

THE MOLECULAR EPIDEMIOLOGY OF MALARIA IN PREGNANCY

Jaymin Chetan Patel

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements of the degree of Doctor of Philosophy in the Department of Epidemiology.

Chapel Hill
2016

Approved by:

Steven R. Meshnick

Stephanie M. Engel

Jonathan J. Juliano

Jeffrey A. Bailey

Steve M. Taylor

© 2016
Jaymin Chetan Patel
ALL RIGHTS RESERVED

ABSTRACT

Jaymin Chetan Patel: The Molecular Epidemiology of Malaria in Pregnancy
(Under the direction of Steven R. Meshnick)

Malaria in pregnancy remains a significant public health problem with an estimated 125 million pregnant women at risk for *Plasmodium falciparum* malaria globally every year. Pregnancy associated malaria (PAM) causes several adverse pregnancy and birth outcomes including maternal anemia, low birth weight (LBW), and small-for-gestational age (SGA). PAM also results in 10,000 maternal and 70,000-200,000 infant deaths annually. The key biological mechanism by which the plasmodium parasite infects pregnant women is through sequestration of infected erythrocytes in the placenta. This sequestration facilitated by a large polymorphic plasmodium surface antigen, VAR2CSA. Efforts are underway to develop the first syndrome-specific malaria vaccine against PAM targeting VAR2CSA. The ID1-DBL2x region of VAR2CSA is the minimal binding epitope and has emerged as a lead vaccine candidate. However, there is limited data on the extent of genetic diversity of ID1-DBL2x in field isolates. Also it is unknown if particular variants are more pathogenic than others.

In this dissertation, we leveraged advanced molecular methods, population genetics, and epidemiology to inform vaccine development efforts. Using samples from *P. falciparum*-infected pregnant women in Malawi and Benin, we characterized the genetic diversity of the ID1-DBL2x vaccine target and identified pathogenic clades. We demonstrated that ID1-DBL2x region is highly diverse in both countries. We found that the entire 1.6kb region is primarily under

balancing selection, confirming its role as an important epitope. Importantly, our phylogenetic analyses showed clustering of ID1-DBL2x variants in multiple distinct clades. Two clades containing the vaccine referent strains (3D7 and FCR3) were found in both countries in addition to three unique clades in Benin. Across multiple birth outcomes we consistently identified variants from 3D7-like clade as pathogenic. We did detect LBW and SGA variants in FCR3-like clade. However, compared to FCR3-like clade, 3D7-like clade was associated with LBW, SGA, and lower infant birth weight.

Overall, our results provide strong evidence for developing a polyvalent VAR2CSA-based vaccine against PAM. A vaccine that includes variants from most common and pathogenic clades will be more efficacious than current monovalent vaccines in phase I trials. The integrative approach used here can be employed to inform development of future malaria vaccine candidates targeting polymorphic antigens.

ACKNOWLEDGEMENTS

There are several people I would like to thank for all the help and support they have given me not only during this dissertation work but throughout my time in the doctoral program. The incredible support that I received made my time at UNC a tremendous learning experience.

First, and foremost, I would like to thank my advisor and dissertation chair, Steve Meshnick, for being best mentor anyone could ask for. Steve has guided my academic career and really fostered my curiosity from day one in the doctoral program. He provided me with opportunities to do exciting field work, gain valuable lab skills, and work with large datasets. Steve really helped me develop as an independent scientist; he gave me autonomy over one of his NIH grants which turned into this dissertation project, encouraged me to forge new collaborations, and challenged me to get creative and merge approaches from different fields to answer complicated questions in malaria epidemiology. I am truly thankful for all that he has done for me.

I would also like to thank the rest of my committee; Jonathan Juliano, who has been an amazing teacher, advisor, and always encouraged me to use new methods and approaches to answer important questions relating to malaria genomics and epidemiology; Jeff Bailey, who has provided great bioinformatics support for this project and all other projects using next-generation sequencing data in our lab; Stephanie Engel, who has been simply amazing, helping me think through the modeling approaches and how to best merge molecular data into epidemiologic models. She has always been positive, insightful, encouraging, and just the best mentor a

graduate student can ask for; and Steve Taylor, who has been an incredible mentor throughout my time here at UNC. I have learned tremendously from his unique perspective on scientific research, emphasis on a healthy mix of research and applied public health, and his brilliant ability to write and communicate effectively.

I would remiss if I didn't thank Dr. Udhayakumar and Kim Lindblade from CDC, both of whom have been amazing mentors from the start of my academic career. Working in Dr. Kumar's lab as a curious graduate student, I got my introduction into molecular epidemiology research. Kim has just been an incredible mentor and a role model. She exemplifies the perfect scientist who merges the rigors academia and epi methods with public health practice. Working with her on the WASH study in Guatemala was my first exposure to field work and I have learned a great deal from her over the years.

