lab, Center for Vaccine Development Department of Medicine, University of Maryland Baltimore, MD Mentor: Dr. Matthew Laurens
2021	Rotation M.S Candidate Immunology lab Department of Microbiology and Immunology, University of Maryland Baltimore, MD Mentor: Dr. Nevil Singh
2016	Independent Study in Entomology SUNY New Paltz Mentor: Dr. Aaron Haselton
2016	Research Experience for Undergraduates, University of Hawaii at Hilo Mentor: Dr. Jonathan Koch
2014	Biochemistry Research

SUNY New Patz, NY
Mentor: Dr. Megan Ferguson

2013 Hydrogeology Research
SUNY New Patz, NY
Mentor: Dr. Shafiul Chowdery

2013 Independent Study in Ecology
SUNY New Paltz, NY
Mentor: Dr. Eric Keeling

Presentation and Conferences

2022 Graduate Research Conference
Baltimore, MD
Poster: Evaluating the Difference of Circumsporozoite Protein (CSP) Sequences Between Vaccinated and Unvaccinated Persons Infected with Malaria

2016 PIPES REU Annual Symposium
Hilo, Hawaii
Oral/Poster: The invasion biology of *Drosophila suzukii* in Hawaii.

2015 C-Step Annual Conference
Albany, NY
Poster: Examining Bacterial Cell Interactions using Atomic Force Microscopy

2013 Student Symposium.
SUNY New Paltz, NY
Poster: Potential Flooding on SUNY New Paltz Campus and the Use of Green Infrastructure

Publications

1. Koch, J., Dupuis, J., Jardeleza M., Geib, S., Follett, P., Price, D. Ouedraogo, N. (2020). Population genomic and phenotype diversity of invasive Drosophila suzukii in Hawai'i. Biological Invasions, 1-18.

2. Ouedraogo, N., Aucapina, R., & Ferguson, M. A. (2015). Examining Bacterial Cell Interactions using Atomic Force Microscopy. *Biophysical Journal*, *108*(2), 633a.

Honors and award

2021 HJF Superior Performance Award

2016	National Science Foundation (NSF) - Research Experience for Undergraduates recipient
2016	CStep/AMP Outstanding Scholar Award
2014	SUNY New Platz Honoree of Chi Alpha Epsilon

Techniques

Polymerase Chain Reaction
Cell culture
Sample extraction from Dried Blood Sampling
RT-PCR
Gel Electrophoresis
DNA extraction
Use of SnapGene
Use of BioEdit
Use of Atomic Force Microscopy
Filed survey techniques to determine forest community structure and composition,
Field techniques for determining soil and topographic variables,
Field dendrochronology (tree-ring) techniques to determine tree ages and growth rates,
Identification of common songbirds,
Collecting insects,
Making insets baits,
Measuring insect body dimensions
Creating vibration sensor using Arduino

Relevant Course work

Cellular and Molecular Basis of Translational Research
Biostatistics
Basic Immunology
Advanced Microbiology Pathogenesis
Genomics and Bioinformatics

Abstract

Title: Using PacBio SMRT platform to assess the genetic variation of *Plasmodium falciparum* Circumsporozoite Protein (*Pf*CSP) in Kalifabougou, Mali

Nadiatou Ouedraogo, Master of Science, 2022

Thesis Directed by: Matthew Laurens, MD, MPH, Professor

Objectives: In this study, we assessed the genetic variation in CSP sequences from Kalifabougou, Mali, by evaluating the diversity of haplotypes in the N- and C- Terminal region, the repeat regions, and the T- and B-cell epitopes.

Methods: Sequences used for the study were generated from 59 sequences collected during a malaria transmission season conducted in Kalifabougou, Mali. Generated using the PacBio SMRT platform.

Results: Results showed that the central repeat and C-terminal Th2R/Th3R epitope regions were highly polymorphic in this study, whereas the N-terminal non-repeat region was less variable. Furthermore findings highlighted CSP polymorphism and suggested that the repeat and B-cell epitope of our samples is similar to those of 7G8 *Pf*CSP, whereas the C-terminal is similar to that of the 3D7 *Pf*CSP strain. Finally, findings the CSP sequences from the Kalifabougou site are more similar to those from Navrongo than to those from Cape Coast.

Using PacBio SMRT platform to assess the genetic variation of *Plasmodium falciparum* Circumsporozoite Protein (*Pf*CSP) in Kalifabougou, Mali

by
Nadiatou Ouedraogo

Thesis submitted to the faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Master of Science
2022

DEDICATION

I would like to dedicate this work to my beloved Grandfather Abibou Galbane. He taught me the value of education which is the reason why I try my best to succeed in it. Yaaba, Thank you for your unwavering support. Thank you for showing and showering me with unconditional love. Thank you for your care, protection, advises, support and the finest omelet I've ever had. Thank you for taking me to school every day when I was little. I wish you were here for me to share this work with you. But I do hope that you can still see me from above, and that you are beaming with pride. I love you beyond words and I miss you tremendously.

ACKNOWLEDGEMENTS

I am grateful to my mentor, Dr. Matthew Laurens, for his commitment to my work and growth as a scientist. This study would not have been possible without his help, passion for his field and intellect. I'm also grateful for the opportunities he gave me during my rotation and research studies, as well as the opportunity to present my work during the Malaria lab meeting and the Graduate Research Conference. He was the driving force behind my decision to apply the CMBS program at the school of Medicine in the first place and has been guiding and advising me way before my Master's journey. It has been a tremendous Honor to have him as a mentor. I would like to say a huge thank you to Dr. Amed Ouattara for his mentorship and guidance through every step of my thesis project. My investigation was made feasible by his very impressive prior work in the field of malaria. He spent a lot of time teaching me how to utilize analysis software, such as BioEdit, and evaluate my results critically, read my drafts, and provide me helpful input to improve my writing and help me grow as a scientist. Without his help, I would not have been able to conduct and complete my thesis.

I'd like to thank the Center for Vaccine Development and Global Health, as well as the Malaria Lab, particularly Biraj Shrestha for teaching me so many wet lab techniques, and Sudhaunshu Joshi and Kathy for allowing me to shadow them and learn about their work. I want to thank Tina Williams for always being willing to assist me with organizing my meeting schedules and resolving IT issues. I want to express my gratitude to Melissa Myers and Ana Raquel for always ensuring that I have all of the resources I needed to complete my work. I want to express my gratitude to the Mali Research Team and the Malaria lab for gathering and processing the samples that were detrimental to my study.

I want to say how much I appreciate my committee' members contributions to this project. Dr. David Serre gave me important insights into my experimental method design, Dr. Miriam Laufer gave me great guidance on the development of my study and my scientific writing, and Dr. Eileen Barry helped me refine my thesis proposal. I'd like to thank the GPILS program at the University of Maryland, particularly Dr. Ivy Dick and Elice Garcia, for their encouragement and support in my pursuit of my Master' studies over the past few years. I would like to thank my undergraduate mentors Dr. Eric Keeling and Dr. Jonathan Koch for inspiring my passion of scientific research and for their continuous support over the years, even though I am no longer their mentee. If our paths had not crossed, I would not be here. Last but not least, I'd like to thank my family, both in Burkina Faso and in the United States, for their faith in me, their support, and their pride. I want to thank my best friend and partner Boris Djiguemde for everything he has done for me over the years, including listening, reading, supporting me, seeing that I don't starve myself, and helping me overcome obstacles. I love each and every one of you very much.

Table of Contents

LIST OF FIGURES

LIST OF ABBREVIATIONS

Aa	Amino Acid
ACTs	Artemisinin-based combination therapies
AMA1	Apical membrane antigen 1
AS01	Adjuvant system including in the RTS,S vaccine
CS	Circumsporozoite
CSP	Circumsporozoite Protein
GAP	Genetically Attenuated Parasites
GLA-SE	glucopyranosyl lipid adjuvant-stable emulsion
GPI	Glycosylphosphatidylinositol
HSPG	Heparin Sulfate Proteoglycans
LSA	Liver Stage Antigen
MSP	Merozoite Surface Protein
NANP/NVDP	Tetramer Repeat of Targeted by Antibodies
Pf	*Plasmodium falciparum*
***Pf*CSP**	*Plasmodium falciparum* Circumsporozoite Protein
PfSPZ	Radiation-attenuated *Plasmodium falciparum* Sporozoite
PfSPZ-CVac	non-replicating irradiated whole sporozoites Vaccine
PRIMVAC	Placental malaria vaccine
RAS	Radiation attenuated sporozoites
RTS,S	Malaria vaccine composed of the 3D7 CSP repeat region, T Cell epitope and Hepatitis B surface antigen
SMRT	Single Molecule, Real-Time

SNP	Single Nucleotide Polymorphism
TBV	Transmission-blocking vaccines
TH2R	Helper T-cell epitope of the circumsporozoite protein
TH3R	Helper T-cell epitope of the circumsporozoite protein
TRAP	Thrombospondin-related Adhesion Protein
TSR	Thrombospondin Type 1 Repeat
ZMW	Zero-mode Waveguide

I. INTRODUCTION

The *Plasmodium falciparum* (*Pf*) circumsporozoite protein (CSP) is a secreted protein of the sporozoite stage of the malaria parasite. The genetic structure of CSP (figure 1) has an immunodominant core repeat region flanked by conserved motifs at the N- and C-termini, which are thought to be involved in protein processing as the parasite travels from the mosquito's midgut to the mammalian carrier (1,2). Dame and colleagues identified CSP in *P. gallinaceum* in 1984 (3). CSP is the sporozoite's primary surface protein. The protein is multifunctional as it is necessary for sporozoite growth and mediates numerous aspects of its journey from the mosquito's saliva to the liver of the mammalian host (2). CSP has two structural states: an adhesive conformation with the C-terminal cell-adhesive domain exposed, and a non-adhesive conformation with the N terminus masking this domain (2). The cell-adhesive domain plays a role in sporozoite formation and hepatocyte invasion, such as facilitating the interaction of CSP with heparin sulfate proteoglycans (HSPGs) during the primary attachment of sporozoites to hepatocytes. (4)

Figure 1: Diagram of the genetic structure of plasmodium falciparum circumsporozoite protein (Adapted from Gandhi et al, 2014 (5)).

CSP is the active component of the only licensed malaria vaccine RTS,S (5) and a relatively new formulation (R21/Matrix-M) that has shown a 77.0% efficacy against clinical episodes in a phase 2 trial (7).

The circumsporozoite (CS) gene is diverse, thus CSP is polymorphic as well, with up to 454 haplotypes that can be found in its sequence (8). When compared to the envelope protein found on the Dengue Fever virus, CSP is more variable. The Dengue virus envelope protein, like CSP, is an important target for neutralizing antibodies and vaccine development since it mediates host cell receptor interactions. Similar to CSP, it is believed that amino acid changes in the envelope protein affect antibody recognition and binding, as well, (9,10). In a study performed in in Zhejiang in 2019, researchers discovered that among the 10 sites of E protein amino acid substitutions observed in Dengue Virus Type1 (DENV-1) strains compared to the D1 Hawaii reference strain, six sites of amino acid substitutions were observed in strains isolated from two cities, while two sites of amino acid substitutions were observed in strains isolated from two other cities (10). Due to the high number of haplotypes (454) identified in the CSP sequence, CSP is a far more diverse protein than the E protein, which presents only a few mutation sites.

Determining the genetic variation of malaria vaccine antigens is an important element in understanding and ascertaining vaccine escape, a phenomenon where some strains of the parasite are not affected by vaccine-induced immunity, therefore making the vaccine ineffective in preventing the disease, especially in endemic areas where malaria vaccines

would be deployed. Due to different malaria transmission patterns and variations in human immune response to malaria, parasite selection patterns vary accordingly (6). Thus, determining the genetic diversity of vaccine antigens and/or epitopes can be advantageous for the design of an efficacious malaria vaccine. For vaccines with rare alleles, allele-specific efficacy of vaccines based on major strain may select for these uncommon variants, affecting the overall efficacy (8).

An assessment of CSP sequence variation will provide knowledge and insight, as well as a comprehensive examination of the impact of local and eventually broader diversity on the use of CSP-based vaccines. Vaccine antigen' repeats and epitopes are important because they are the regions recognized by the human immune system to fight infections. That is why, when assessing genetic diversity, these antigenic regions must be considered.

Study Objective

Failure to consider the genetic variation, specifically in the key amino acid residues which can lead to malaria vaccine-generated immune responses, can result in a lack of efficacy in field trials. Therefore, an assessment of the genetic diversity of potential vaccine candidate antigens is an essential step for designing and deploying a candidate malaria vaccine in the field. For the purpose of this study, we aimed to:

- Assess the genetic variation in the **CSP** sequences from samples collected in Kalifabougou, Mali. Thus, we sought to:
    - Measure the distribution of amino acid mutations in the N and C terminal regions

    - Evaluate the number of repeats across the samples
    - Evaluate the diversity of T cell epitopes
    - Evaluate the diversity of B cell epitopes

- Compare the variations in CSP sequences from Kalifabougou, Mali, to those observed in *Pf* laboratory strains 3D7 and 7G8.
    - Assess variations of the major haplotype from Kalifabougou compared to 3D7 and 7G8.

With these measures we hypothesize the following:

- CSP sequences from Kalifabougou samples will be polymorphic, particularly in the C terminal region
- CSP sequences from Kalifabougou samples will have similar amino acid sequences to those found in the immunodominant regions of the 3D7 PfCSP
- CSP sequences from Kalifabougou will differ from CSP sequences from othe regions such as Navrongo and Cape Coast, in Ghana

II. BACKGROUND AND SIGNIFICANCE

Malaria is an infectious disease transmitted by infective female Anopheles mosquitoes. According to the World Health Organization, in 2020 alone, an estimated 241 million clinical cases of malaria occurred, and 627,000 people died of malaria (11).

Though four species of human *Plasmodia* exist, *Plasmodium falciparum* is the malaria species that most often causes severe and life-threatening malaria (11,12). Humans and female Anopheles mosquitoes are infected cyclically in the natural history of malaria.

A. Parasite Life Cycle

Two hosts participate in the malaria parasite's life cycle. A malaria-infected female Anopheles mosquito injects sporozoites into the human host during a bloodmeal. Sporozoites released by these mosquitoes will infect liver cells, grow into schizonts then burst, releasing merozoites. Dormant stage, hypnozoites, in *P. vivax* and *P. ovale* can persist in the liver if it is untreated and trigger relapses by infecting the bloodstream weeks or even years later. The parasites undertake an asexual multiplication in the erythrocytes after initial replication in the liver. Merozoites are asexual stage parasites that are released by infected liver or blood cells and which go on to infect other red blood cells. Infected red blood cells parasites develop into trophozoite or ring stages, then can develop into schizonts, which rupture and release individual merozoites. Some infected red blood cells parasites develop into gametocytes, which are sexual erythrocytic stages. In humans, clinical signs of disease are observed during the blood stage of parasite development. To get infected, an Anopheles mosquito will ingest male (microgametocytes) and female (macrogametocytes) gametocytes during a blood meal. During the sporogony cycle, the microgametes penetrate the macrogametes in the mosquito's stomach, resulting in zygotes. The zygotes then become motile and elongated (ookinetes), invading the mosquito's midgut wall, and developing into oocysts. The oocysts mature, burst, and release

sporozoites, which enter the salivary glands of the mosquito (Figure 2). The malaria life cycle is perpetuated by injecting sporozoites into a new human host. Thus, the infected mosquito acts as a vector for the disease, while infected humans transmit the parasite to the mosquito. Unlike the human host, the mosquito vector is immune to the parasite, meaning that the mosquito is the parasite carrier without being affected by it (12,13).

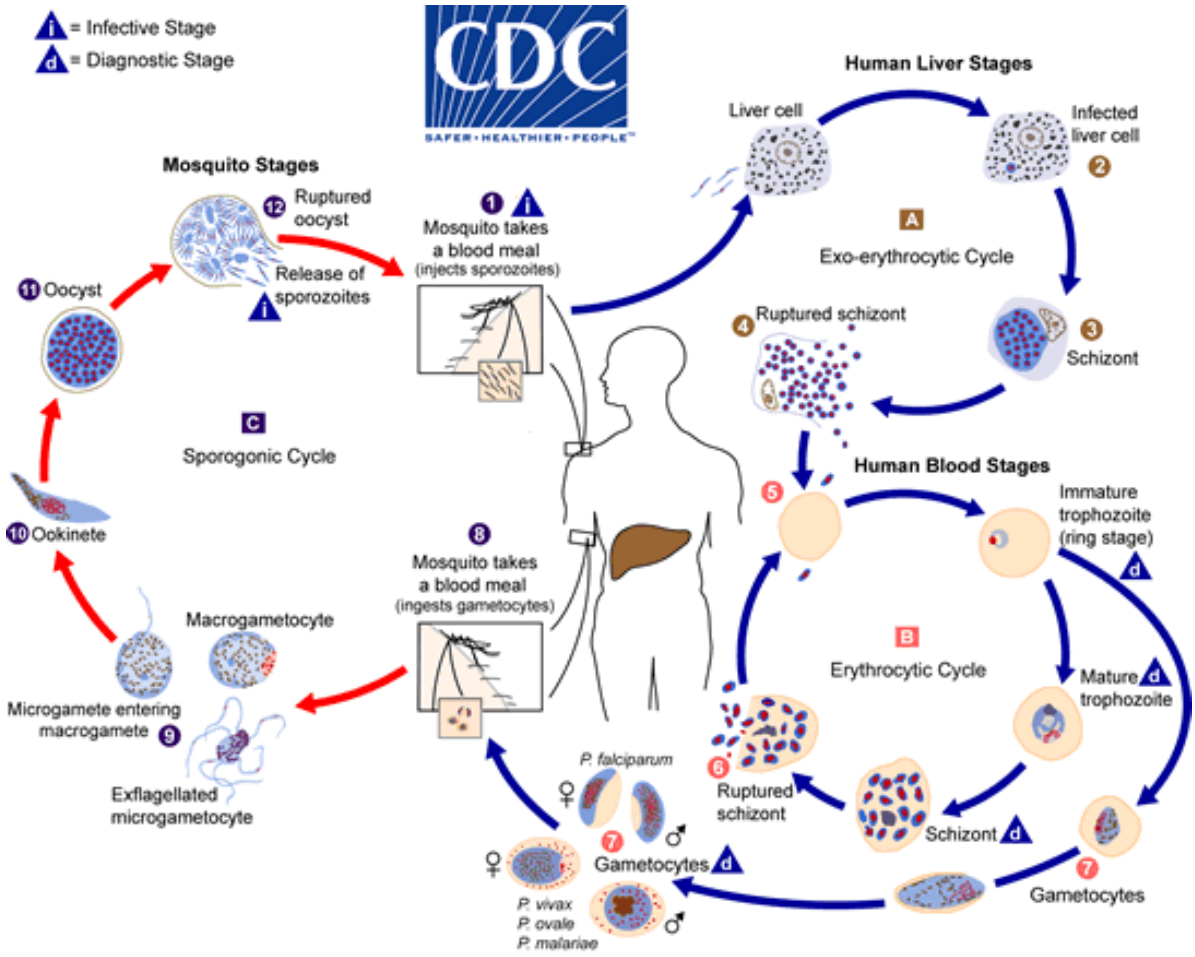

Figure 2: CDC Malaria transmission diagram. (Adapted from CDC Malaria (11))

Preventive methods have been developed to eradicate malaria. However, the parasites developed resistance to artemisinin-based combination therapies (ACTs) (14) and this,

along with mosquito resistance to insecticides, may be linked to increases in clinical malaria rates (15).

In addition, the disease incidence has stalled since 2015. Although this could be related to the various social and financial issues that endemic areas are facing, it could also be due to limited access to quality medical care, and to the inability of countries to fund their malaria control programs. These challenges, together with the increase in drug and insecticide resistance, support the idea that we will need an efficacious malaria vaccine to effectively achieve malaria elimination, as this disease not only affects the health of the impacted nation but also hinders them financially.

B. Vaccine against malaria

Previous malaria vaccine clinical trials showed that after immunization, vaccine-induced selection, and antigenic drift (the change in gene frequencies due to chance,) can impair overall vaccine efficacy in the long run (Figure 3) (5, 16). This is one of the main challenges for the current and developing vaccines.

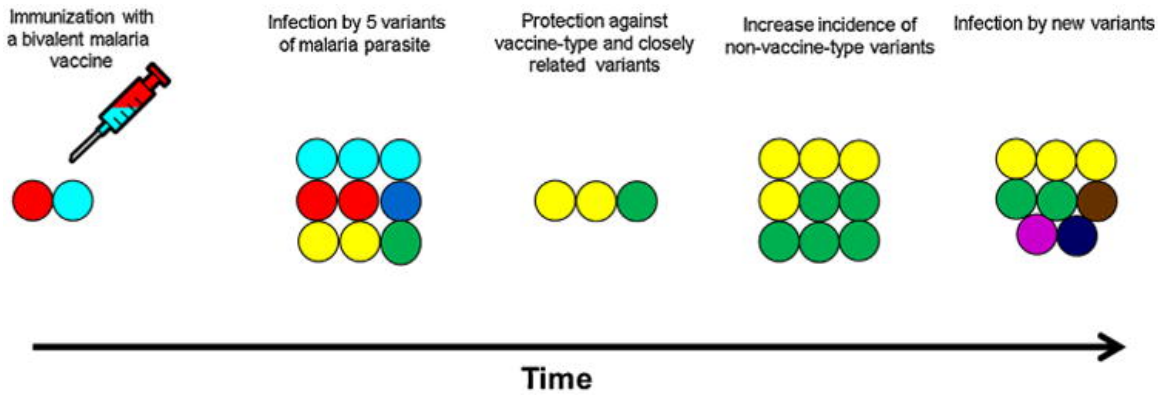

Figure 3: Effect of allele-specific efficacy on breakthrough infections. The malaria vaccine contains two variants (red and cyan) of a polymorphic antigen (A). When an individual is infected by five different parasites with five variants (cyan, red, dark blue, yellow and green) of the same proteins (B), malaria clinical episodes with homologous (red and cyan) and closely related variants (dark blue) are prevented (C). Heterologous variants (yellow and green) may increase within the individual and the whole population (D, E).(Adapted from Ouattara et al, 2015. (5))

Malaria vaccines candidates are being developed to improve the efficacy of RTS,S, and/or to help eradicate the disease. RTS is the only vaccine that has been authorized for use, while research is ongoing to improve its effectiveness.

1. Pre-erythrocytic vaccines

Pre-erythrocytic vaccines, also known as anti-infective vaccines, are designed to specifically target the parasite during the liver stage (figure 4). These vaccines may target proteins on the surface of sporozoite before the invasion of the hepatocyte (motility and

transversal blockade) or by preventing the development of the parasite within the hepatocyte (destruction of infected hepatocytes). This could be done through specific antibodies that kill infected liver cells or interfere with the malaria parasite during liver cell proliferation, preventing the release of infectious merozoites and providing anti-infection protection. Proteins that are the target of pre-erythrocytic malaria vaccines include the *Plasmodium falciparum* circumsporozoite protein (*Pf*CSP) (17), the thrombospondin-related adhesion protein (TRAP), and the liver stage antigen (LSA) (17).

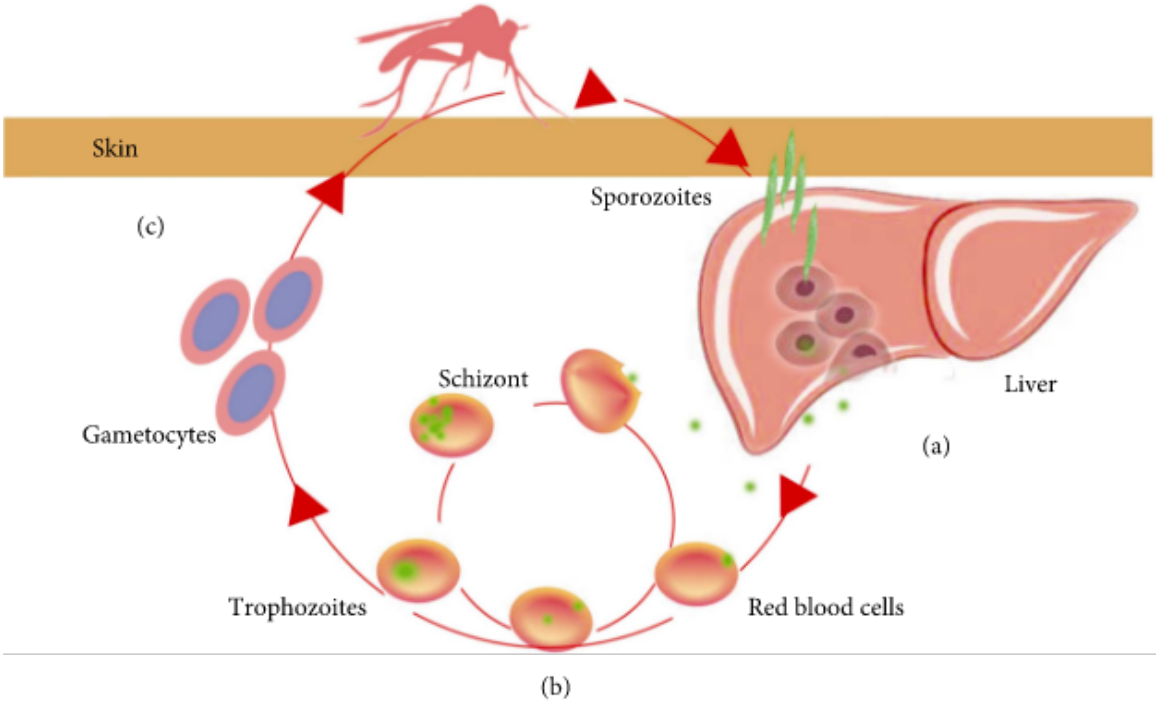

Figure 4: Vaccines target the life cycle of Plasmodium falciparum. (a) Pre-erythrocytic malaria vaccine: vaccine produces antibodies that kill infected liver cells or interfere with the malaria parasite during liver cell proliferation; (b) asexual blood-stage vaccine: vaccine aims primarily at reducing parasite load or eliminating circulating parasites; (c) vaccine aims at controlling the transmission of malaria

parasites from human hosts to the mosquito vectors. (Adapted from Zheng et. al, 2019. (17))

Plasmodium falciparum circumsporozoite protein (*Pf*CSP), a 40–60 kDa pre-erythroid antigen found on the surface of mature sporozoites, plays an important role in sporozoite invasion of liver cells (17). RTS,S/AS01, the only licensed malaria vaccine, targets immunogenic epitopes of the protein, which is encoded by the circumsporozoite (CS) gene. The CS gene has polymorphisms in areas that encode epitopes recognized by the human immune system (8). The RTS,S/AS01 vaccine comprises a segment of the core NANP-NVDP repeat, polymorphic B-cell epitope areas and a highly polymorphic C-terminal non-repeat epitope region that induces cell-mediated immunity. Th2R and Th3R, the 3′ sections of the CS gene, encode epitopes that CD8+ and CD4+ T-cells recognize. In natural parasite populations, several studies have found increased polymorphism in T-cell epitopes within the C-terminal region of CSP (6,8).