The aid and support I received from Nancy Colvin, Carmen Woody, Jennifer Moore, and Valerie Hudock really made this process go a lot smoother. Their dedication to the students in our department is outstanding.

To the Meshnick lab members, you have been like family these past five years and I have learned so much from each one of you. I cannot thank Jolly enough for all his help in the lab over the last three years. He has been on this project since the first day when we started to design the PCRs and has been through the ups and downs with me. He is a master with lab work and I am glad to have had an opportunity to work with him. Thanks to Sujata and Stephanie, who always helped me with lab work at moment's notice; perhaps out of the concern that I might accidentally set fire in the lab! I would also like to thank the senior members of the lab, Carla Cerami, Natalie Bowman, and Jessica Lin; throughout my time at UNC, they provided me with

opportunities to gain valuable experience working on a variety of projects and apply what I learned in class to actual data. My time in the Meshnick lab has been incredibly fruitful.

I have been fortunate to receive financial support during my tenure at UNC, including NIH R56 AI106129 01A1, the National Evolutionary Synthesis Center (NESCent) graduate fellowship, Carolina Asia Center pre-dissertation award, GPSF travel award, David J. Sencer Scholarship and the Koch Travel Award. I am also grateful for our collaborators who agreed to share data and samples with us. Feiko O ter Kuile, who sent us samples and data from Malawi. Nicaise Ndam, who has been an enthusiast collaborator and generously hosted me in Paris to process the necessary samples from their Benin cohort study, which have really strengthen the results of this research. I hope to continue working with them in the future. I also want to thank the field staff and investigators at both field sites in Malawi and Benin. It is because of their hard work and dedication that we were able to complete this project. I am indebted to the women and infants who participated in the study, besides participating in the study; they have provided me a source of motivation to continue working in public health with the goal to improve health and lives of people. Without them, this project would not have been possible.

Having a group of peers and friends with whom you can collaborate with, use as a sounding board, and learn from is vital to a graduate student, and I have been extremely fortunate in that category. I cannot thank Nicholas Hathaway enough for all his support with the bioinformatics, his help with clustering next-generation sequencing data was absolutely critical to this project. He is a wizard with coding and coming up with new ways to visualize and process large amounts of data. Christian Parobek has been an absolute rock-star, he has helped me immensely with the genetic approaches on this project and participated in numerous exercises where we came up with ways to merge and analyze molecular and epi data together. Catherine

Lesko and Elizabeth Cromwell have been a great sounding board for figuring out the best epidemiologic approaches for this project and beyond. I have learned so much from my peers and colleagues, and I look forward to collaborating with them in the future.

Last but most importantly, I want thank my family for their unconditional love and support - my brother, Hardik, who has been my role model since I was ten and someone I try to emulate every day; my sister-in-law, Jigna, who has been constantly loving and supportive; my nephew, Akhil, who is the best remedy for stress and a constant reminder of the bigger picture. I really want to thank and acknowledge the role of my parents in this; they dropped everything and moved to the United States just so my brother and I could have the best educational opportunities. I learned how to work hard, challenge myself, and push myself to be better all the while staying humble from them. They have always put their family before themselves and the sacrifices they made for us have been immense. Without them, I wouldn't be who I am today or be where I am today.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS.....	xv
CHAPTER ONE: SPECIFIC AIMS	1
Aim 1:	2
Aim 2:	2
CHAPTER TWO: BACKGROUND AND SIGNIFICANCE	4
Malaria disease burden	4
Malaria in Pregnancy.....	6
Pathogenesis of malaria in pregnancy: role of var2csa	8
Key gaps in our understanding of var2csa.....	12
Rationale	13
Innovation	15
CHAPTER THREE: DESCRIPTION OF DATA SOURCES	19
Study design.....	19
Source population and Study area	21
Study population.....	22
Demographic and clinical data	23
Outcome assessment.....	25
Exposure ASSESSMENT.....	26
CHAPTER FOUR: METHODS	28
DNA extraction.....	28
ID1-DBL2x PCR assay.....	28
PacBio Circular Consensus Sequencing (CCS).....	29
Clustering of next-generation sequencing data.....	30