In field studies, CSP has been shown to be remarkably diverse. In an incidence study conducted in a Navrongo and Cape Coast, two sites in Ghana where RTS,S is being tested in a phase 4 trial, Th2R and Th3R epitopes showed differences in amino acid composition, with 1 to 6 amino acid variations in the Th2R and Th3R epitope regions, compared to the vaccine strain (6). Moreover, the study shown that only 5.9% and 45.7% of the Navrongo and Cape Coast sequences, respectively, had the CSP vaccine strain C-terminal epitope. Similarly, in a study conducted in Bandiagara, Mali, T-cell epitope-encoding regions Th2R

and Th3R were found to be polymorphic though the distribution were stable across seasons, age groups, and clinical and asymptomatic illnesses (8).

In contrast to the RTS vaccine, PfSPZ Vaccine is a live, radiation-attenuated, whole organism pre-erythrocytic malaria vaccine. For this whole sporozoite vaccine, attenuated sporozoites are injected into the blood stream. These sporozoites will grow in hepatocytes and trigger an immune response that targets subsequent sporozoites. Compared to subunit vaccines, vaccination with attenuated sporozoites provides a greater range of immunogens spanning at least two parasite life cycle stages. There are three whole sporozoite vaccines in development including radiation attenuated sporozoites (RAS), genetically attenuated parasites (GAP) and wild type *Pf* sporozoites administered under chemoprophylaxis (PfSPZ-CVac) (17,18). Malaria vaccine studies conducted with controlled human malaria infection have demonstrated higher protection against homologous versus heterologous parasite strains, which implies that there is a strain-specific vaccine-induced protection. This strain-specific effect a potential reason why PfSPZ Vaccine had an inadequate overall efficacy in clinical trials (19).

2. Erythrocytic vaccines

T cells and B cells have been demonstrated to provide protection against malaria infection at the blood stage. For instance, T-cell-depleted animals that were infected with parasites, had lower levels of IFN- and IgG2a than in control mice (20). Limiting blood stage infection can limit transmission of the disease as it will inhibit erythrocyte invasion and

therefore reduce the blood-stage parasitemia. As a result, blood-stage malaria vaccines have become a focus area for vaccine development (17).

The *P. falciparum* merozoite surface protein (MSP) can persist in the red blood cell membranes after entering the human red blood cells from the liver, where they are exposed the host's circulating immune responses (Figure 4) (17). MSP1 and MSP2 are two major *P. falciparum* blood-stage malaria vaccine targets that are diverse in the nature. The proteins participate in erythrocyte invasion and are the target of the immune responses. MSP1 is a 180 to 210 KDa surface protein that is encoded by the msp1 gene on chromosome nine and consists of 17 blocks of sequence flanked by conserved regions (21, 22). On the other hand, MSP2 is a glycoprotein encoded by the msp2 gene on chromosome two that is made up of five blocks, the central block being the most polymorphic. The MSP2 alleles are divided into two groups, the FC27 and 3D7/IC1 families (21). In malaria-endemic nations, both MSP1 and MSP2 have been demonstrated to be highly polymorphic. In *P. falciparum* isolates from symptomatic malaria patients in Bobo-Dioulasso, a substantial genetic diversity has been observed that has limited efficacy of MSP-based candidate malaria Vaccines to date (21).

The apical membrane antigen 1 (AMA1) is perhaps the most tested blood stage malaria vaccine candidate (5,20,21,23,40). AMA1 is expressed in the sporozoite, hepatic, and erythrocytic stages where it plays an essential role in parasite survival (17). In a study conducted in Mali in 2009, 214 distinct AMA1 haplotypes were identified in 506 human

infections, and amino acid variations around a putative invasion machinery binding location were highly linked to the development of clinical symptoms (24).

3. Transmission-blocking vaccines

Transmission-blocking vaccines (TBVs) are used to prevent malaria parasites from spreading from mosquito vectors to human hosts by targeting the sexual and sporogony stages of the parasite (Figure 4). TBVs do not provide direct protection against infection. In contrary, they seek to prevent an infected human from transmitting malaria parasites to a feeding mosquito, i.e., preventing parasites from successfully infecting the mosquito (25).

TBV antigens are split into two groups: pre-fertilization antigens and post-fertilization antigens. Pfs48/45, Pfs47, and Pfs230 are pre-fertilization antigens found on the surface of malaria parasite gametocytes and gametes. These proteins are members of a family with six cysteine domains. Pfs230 and Pfs48/45 are gamete surface antigens that elicit antibody responses in naturally exposed people and are linked to transmission-reducing immunity. Pfs25 is a post-fertilization antigen that is expressed on the surface of the zygote and ookinete, and it has a high immunogenicity and a low antigenic polymorphism. The appropriate folding of cysteine-rich proteins and the precise production of disulfide bridges are required by antibodies that only target conformational epitopes of these proteins (25). These sexual stage antigens elicit antibodies in the human host, which inhibit parasite development (25). Thus, transmission inhibition occurs within the mosquito vector and is mediated by antibodies (17,25).

Pfs25 is regarded as one of the most important transmission blocking vaccine candidate antigens (25). Experiments in gene knockout mice indicate that this protein is required for the parasite to survive inside the mosquito midgut (25). Furthermore, a double-knockout study on *P. berghei* reveals that removing the P25 antigen reduces ookinete invasion into midgut epithelial cells (25). The advantages of post-fertilization antigens such as Pfs25 are their ability to induce a durable immune response and low antigenic variation (25).

4. Pregnancy Associated Malaria Vaccines

Malaria is a fatal disease that affects pregnant women and newborns. In malaria-endemic areas of the world, pregnant women who get the disease suffer significant health problems as well as pregnancy difficulties (26). Furthermore, *Plasmodium falciparum* parasite-infected blood cells tend to aggregate in the placenta, negating the childhood immunity that people build in malaria-endemic areas, which would normally lessen the disease's effects. Though there are many vaccines being developed, no malaria vaccine has ever been tested in pregnant women (26).

In a study published in 2020, researchers evaluated the safety and immunogenicity of PRIMVAC adjuvanted with Alhydrogel or glucopyranosyl lipid adjuvant in emulsion GLA-SE (emulsion) in non-pregnant French and Burkinabe women who were enrolled in the trial between 2016, and 2017, (27). PRIMVAC is a placental malaria vaccine candidate generated from the VAR2CSA gene that aims to avoid clinical consequences from

Plasmodium falciparum infection during pregnancy. According to the findings, PRIMVAC exhibited a favorable safety profile, was immunogenic, and produced functional antibodies interacting with the homologous VAR2CSA variant expressed by NF54-CSA infected erythrocytes. (26,27)

C. CSP

CSP is a protein secreted by the malaria parasite's sporozoite stage. CSP is embedded in the plasma membrane via a Glycosylphosphatidylinositol (GPI) anchor at its C-terminus, allowing it to reach the extracellular space (4). The protein is made up of an immunodominant core repeat region flanked by conserved motifs of an N-terminal domain, a tandem repeat region, and a C-terminal domain, all of which are thought to be involved in protein processing as the parasite moves from the mosquito to the mammalian carrier. CSP is required for the formation of sporozoites in the mosquito midgut, the release of sporozoites from the oocyst, the invasion of salivary glands, the attachment of sporozoites to hepatocytes in the liver, and the invasion of hepatocytes by sporozoites (4).

In its structural state (figure 5), the CSP C-terminal cell-adhesive domain is exposed and adhesive, whereas the non-adhesive conformation of the N terminus masks the C-terminal domain while sporozoites are in the mosquito salivary glands. The cell-adhesive domain's N-terminal masking keeps the sporozoite migratory during sporozoite formation and hepatocyte invasion. The flip to an adhesive conformation in the mammalian host is

controlled by proteolytic cleavage of CSP, and the highly conserved region I is important in this process (2). Primary contacts between the N-terminus of CSP and heparin sulfate proteoglycans (HSPGs) arrest sporozoites on the hepatocyte surface in the mammalian liver. The TSR domain is revealed by subsequent cleavage of CSP in region I. The TSR domain has cell adhesive properties and is thought to mediate the secondary attachment of sporozoites to hepatocytes, resulting in sporozoite invasion of hepatocytes (2, 4).

Different regions of CSP have been revealed to be highly polymorphic (6,8,28). The tetrameric central repeats, which can vary in sequence and number of repeats, are found in these polymorphic locations. It is hypothesized that the length of the repeat affects the protein's stability, which in turn affects the B-cell epitopes. The Th2R and Th3R regions encode epitopes that are recognized by CD8+ and CD4+ and can be expressed as non-synonymous single nucleotide polymorphism, SNPs. The variation in these regions grows as malaria transmission spreads across different geographic areas with Africa having the most diversity (8).

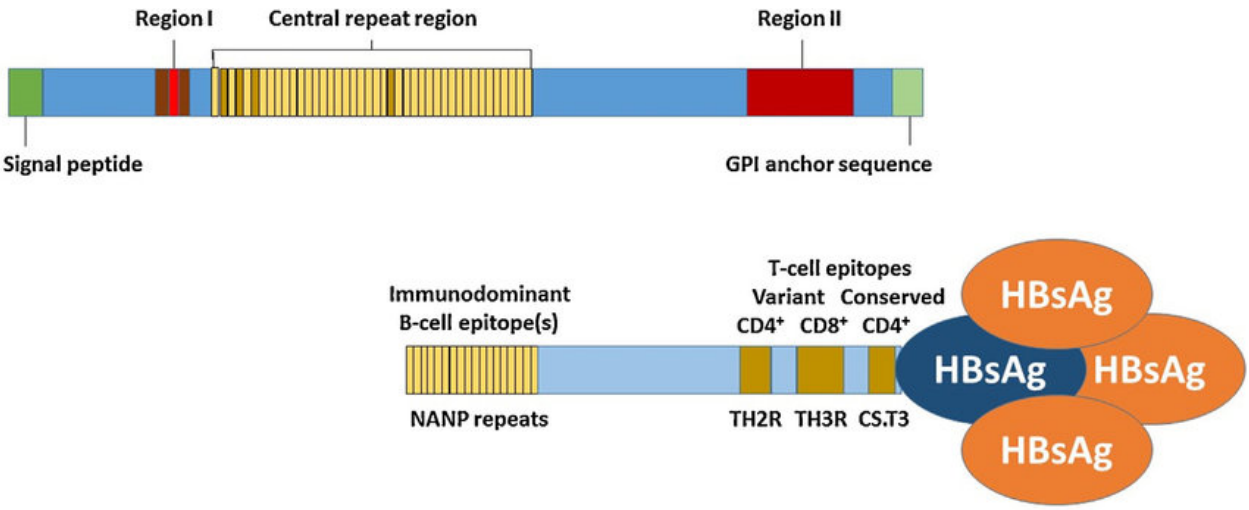

Figure 5: Graphical depiction of circumsporozoite (CSP) and RTS,S structures. (Adapted from Kaslow et al, 2015. (41))

CSP can be found in all *Plasmodium* species. Though its amino acid sequence varies between species, the overall domain structure is well conserved (2) which may explain its choice as a malaria vaccine candidate.

D. PacBio SMRT

The key technology for long-read sequencing platforms is Single Molecule, Real-Time (SMRT) sequencing. This groundbreaking approach is today's technology utilized across the board in the biological sciences.

PacBio sequencing captures sequence information while the target DNA molecule is replicating. SMRTbell is a closed, single-stranded circular DNA template formed by ligating hairpin adaptors to both ends of a target double-stranded DNA (dsDNA) molecule. When an SMRTbell sample is loaded into a SMRT cell, it diffuses into a sequencing unit known as a zero-mode waveguide (ZMW), which has the smallest available volume for light detection. At the bottom of each ZMW, a single polymerase is immobilized, which can bind to either hairpin adaptor of the SMRTbell and initiate replication. The SMRT cell receives four fluorescent-labeled nucleotides with distinct emission spectra. As the polymerase holds a base, a light pulse is produced that identifies the base. A "movie" of light pulses records the replication processes in all ZMWs of a SMRT cell, and the pulses corresponding to each ZMW can be interpreted as a sequence of bases (called a continuous long read, CLR) (29).

Due to its ability to generate lengthy reads, PacBio SMRT has the potential to create full sequences. PacBio sequencing is especially appealing because samples may be multiplexed, allowing for high throughput. Furthermore, the assembly does not necessitate the use of a reference sequence, making it ideal for a highly polymorphic gene including CSP. Complex loci in humans and *P. falciparum* have been sequenced using targeted PacBio sequencing (30).

III. RESEARCH DESIGN AND METHODS

A. Overall study design

Sequences used in this study were generated from samples collected during a malaria transmission season conducted in Kalifabougou, Mali. To generate long read covering the repeats region in addition to the N- and C-terminal the PacBio SMRT platform was used. Sequences were separated by barcodes, assembled by de novo assembly, and aligned against the 3D7 reference strain.

B. Study site & Sample origin (30)

Samples utilized in this study are comprised of 59 sequences generated from samples collected during a single malaria transmission season conducted in Kalifabougou, a rural community located 45 kilometers northwest of Bamako, Mali. Samples were collected at

the beginning of malaria transmission in May 2011 from healthy children and adults, ranging in age from 3 months to 25 years. The samples were collected from May 2011 to January 2012.

C. DNA extraction and PCR (6, 30)

Malaria parasite DNA were extracted using a Qiagen kit according to the manufacturer protocol. Whole-blood gDNA was extracted from blood samples and DBSs using the QIAamp 96 DNA kit (Qiagen, Valencia, CA). To genotype polymorphisms in *Pf* CSP, a PCR was used to amplify the major exon in all the samples. Primary and secondary primers were designed to cover the whole encoding gene of CSP. The Secondary PCR (BioRad) products were run on a E-gel to assess the presence of malaria parasite DNA.

Primers CSP-F1 (5′ GTAGAAACCACGTATATTATAAATT 3′) and CSP-R1 (5′ ATATTGTACAACTCAAACTAAGATG 3′) were used in the primary PCR reaction. Conditions for this amplification were 94°C for 2 minutes, followed by 35 cycles of denaturation at 94°C for 20seconds, primer annealing at 48°C for 20second, and extension at 60°C for 90 seconds with a final extension at 60°C for 5 minutes. All PCR products from the first amplification underwent a new secondary PCR with the secondary PCR primer to which sample specific barcodes were attached. The primers were CSP-F2 (5′ ATGATGAGAAAATTTAGCTATTTTAT 3′) and CSP-R2 (5′ CTAATTAAGGAACAAGAAGGATAAT 3′). Amplification conditions were the same as in the primary amplification with only 34 cycles instead of 35. PCR reactions were

carried out in a total volume of 25 μL with 23.75μL of Platinum Super Mix in 96-well plates (Invitrogen, Carlsbad, CA), 1 μL of each primer (1 μM final concentration) and 1 μL of malaria parasite DNA. Secondary PCR also had 23.75 μL of Platinum Mix, 1 μL of each (1 μM final concentration) primer and 1 μL of primary PCR product. Secondary PCR products were electrophoresed on a high-throughput precast gel system (Invitrogen), then visualized and photographed under ultraviolet light with ethidium bromide as the stain.

After running a E-gel to confirm samples positivity, parasite DNA were purified using a Millipore filter plate. DNA were quantified using PicoGreen following the manufacturing protocol. An equimolar multiplex DNA solution comprised of the 48 individual sample was made and send to Institution for Genome Sciences (IGS) for sequencing using PacBio SMRT.

D. PacBio SMRT (31)

PacBio was utilized to sequence the CSP data used in this study. The amplicons were sequenced using PacBio sequencing after PCR primers were designed to target a 5 kb sequences covering the whole CSP protein.

The SMRT-analysis consensus tools package was used to put together sequence readings. Purified, pooled amplicons were sequenced on a PacBio RS II sequencer (Menlo Park, Pacific Biosciences, California, USA) at the Genomics Resource Center of the Institute for

Genome Sciences, University of Maryland School of Medicine, USA. SMRTbells™ adapters were ligated to the barcoded-pooled amplicon template to create libraries. P4-C2 chemistry (P4 polymerase with C2 sequencing chemistry) was used in one SMRTTM cell. The readings were generated using a 180-minute video on the PacBio RS II.

The secondary analysis was carried out with the PacBio SMRT Analysis v2.3.0 package, utilizing the long amplicon analysis (LAA) algorithm and the following parameters: minLength 3250—minSnr 4. The barcode de-multiplexed the pooled amplicon data. To filter high-quality consensus sequences for downstream analysis, a projected accuracy of 95% was employed, which is defined as the threshold below which a haplotype (consensus sequence) is considered noise.

E. Human subject protection

The Intuitional Review Board (IRB) of the University of Maryland at Baltimore, as well as the University of Sciences Techniques and Technology of Bamako Mali, and the National Institute of Allergy and Infectious Diseases all evaluated and approved the study protocols. Participants in the study gave their consent and/or assent to take part in the investigation. The data utilized in this investigation was already de-identified.

IV. RESULTS

A. Overall summary of results

Overall, all 48 samples were sequenced by PacBio SMRT platform resulting into 59 individual sequences. Samples had on average a 295X coverage and a 95% accuracy. Haplotype diversity at the whole gene level was one. *Pf*CSP for both 3D7 and 7G8 sequences were used as a reference in the analysis.

B. Haplotype polymorphism

Haplotype definitions were based on previously conducted studies. T cell epitope haplotypes were determined based on a study conducted in Bandiagara, Mali (8), whereas the B cell epitope haplotypes were determined based on a study conducted in 1993 (32,33).

Briefly haplotypes were defined based the N terminus (21aa), the C Terminus (93aa (34)), the Th2R (17aa) and Th3R (24aa) sections of the gene, and the NANP/NDVP tetrameric Repeat (168aa) of the *Pf*CSP sequence, to access the overall samples genetic diversity. The study samples displayed great variability in the C Terminal, with a total of 43 haplotypes, of which 10 were the most prevalent, figure 7. For other *Pf*CSP regions, N Terminal, Th2R, ThR3, and Repeats, 3 haplotypes, 8 prevalent haplotypes (total of 35 haplotypes), 7 prevalent haplotypes (total of 22 haplotypes), and 6 haplotypes respectively were identified.

C. N Terminal and C Terminal

Using the 3D7 *Pf*CSP strain as a reference, we observed 5 to 52 amino acid variations in the C terminal region, figure 6. Most of the CSP sequences in our sample had 13 mutations in their C Terminal region, followed with 8 and 10 mutations. The sites with the most mutations were in the Th2R and Th3R regions.

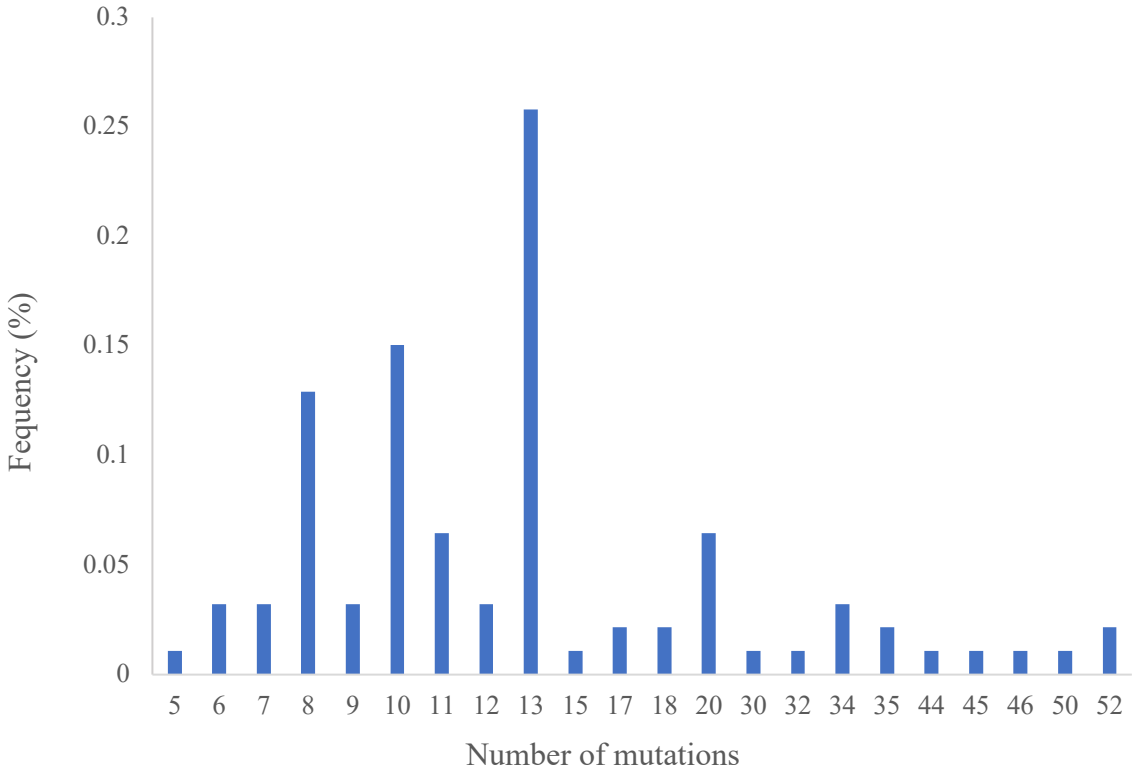

Figure 6: The frequency of mutation at the C Terminus. Amino acid mutations at each sequence site have been calculated using the 3D7 *Pf*CSP strain sequence as reference.

In the C-terminal region, 10 major haplotypes were identified. Haplotype A, B, C and D were the most prevalent with frequencies of 8.3%, 6.7%, and 5.0%, respectively. The 3D7 *Pf*CSP haplotype was not among the most prevalent haplotypes and had a frequency of 1.7%, figure 7.

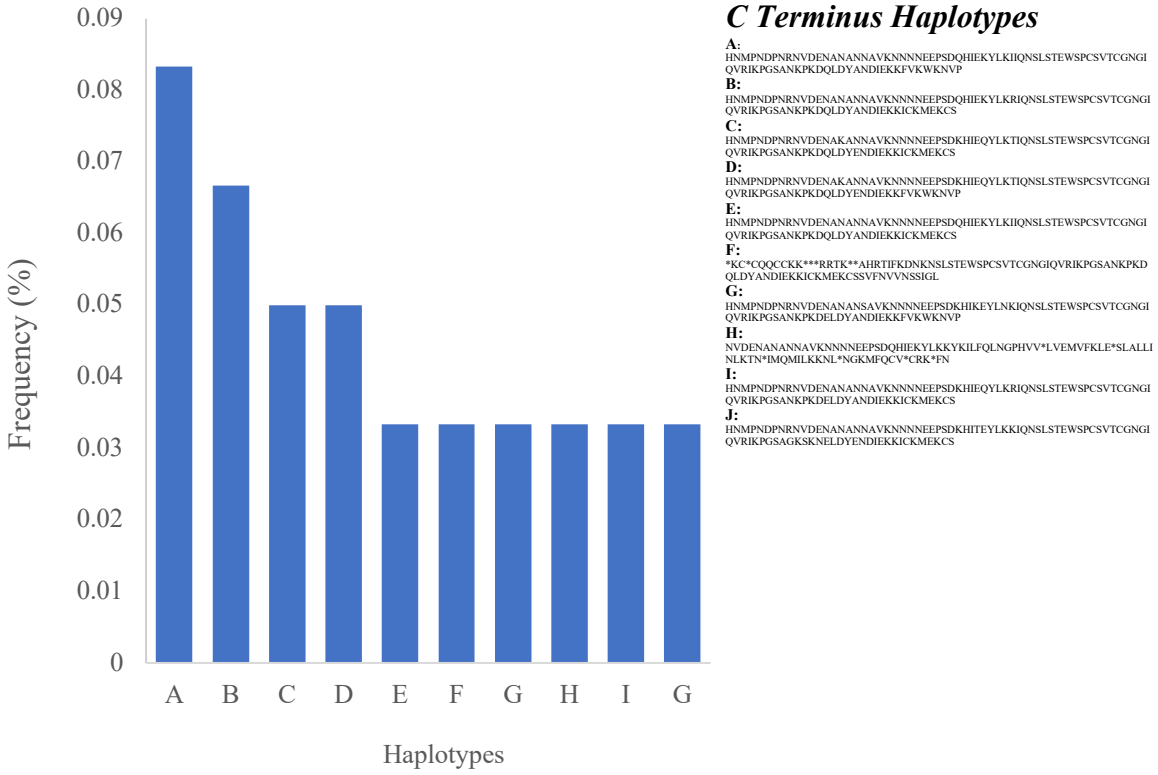

Figure 7: Frequency of the C-terminal haplotypes.