Validation study.....	31
Specific Aim 1	31
Specific Aim 2	37
CHAPTER FIVE: RESULTS. GENETIC DIVERSITY AND	
POPULATION STRUCTURE OF VAR2CSA ID1-DBL2X	
IN MALAWI AND BENIN: IMPLICATIONS FOR VACCINE	
DEVELOPMENT.....	
	41
Study population.....	41
Validation of genotyping approach	41
ID1-DBL2x deep-sequencing among clinical isolates	42
Within- group (alpha) diversity of ID1-DBL2x	42
Phylogenetic clustering.....	43
Between- group (beta) diversity of ID1-DBL2x	44
Detecting signatures of selection within ID1-DBL2x	44
Discussion.....	45
CHAPTER SIX: RESULTS. IDENTIFICATION OF PATHOGENIC	
VAR2CSA CLADES: ASSOCIATIONS BETWEEN	
ID1-DBL2X CLADES AND BIRTH OUTCOMES	
	63
Study population.....	63
ID1-DBL2x distribution – phylogenetic analyses	64
Association between ID1-DBL2x clades and birth outcomes	65
Discussion.....	67
CHAPTER SEVEN: DISCUSSION.....	
	77
Summary of findings	78
Conclusions.....	80
Future Directions	82
APPENDIX A: POWER CALCULATION	84
APPENDIX B: LIST OF BARCODED PRIMERS USED	86
APPENDIX C: DEVELOPMENT AND OPTIMIZATION OF ID1-DBL2X	
GENOTYPING APPROACH	87
REFERENCES	89

LIST OF TABLES

Table 4.1: ID1-DBL2x hemi-nested PCR primer sequences	39
Table 5.1: Description of the study population.....	49
Table 5.2: Characteristics of PCR amplified and unamplified samples	50
Table 5.3: Nucleotide diversity (π) and Tajima's D	51
Table 5.4: Within-group (alpha) diversity of ID1-DBL2x populations.....	52
Table 6.1: Study population characteristics	71
Table 6.2: Distribution of infant birth weight and adverse birth outcomes by ID1-DBL2x clades in Malawi and Benin	72
Table 6.3: Associations between ID1-DBL2x clades and infant birthweight (grams)	73
Table 6.4: Associations between ID1-DBL2x vaccine clades (3D7 & FCR3) and small-for-gestational age (SGA)	74

LIST OF FIGURES

Figure 2.1: Global distribution of malaria deaths – 2010. Size of each country is proportional to the malaria deaths in the respective country.	17
Figure 2.2: Schematic representation of the var2csa gene. VAR2CSA consists of an N-terminal segment (NTS), six Duffy Binding Like Domains (DBL), four cysteine-rich interdomain regions (ID), a transmembrane domain (TM), and an intra-cellular acidic terminal segment (ATS). Current vaccine development efforts (PlacMalVac and PriMalVac) are targeting the NTS-DBL2x region. The ID1-DBL2x region has recently been identified as the minimal CSA-binding region that contains major protective epitopes and elicits a strong host immune response	18
Figure 3.1: Malawi Study Site	27
Figure 3.2: Benin Study Site	27
Figure 4.1: Schematic representation of the bioinformatics clustering pipeline employed in this study. Pools of amplified ID1-DBL2x from individual women in technical replicates using barcoded primers were sequenced on PacBio CCS platform. Reads were clustering using a k-mer algorithm to identify unique ID1-DBL2x variants in the study population.	40
Figure 5.1: Results from the validation study - expected and observed frequencies of the seven genomic parasite lines in five pools using referent-based and k-mer clustering.	53
Figure 5.2: Scatterplot of expected and observed frequencies of ID1-ID2x haplotypes in mixtures of reference parasite lines.	54
Figure 5.3: Flow chart of the processing and filtering raw reads by barcodes and quality.	55
Figure 5.4: Rarefaction curves of ID1-DBL2x variants in Malawi (blue) and Benin (red). Calculated rarefaction curves are represented by solid lines. Extrapolated curves are depicted by dotted lines. The rarefaction curves were bootstrapped 1000 times to generate the 95% confidence intervals (CIs), which are represented by lightly shaded red and blue.....	56
Figure 5.5: Expected heterozygosity (H_e) at each amino acid position along the ID1-DBL2x region among variants populations in Malawi (blue) and Benin (red).	57