Using the 3D7 strain as a reference sequence, we found that the N-terminal region of Kalifabougou sequences, were mostly similar to that of the 3D7 *Pf*CSP sequences, figure 8. The CSP sequences that differed from the reference sequence had a mutation at position 98 where an adenine mutated to glycine.

Figure 8: The frequency of mutation at the N-Terminal region of CSP encoding gene. Amino acid mutations at each sequence site have been calculated using the 3D7 *Pf*CSP strain sequence as reference.

In the N-terminal region, three major haplotypes were identified. Polymorphisms were primarily in the last few amino acids of the N-terminal region. Figure 9 shows that haplotype A, which is also the haplotype found in 3D7 *Pf*CSP, had a frequency of 75.0%, while haplotype B had a frequency of 23.3%. Finally, the frequency of haplotype C was 1.7%, figure 9.

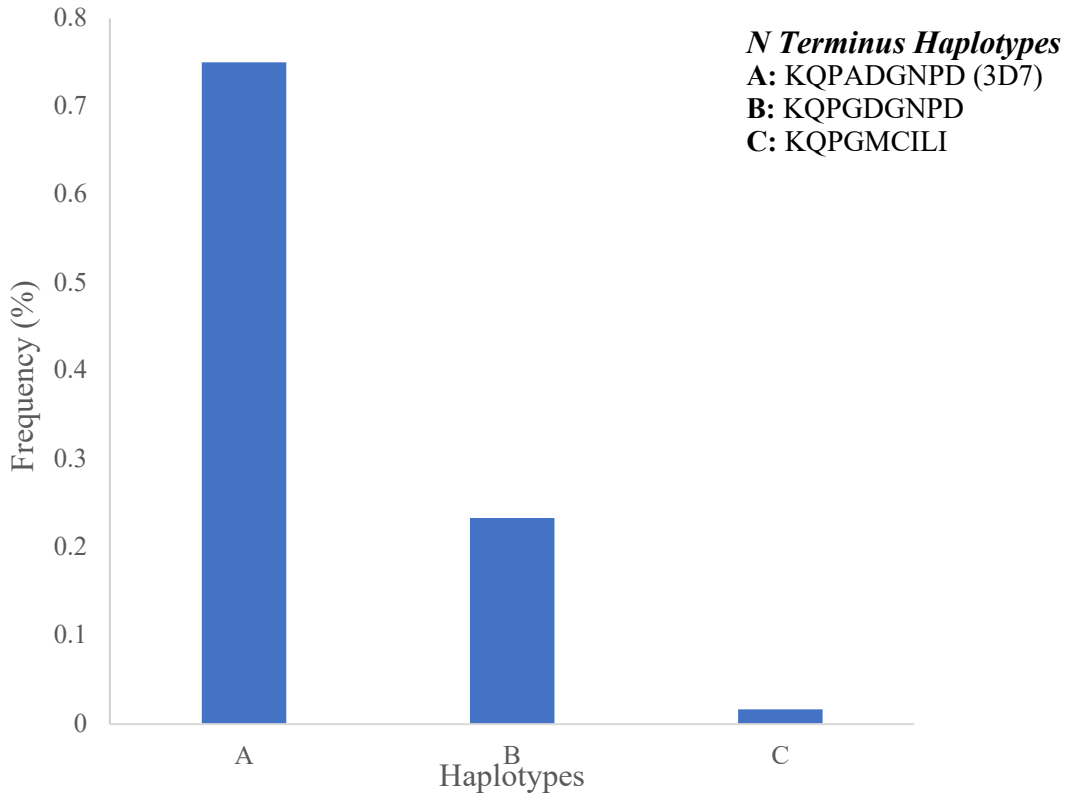

Figure 9: N-Terminal haplotypes frequency

D. T cell Epitopes Diversity

With 35 and 22 haplotypes, respectively, the Th2R and Th3R sites were the most polymorphic regions of the *Pf*CSP analyzed.

Eight Th2R predominant haplotypes out of a total of 35 were identified in the Kalifabougou dataset. Within the eight haplotypes, three of the haplotypes were most prevalent within the sequences at the Th2R site. The most common haplotypes were found in 13.3% (Haplotype A) , 11.7% (Haplotype B), and 10.0% (Haplotype C) of our samples, figure 10.

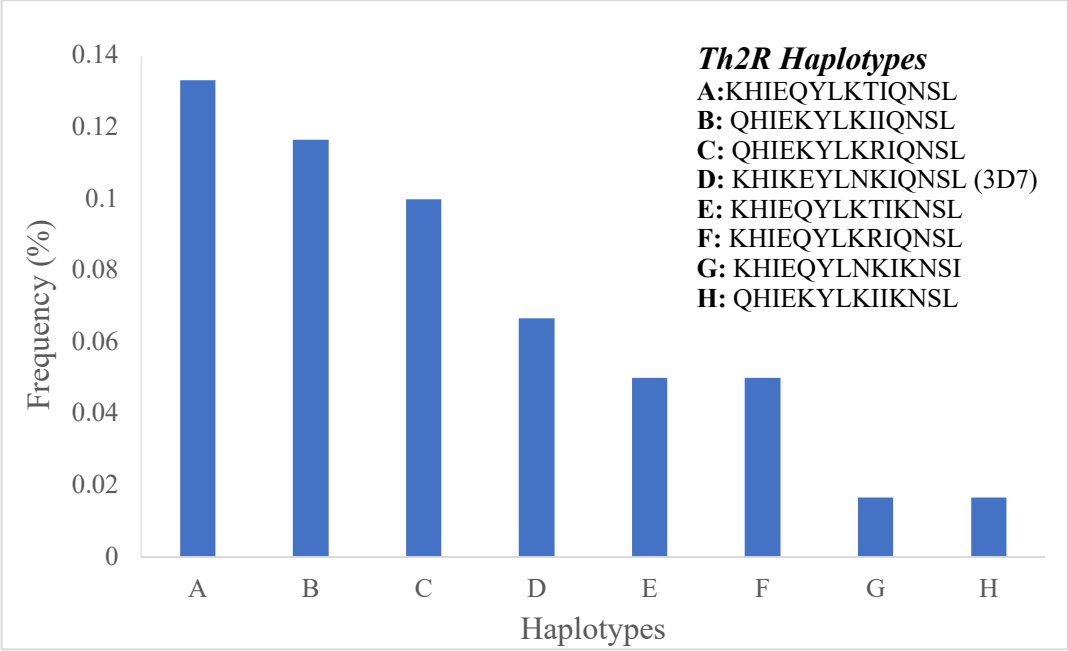

Figure 10: Distribution of Kalifabougou Th2R haplotypes (17aa).

In the Th3R region, seven major haplotypes were determined. Of which, the most common haplotypes were present at the frequencies of 31.7% (A), 15% (B), and 13.3% (C), which is the 3D7 *Pf*CSP-Type TH3R haplotype, figure 11.

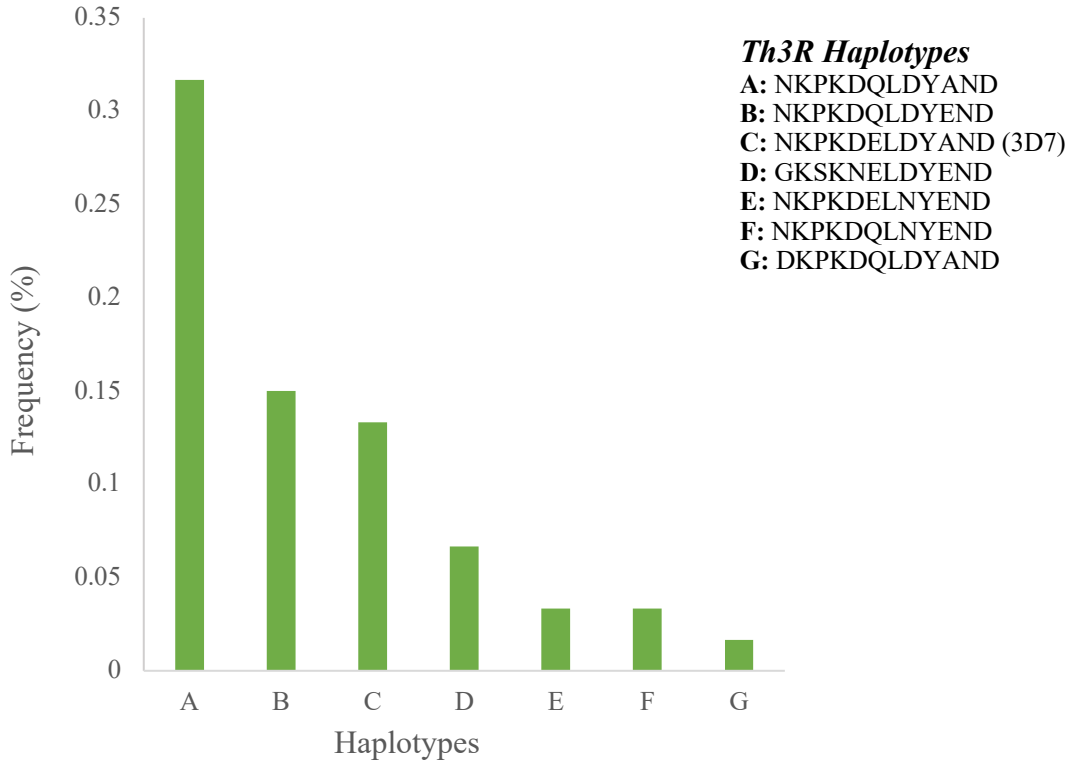

Figure 11: Distribution of Th3R haplotypes in Kalifabougou, Mali.

In addition, we investigated the diversity in amino acids at the TH2R and TH3R regions between our samples and the 3D7 reference strain and found three instances of a perfect match at the analyzed sites. Besides, most amino acids are conserved in 80.0% of the cases between our samples and reference strain, while most variations were found in the amino acids lysine, asparagine, and glutamic acid, figure 12.

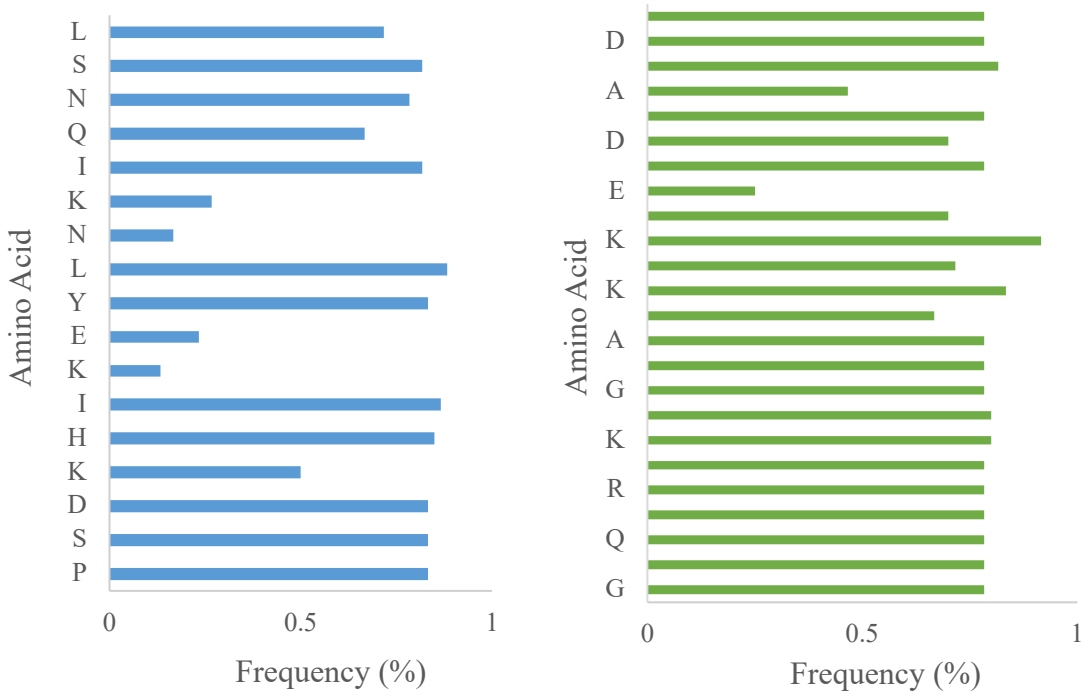

Figure 12: Frequency of Amino acid polymorphisms within Kalifabougou TH2R and TH3R sequences.

E. B-cell epitopes diversity

To access the diversity of B cell epitopes, we evaluated the frequency of DV10 haplotypes (positions 296-303aa), which has been shown to be the both T- and B-cell epitopes (28,29). There were four major haplotypes (total of 11 haplotypes) identified, with NANANNAV (A) accounting for 53.3% of the haplotypes, 15.0% for NAKANNAV (B) haplotype, 13.3% for NANANSAV (C) haplotype which is the haplotype found in 3D7 *Pf*CSP, and 3.3% for NNNNEEPS (D) haplotype, Figure 13.

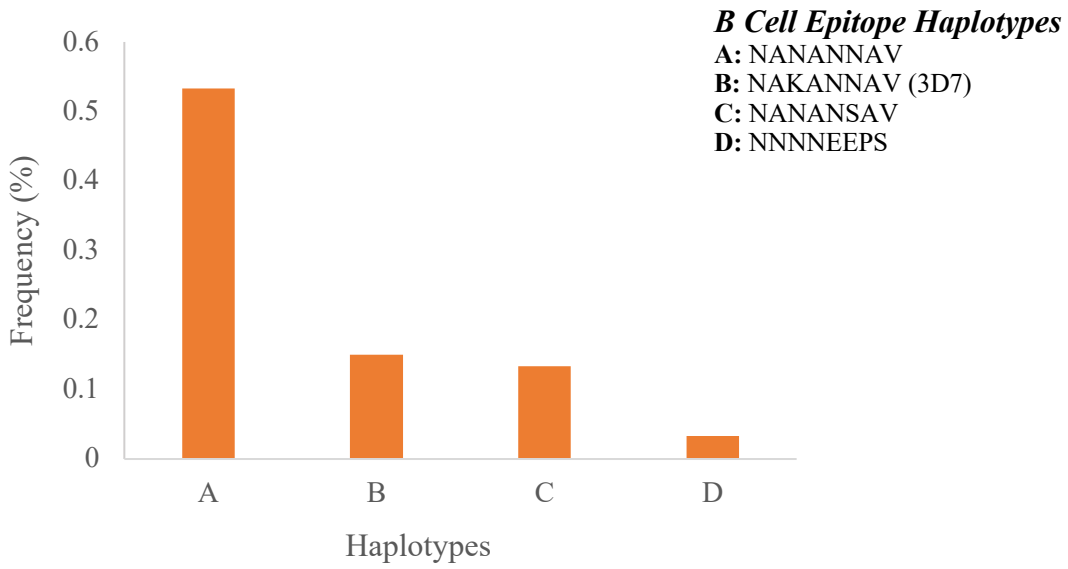

Figure 13: B cell Epitope haplotype frequency in Kalifabougou, Mali. Haplotypes (8aa) were determined using amino acids in the DV10 region (position 296-303).

F. Polymorphism in the Repeat Region

The frequency NANP/NVDP tetrameric repeats, as well as the number of NANP or NVDP repeats were accessed to determine the diversity within *Pf*CSP repeat region.

Six haplotypes were determined based on the length of the tetrameric repeats. Two of the haplotypes were the most common, with frequencies of 30.0% and 28.3% for NANP/NVDP repeat lengths of 40 and 39 tetrameric repeats, respectively. 3D7 *Pf*CSP sequences had the longest tetrameric repeats (43), figure 14.

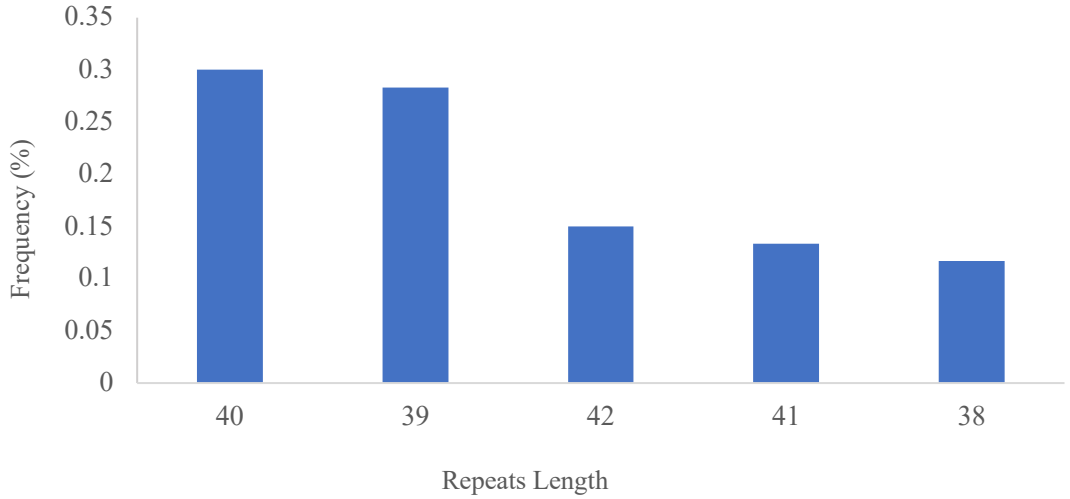

Figure 14: Total number of NANP/NVDP Tetrameric Repeat region (168aa) and their frequencies in Kalifabougou, Mali.

By assessing the length of the NANP or NVDP separately, we found that the number of NANP repeats were variable across samples. However, except for four samples with three NVDP sequences and one with no NVDP in the repeat region, most repeats had four NVDP sequences, figure 15 and 16.

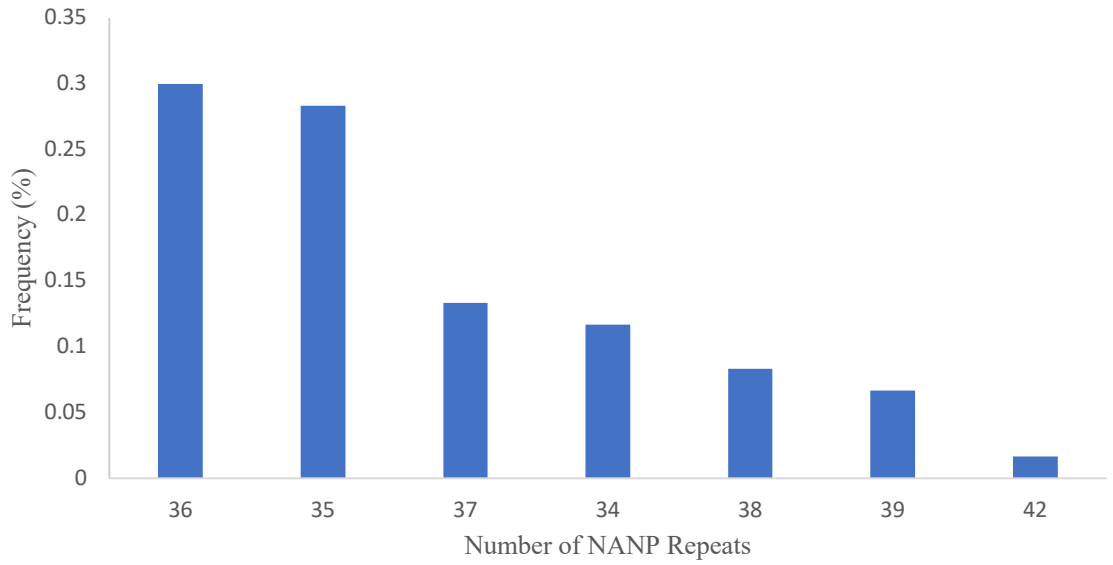

Figure 15: NANP tetrameric repeats frequency in Kalifabougou, Mali.

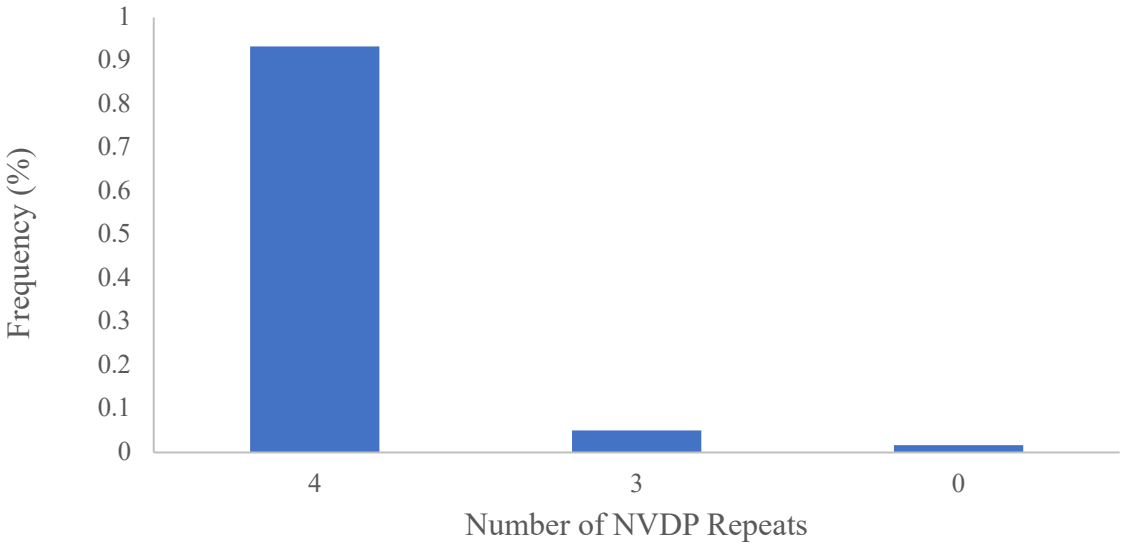

Figure 16: NVDP tetrameric repeats frequency in Kalifabougou, Mali.

G. Assess the variation of the major Haplotype from Kalifabougou compared to 7G8

Finally, with the CSP sequences in our dataset being different from the Pf3D7 CSP in terms of haplotypes, we decided to compare our sequences to the *P. falciparum* 7G8 strain (*Pf*7G8) as a reference strain from the Americas, to access whether there was a difference in diversity compared to *Pf*3D7 CSP.

After Data analysis, the findings suggest similarities between the CSP sequences of our study samples and the 7G8 *Pf*CSP was withing the haplotypes of B cell epitopes, the N Terminus and the number of tetrameric repeats. The C Terminus haplotypes derived from our dataset were different from those of the reference 7G8 *Pf*CSP.

We then determined the frequency of the CSP haplotypes based on the length of the NANP/NVDP tetrameric repeats of each sequence of our samples compared to 7G8 *Pf*CSP. For the length of the repeat region, four haplotypes were identified, with two of the haplotypes being the most common, with frequencies of 30.0% and 28.3% for lengths of 40 and 39 tetrameric repeats, respectively. There were 41 tetrameric repeats in the 7G8 *Pf*CSP sequence which was identical to nine other CSP sequences in our dataset (15.0% frequency), figure 17.

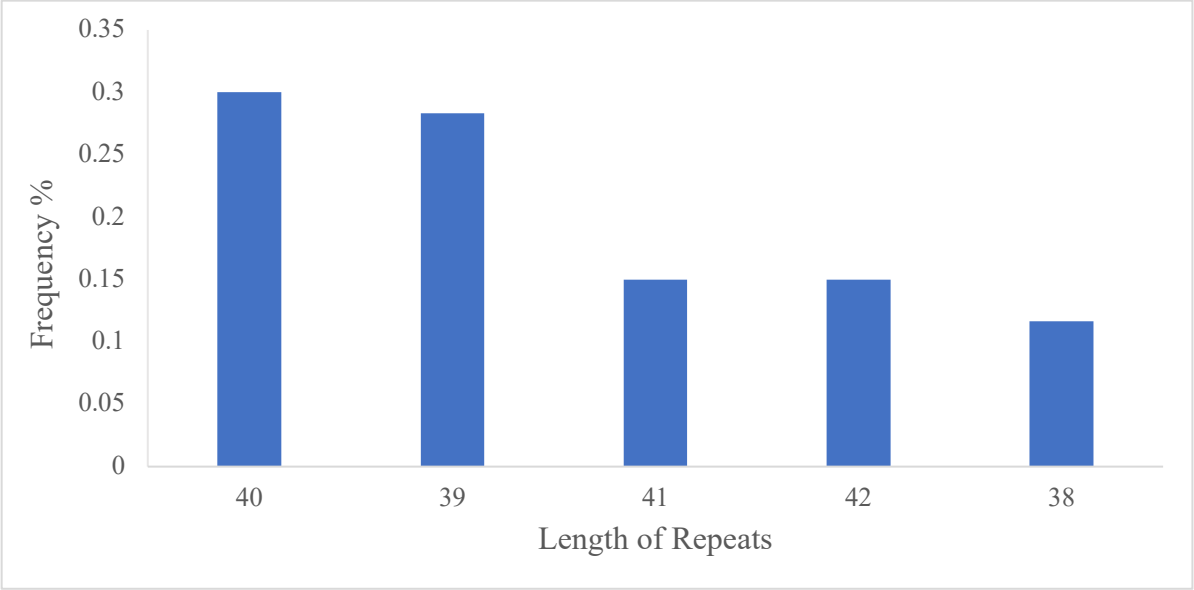

Figure 17: Total number of Pf7G8 NANP/NVDP Tetrameric Repeat region (168aa) and their frequencies in Kalifapougou, Mali.

In addition, the three haplotypes identified in the N Terminal include KQPADGNPD which had a frequency of 73.3%, KQPGDGNPD which was the also the haplotype found in 7G8 *Pf*CSP sequence had a frequency of (25.0% frequency), and KQPGMVILI with the lowest frequency of 1.7%. Here the haplotype analysis found that the 7G8 *Pf*CSP had a lower frequency 25.0% in the N Terminal region than the 3D7 *Pf*CSP haplotype which had a frequency of 75.0%, figure 18.

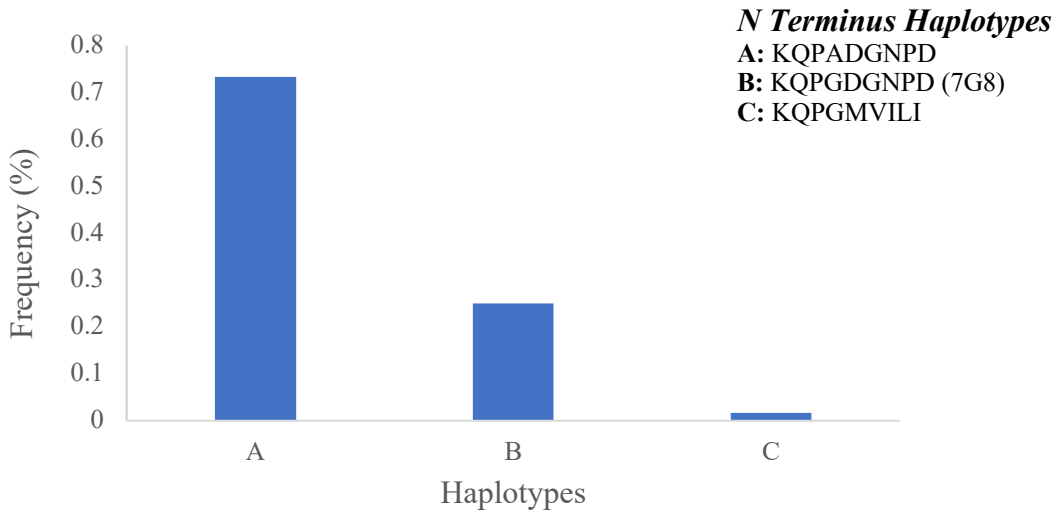

Figure 18: Pf7G8 N-terminal haplotype frequency

Finally, the B cell epitope region yielded 4 haplotypes, with NANANNAV (7G8 haplotype) being the most common (55.0%), while NAKANNAV, NANANSAV and NNNNEEPS had frequencies of 15.0% and 11.7 % respectively, figure 19.

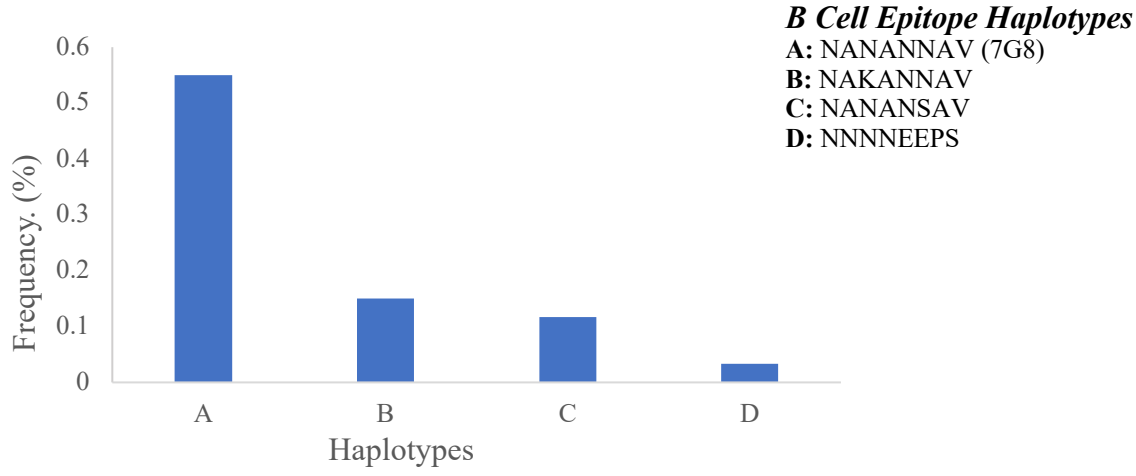

Figure 19: Pf7G8 B Cell Epitope haplotypes' frequency

V. DISCUSSION AND CONCLUSION

Determining genetic variation of malaria vaccine antigens can be an important element in ascertaining allele-specific efficacy and vaccine escape, particularly in endemic areas where vaccinations would be used. Thus, conducting molecular epidemiology research to determine the genetic variety of vaccine antigens and/or epitopes can aid in the development of an effective malaria vaccine.

In this study, we assessed the genetic variation in CSP sequences from Kalifabougou, Mali, by evaluating the diversity of haplotypes in the N- and C- Terminal region, the repeat regions, and the T- and B-Cell epitopes.

The central repeat and C-terminal Th2R/Th3R epitope regions were highly polymorphic in this study, whereas the N-terminal non-repeat region was less variable. There were eight common Th2R haplotypes and seven haplotypes for the Th3R region. Despite the small size of our sample, the distribution of haplotypes was diverse for both allelic variants, which is consistent with previous findings conducted in Africa (35,37). The diversity observed has been attributed to the immune pressure exerted on these two immunodominant T-cell epitopes (6, 36, 37).

In this dataset, the 3D7 *Pf*CSP-type haplotype comprised 5.0% and 13.3% of the Th2R and Th3R epitopes, respectively. These findings are consistent with results from studies that observed 5.0% of Th2R and Th3R allelic variants 3D7 type in Africa and Asia (38).

An analysis of the C terminal regions indicated that it had 43 haplotypes, with 10 variants with high frequencies. A comparison of the N terminal region haplotypes in our dataset with the 3D7 *Pf*CSP reference sequence indicated that the amino acid composition was identical, with the exception of a point mutation at position 98 in 14 of the samples. However, haplotypes in the C-terminal region were more diverse indicating that the polymorphisms in this region should be taken into account when developing vaccines based on CSP to avoid vaccine escape.

B-Cell epitope were very polymorphic, with 12 haplotypes identified. Five haplotypes were the most common, although the 3D7 *Pf*CSP-type haplotype was the second most common, with a frequency of 13.3%, while the most common haplotype had a frequency of 53.3%. These findings suggest that as the most frequent strain has a different B-cell epitope sequence, a vaccine based on the 3D7 *Pf*CSP may not be very effective in the Kalifabougou region.

The repeat region was comprised of 6 haplotypes, with the most common CSP sequences having 40 and 39 NANP/NVDP repeats. Moreover, none of the CSP sequences in the dataset had the same repeat length as that of 3D7 *Pf*CSP, which had 43 repeats. This is significant because the length of the tetrameric repeat is thought to influence the protein's stability, which in turn influences how B-cell epitopes are presented to the immune system (39). Thus, it could affect the effectiveness of a 3D7 *Pf*CSP-based vaccine in the Kalifabougou region.

We decided to compare our dataset to the 7G8 *Pf*CSP sequence, a strain from the Americas. 7G8 *Pf*CSP is now being employed in a whole sporozoite malaria vaccine (19). These results indicated that the dataset CSP sequence differed significantly from the 7G8 *Pf*CSP sequence. Only the repeat region, the B-cell epitope, and the N-terminal regions from our similar were similar to the 7G8 haplotypes. Even though the 7G8 *Pf*CSP-type T Cell epitope and C terminal region haplotypes did not match those of our samples, this findings suggests that the 7G8 *Pf*CSP B-cell epitope was the most abundant in our samples with a frequency of 55.0%. When compared to our dataset, the 7G8 *Pf*CSP-tetrameric repeat lengths was the third most common, with a frequency of 15.0% which is greater than the frequency (1.7%) of 3D7 *Pf*CSP-repeat region. These findings highlight CSP polymorphism and suggest that the repeat and B-cell epitope of our samples is similar to those of 7G8 *Pf*CSP, whereas the C-terminal is similar to that of the 3D7 *Pf*CSP strain. The N-terminal, on the other hand, is consistent across all strains. This should be taken in consideration when developing a vaccine that will be effective in our study region.

A study conducted in 2020 in two distinct ecological zones of Ghana (Navrongo, Cape Coast)(6) indicated that the frequency of the C-terminal epitope of the *PfCSP* was 5.9% in Navrongo compared to 1.7% in Kalifabougou . In Cape Cost, however, the frequency was 45.7%. Furthermore, 35 haplotypes were identified in the Th2R sites of Navrongo and Cape Coast, which is the same number of haplotypes discovered in the Kalifabougou site. Only 12 haplotypes were identified in the Th3R region at the Navrongo and Cape Coast sites, whereas 22 haplotypes were discovered at Kalifabougou. In addition, 1 to 6 amino

acid variations were observed in the CSP's Th2R and Th3R epitope regions, whereas 5 to 52 amino acid variations were observed in the Kalifabougou sites.

Finally, 3D7 *Pf*CSP-type Th2R epitope haplotypes were found in 7.4% of Navrongo samples and 56.5% of Cape coast samples, while 3D7 *Pf*CSP -type Th2R haplotypes were found in 5.0% of Kalifabougou samples. The Th3R epitope haplotype was found in 18.4% and 79.9% of the Navrongo and Cape Coast sequences, respectively, while 3D7 *Pf*CSP - Type Th3R epitope was found in 13.3% of the Kalifabougou sequences.

These findings indicates that the CSP sequences from the Kalifabougou site are more similar to those from Navrongo than to those from Cape Coast, which rejects our hypothesis.

Navrongo is located in northern Ghana, bordering Burkina Faso, which then borders Mali in the North. Malaria is a perennial threat in Navrongo, with peak transmission from June to October. Kalifabougou, on the other hand, is located in Mali's southwest. Malaria transmission is seasonal, peaking during the rainy season of July to October.
The similarity of transmission seasons and climate could explain the parallel in diversity observed in CSP sequences from the two regions.

Our study has limitations in terms of sample size, making it difficult to determine the effect of environment on CSP diversity. In fact, researchers from Ghana were able to discriminate between CSPs of Navrongo and Cape Coast, two regions with different transmission

patterns. As a result, future research should be conducted to determine whether CSP diversity is affected by environmental factors including climate (such as temperature and humidity) and regional vegetation. The study may comprise of CSP sequences form different regions of Africa to compare the genetic diversity frequencies.

These findings suggest that CSP sequences from Kalifabougou varied primarily in the C terminal region, which is what we hypothesize and is consistent with previous findings. Most of the sequences differed from the 3D7 *Pf*CSP strain, which is different from our hypothesis. Though, the B cell epitope, and repeat region were similar to the 7G8 *Pf*CSP strain.

When developing a CSP-based vaccine, multiple epitopes of CSP should be used to increase vaccine efficacy. Finally, environmental factors may play a role in CSP diversity and strain distribution in endemic areas.

VI. REFERENCES

1. Aldrich, C., Magini, A., Emiliani, C., Dottorini, T., Bistoni, F., Crisanti, A., & Spaccapelo, R. (2012). Roles of the amino terminal region and repeat region of the Plasmodium berghei circumsporozoite protein in parasite infectivity. PloS one, 7(2), e32524. https://doi.org/10.1371/journal.pone.0032524

2. Coppi, A., Natarajan, R., Pradel, G., Bennett, B. L., James, E. R., Roggero, M. A., Corradin, G., Persson, C., Tewari, R., & Sinnis, P. (2011). The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. The Journal of experimental medicine, 208(2), 341–356. https://doi.org/10.1084/jem.20101488

3. Dame, J. B., Williams, J. L., McCutchan, T. F., Weber, J. L., Wirtz, R. A., Hockmeyer, W. T., Maloy, W. L., Haynes, J. D., Schneider, I., & Roberts, D. (1984). Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite Plasmodium falciparum. Science (New York, N.Y.), 225(4662), 593–599. https://doi.org/10.1126/science.6204383

4. Zhao, J., Bhanot, P., Hu, J., & Wang, Q. (2016). A Comprehensive Analysis of Plasmodium Circumsporozoite Protein Binding to Hepatocytes. PloS one, 11(8), e0161607. https://doi.org/10.1371/journal.pone.0161607

5. Ouattara, A., Barry, A. E., Dutta, S., Remarque, E. J., Beeson, J. G., & Plowe, C. V. (2015). Designing malaria vaccines to circumvent antigen variability. Vaccine.

6. Amegashie, E.A., Amenga-Etego, L., Adobor, C. et al. Population genetic analysis of the Plasmodium falciparum circumsporozoite protein in two distinct ecological regions in Ghana. Malar J 19, 437 (2020). https://doi.org/10.1186/s12936-020-03510-3

7. Datoo, M. S., Natama, M. H., Somé, A., Traoré, O., Rouamba, T., Bellamy, D., Yameogo, P., Valia, D., Tegneri, M., Ouedraogo, F., Soma, R., Sawadogo, S., Sorgho, F., Derra, K., Rouamba, E., Orindi, B., Ramos Lopez, F., Flaxman, A., Cappuccini, F., Kailath, R., … Tinto, H. (2021). Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. Lancet (London, England), 397(10287), 1809–1818. https://doi.org/10.1016/S0140-6736(21)00943-0

8. 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

9. Cruz-Oliveira, C., Freire, J. M., Conceição, T. M., Higa, L. M., Castanho, M. A., & Da Poian, A. T. (2015). Receptors and routes of dengue virus entry into the host cells. FEMS microbiology reviews, 39(2), 155–170. https://doi.org/10.1093/femsre/fuu004

10. Yao, W., Yang, Z., Lou, X., Mao, H., Yan, H., & Zhang, Y. (2021). Molecular Characterization of Dengue Virus Type 1 in Zhejiang in 2019. Frontiers in cellular and infection microbiology, 11, 673299. https://doi.org/10.3389/fcimb.2021.673299

11. World Heath Orgamizatioon, Newsoom, April 6, 2021.

12. CDC Global Healt, Division of parasitic Diseases and Malaria, July 16, 2020.

13. CDC Global Healt, Division of parasitic Diseases and Malaria, December 15, 2021.

14. Nsanzabana C. (2019). Resistance to Artemisinin Combination Therapies (ACTs): Do Not Forget the Partner Drug!. Tropical medicine and infectious disease, 4(1), 26. https://doi.org/10.3390/tropicalmed4010026

15. Tokponnon, F.T., Sissinto, Y., Ogouyémi, A.H. et al. Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: evidence from health facility data from Benin. Malar J 18, 37 (2019). https://doi.org/10.1186/s12936-019-2656-7

16. Ouattara, A., Barry, A. E., Dutta, S., Remarque, E. J., Beeson, J. G., & Plowe, C. V. (2015). Designing malaria vaccines to circumvent antigen variability. Vaccine.

17. Jingtong Zheng, He Pan, Yinuo Gu, Xu Zuo, Nan Ran, Yuze Yuan, Chao Zhang, Fang Wang, "Prospects for Malaria Vaccines: Pre-Erythrocytic Stages, Blood Stages, and Transmission-Blocking Stages", BioMed Research International, vol. 2019, Article ID 9751471, 9 pages, 2019. https://doi.org/10.1155/2019/9751471

18. Itsara, L. S., Zhou, Y., Do, J., Grieser, A. M., Vaughan, A. M., & Ghosh, A. K. (2018). The Development of Whole Sporozoite Vaccines for Plasmodium falciparum Malaria. Frontiers in immunology, 9, 2748.

19. Moser, K.A., Drábek, E.F., Dwivedi, A. et al. Strains used in whole organism Plasmodium falciparum vaccine trials differ in genome structure, sequence, and immunogenic potential. Genome Med 12, 6 (2020). https://doi.org/10.1186/s13073-019-0708-9

20. Inoue, S., Niikura, M., Takeo, S., Mineo, S., Kawakami, Y., Uchida, A., Kamiya, S., & Kobayashi, F. (2012). Enhancement of dendritic cell activation via CD40

ligand-expressing γδ T cells is responsible for protective immunity to Plasmodium parasites. Proceedings of the National Academy of Sciences of the United States of America, 109(30), 12129–12134. https://doi.org/10.1073/pnas.1204480109

21. Somé, A. F., Bazié, T., Zongo, I., Yerbanga, R. S., Nikiéma, F., Neya, C., Taho, L. K., & Ouédraogo, J. B. (2018). Plasmodium falciparum msp1 and msp2 genetic diversity and allele frequencies in parasites isolated from symptomatic malaria patients in Bobo-Dioulasso, Burkina Faso. Parasites & vectors, 11(1), 323. https://doi.org/10.1186/s13071-018-2895-4

22. Eisen, D. P., Saul, A., Fryauff, D. J., Reeder, J. C., & Coppel, R. L. (2002). Alterations in Plasmodium falciparum genotypes during sequential infections suggest the presence of strain specific immunity. The American journal of tropical medicine and hygiene.

23. Malkin, E. M., Diemert, D. J., McArthur, J. H., Perreault, J. R., Miles, A. P., Giersing, B. K., Mullen, G. E., Orcutt, A., Muratova, O., Awkal, M., Zhou, H., Wang, J., Stowers, A., Long, C. A., Mahanty, S., Miller, L. H., Saul, A., & Durbin, A. P. (2005). Phase 1 clinical trial of apical membrane antigen 1: an asexual blood-stage vaccine for Plasmodium falciparum malaria. Infection and immunity, 73(6), 3677–3685. https://doi.org/10.1128/IAI.73.6.3677-3685.2005

24. Takala, S. L., Coulibaly, D., Thera, M. A., Batchelor, A. H., Cummings, M. P., Escalante, A. A., Ouattara, A., Traoré, K., Niangaly, A., Djimdé, A. A., Doumbo, O. K., & Plowe, C. V. (2009). Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. Science translational medicine, 1(2), 2ra5. https://doi.org/10.1126/scitranslmed.3000257

25. Chaturvedi, N., Bharti, P. K., Tiwari, A., & Singh, N. (2016). Strategies & recent development of transmission-blocking vaccines against Plasmodium falciparum. The Indian journal of medical research, 143(6), 696–711. https://doi.org/10.4103/0971-5916.191927

26. Healy, S. A., Fried, M., Richie, T., Bok, K., Little, M., August, A., Riley, L., Swamy, G. K., Wylie, B. J., Menendez, C., Muehlenbachs, A., Doumbo, O., Greenwood, B., Billingsley, P. F., Hoffman, S. L., & Duffy, P. E. (2019). Malaria vaccine trials in pregnant women: An imperative without precedent. Vaccine, 37(6), 763–770. https://doi.org/10.1016/j.vaccine.2018.12.025

27. Sirima, S. B., Richert, L., Chêne, A., Konate, A. T., Campion, C., Dechavanne, S., Semblat, J. P., Benhamouda, N., Bahuaud, M., Loulergue, P., Ouédraogo, A., Nébié, I., Kabore, M., Kargougou, D., Barry, A., Ouattara, S. M., Boilet, V., Allais, F., Roguet, G., Havelange, N., … Gamain, B. (2020). PRIMVAC vaccine adjuvanted with Alhydrogel or GLA-SE to prevent placental malaria: a first-in-

human, randomised, double-blind, placebo-controlled study. The Lancet. Infectious diseases, 20(5), 585–597. https://doi.org/10.1016/S1473-3099(19)30739-X

28. 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., … Wirth, D. F. (2015). Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. The New England journal of medicine, 373(21), 2025–2037. https://doi.org/10.1056/NEJMoa1505819

29. Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. Genomics, proteomics & bioinformatics, 13(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

30. Ouattara, A., Tran, T. M., Doumbo, S., Adams, M., Agrawal, S., Niangaly, A., Nelson-Owens, S., Doumtabé, D., Tolo, Y., Ongoiba, A., Takala-Harrison, S., Traoré, B., Silva, J. C., Crompton, P. D., Doumbo, O. K., & Plowe, C. V. (2018). Extent and Dynamics of Polymorphism in the Malaria Vaccine Candidate Plasmodium falciparum Reticulocyte-Binding Protein Homologue-5 in Kalifabougou, Mali. The American journal of tropical medicine and hygiene, 99(1), 43–50. https://doi.org/10.4269/ajtmh.17-0737

31. Dara, A., Travassos, M. A., Adams, M., Schaffer DeRoo, S., Drábek, E. F., Agrawal, S., Laufer, M. K., Plowe, C. V., & Silva, J. C. (2017). A new method for sequencing the hypervariable Plasmodium falciparum gene var2csa from clinical samples. Malaria journal, 16(1), 343. https://doi.org/10.1186/s12936-017-1976-8

32. Chaudhury, S., MacGill, R. S., Early, A. M., Bolton, J. S., King, C. R., Locke, E., Pierson, T., Wirth, D. F., Neafsey, D. E., & Bergmann-Leitner, E. S. (2021). Breadth of humoral immune responses to the C-terminus of the circumsporozoite protein is associated with protective efficacy induced by the RTS,S malaria vaccine. Vaccine, 39(6), 968–975. https://doi.org/10.1016/j.vaccine.2020.12.055

33. Shi, Y. P., Udhayakumar, V., Alpers, M. P., Povoa, M. M., Oloo, A. J., Ruebush, T. K., 2nd, & Lal, A. A. (1993). Natural antibody responses against the non-repeat-sequence-based B-cell epitopes of the Plasmodium falciparum circumsporozoite protein. Infection and immunity, 61(6), 2425–2433. https://doi.org/10.1128/iai.61.6.2425-2433.1993

34. Beutler, N., Pholcharee, T., Oyen, D., Flores-Garcia, Y., MacGill, R. S., Garcia, E., Calla, J., Parren, M., Yang, L., Volkmuth, W., Locke, E., Regules, J. A., Dutta, S., Emerling, D., Early, A. M., Neafsey, D. E., Winzeler, E. A., King, C. R., Zavala, F., Burton, D. R., … Rogers, T. F. (2022). A novel CSP C-terminal

epitope targeted by an antibody with protective activity against Plasmodium falciparum. PLoS pathogens, 18(3), e1010409. https://doi.org/10.1371/journal.ppat.1010409

35. 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, 120. https://doi.org/10.1186/1475-2875-8-120

36. 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), e43430. https://doi.org/10.1371/journal.pone.0043430

37. Bailey, J. A., Mvalo, T., Aragam, N., Weiser, M., Congdon, S., Kamwendo, D., Martinson, F., Hoffman, I., Meshnick, S. R., & Juliano, J. J. (2012). Use of massively parallel pyrosequencing to evaluate the diversity of and selection on Plasmodium falciparum csp T-cell epitopes in Lilongwe, Malawi. The Journal of infectious diseases, 206(4), 580–587. https://doi.org/10.1093/infdis/jis329

38. 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), 6622. https://doi.org/10.1038/s41598-018-24585-8

39. Escalante, A. A., Grebert, H. M., Isea, R., Goldman, I. F., Basco, L., Magris, M., Biswas, S., Kariuki, S., & Lal, A. A. (2002). A study of genetic diversity in the gene encoding the circumsporozoite protein (CSP) of Plasmodium falciparum from different transmission areas--XVI. Asembo Bay Cohort Project. *Molecular and biochemical parasitology*, *125*(1-2), 83–90. https://doi.org/10.1016/s0166-6851(02)00216-5

40. Ouattara, A., Takala-Harrison, S., Thera, M. A., Coulibaly, D., Niangaly, A., Saye, R., Tolo, Y., Dutta, S., Heppner, D. G., Soisson, L., Diggs, C. L., Vekemans, J., Cohen, J., Blackwelder, W. C., Dube, T., Laurens, M. B., Doumbo, O. K., &Plowe, C. V. (2013). Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications. The Journal of infectious diseases.

41. Kaslow, D. C., & Biernaux, S. (2015). RTS,S: Toward a first landmark on the Malaria Vaccine Technology Roadmap. Vaccine, 33(52), 7425–7432. https://doi.org/10.1016/j.vaccine.2015.09.061

42. Kumkhaek, C., Phra-ek, K., Singhasivanon, P., Looareesuwan, S., Hirunpetcharat, C., Brockman, A., Grüner, A. C., Lebrun, N., Rénia, L., Nosten, F., Snounou, G., & Khusmith, S. (2004). A survey of the Th2R and Th3R allelic variants in the circumsporozoite protein gene of P. falciparum parasites from western Thailand. The Southeast Asian journal of tropical medicine and public health, 35(2), 281–287.