- Figure 5.6: Maximum-likelihood phylogenetic trees of ID1-DBL2x variants in (A) Malawi and (B) Benin. All trees were bootstrapped 1000x and bootstrap values >80 were used as cutoff points for significant branch differentiation. In Malawi, variants clustered in two distinct clades with the two referent strains (3D7 and FCR3). In Benin, in addition to the 3D7 and FCR3 clades, three other clades were detected. 58
- Figure 5.7: Principal Coordinate Analysis (PCoA) of ID1-DBL2x variants in the study population. (A) PCoA plot using principal coordinates 1 and 2 (PC1 & PC2) (B) PCoA plot using principal coordinates 1 and 3 (PC1 & PC3). Each dot represents the unique ID1-DBL2x variants colored according to the clade. The axes indicate which coordinates are being plotted and the percentage of variation explained by that particular axis. 59
- Figure 5.8: F_{ST} values using a sliding window approach (window size = 10bp, step size = 10bp) across the 1.6kb ID1-DBL2x region of *var2csa*. F_{ST} values were calculated between the two major vaccine clades (3D7 and FCR3). F_{ST} values range from 0 to 1, where 0 signifies a genetically identical population and 1 signifies completely differentiated populations. Sliding window F_{ST} analyses show that scanning across the region, the parasite populations in the two clades are very similar except for a ~100 bp region (highlighted in gray) where the two populations differ substantially. 60
- Figure 5.9: Tajima's D using a sliding window approach (window size = 100bp, step size = 25bp) across the 1.6kb ID1-DBL2x region of *var2csa*. (A) Tajima's D calculated for ID1-DBL2x populations in Malawi and Benin and (B) 3D7 and FCR3 clades (dimorphic region highlighted in gray). Tajima D values less than 0 indicate directional selection, equal to 0 indicate genetic drift, and greater than 0 indicate balancing selection. 61
- Figure 6.1: Maximum-likelihood phylogenetic trees of ID1-DBL2x variants in (A) Malawi and (B) Benin by low birth weight (LBW). All trees were bootstrapped 1000x and bootstrap values >80 were used as cutoff points for significant branch differentiation. LBW was defined according to World Health Organization's criteria as infant with a birth weight of less than 2500 grams. 3D7 (green) and FCR3 (purple) were detected in both countries. Majority of the LBW variants (red) in Malawi were present in the 3D7-like clade whereas in Benin, the LBW variants were more evenly distributed between 3D7-like and FCR3-like clades. 75

Figure 6.2: Maximum-likelihood phylogenetic trees of ID1-DBL2x variants in (A) Malawi and (B) Benin by small-for-gestational age (SGA). All trees were bootstrapped 1000x and bootstrap values >80 were used as cutoff points for significant branch differentiation. SGA was defined as infant with a birth weight below the 10th percentile for babies of the same gestational age at delivery. 3D7 (green) and FCR3 (purple) were detected in both countries. SGA variants (red) clustered more readily in 3D7-like clade in Malawi whereas in Benin, the distribution was more even between 3D7-like and FCR3-like clades. We also identified four SGA variants clustered in clade 3. 76

LIST OF ABBREVIATIONS

ACE	abundance coverage estimator
ACT	artemisinin-based combination therapy
AIDS	Acquired Immune Deficiency Syndrome
ATS	intracellular acidic terminal segment
CCS	circular consensus sequencing
CI	confidence interval
ID	cysteine-rich interdomain regions
CSA	chondroitin sulfate A
DBLs	Duffy Binding Like Domains
DBS	dried blood spots
dhfr	dihydrofolate reductase
dhps	dihydropteroate synthetase
DP	Dihydroartemisinin-Piperaquine
EMA	European Medicines Agency
F_{ST}	Wright's Fixation index
H'	Shannon index
H_e	expected heterozygosity
HIV	Human Immunodeficiency Virus
ICE	incidence coverage estimator
IE	infected erythrocytes
IgG	Immunoglobulin G
IPTp	intermittent preventive treatment in pregnancy

IPW	inverse probability weights
IQR	interquartile range
IRS	indoor residual spraying
ISTp	intermittent screening and treatment in pregnancy
ITN	insecticide treated bed nets
kdr	Knockdown resistance
LBW	low birth weight
MiP	Malaria in pregnancy
MLE	maximum composite likelihood
MOI	multiplicity of infection
MUAC	mid-upper arm circumference
NTS	N terminal segment
OR	odds ratio
<i>P.</i>	<i>Plasmodium</i>
PAM	pregnancy associated malaria
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
PfEMP1	<i>Plasmodium falciparum</i> erythrocyte membrane protein 1
RDT	rapid diagnostic test
SD	standard deviation
SGA	small-for-gestational age
SMRT	Single Molecule, Real-Time
SP	sulphadoxine–pyrimethamine
TM	transmembrane domain

WHO World Health Organization

π nucleotide diversity

CHAPTER ONE: SPECIFIC AIMS

Malaria in pregnancy is a significant public health problem with high disease morbidity and mortality. Each year an estimated 125 million pregnant women are at risk of *Plasmodium falciparum* malaria infection [1]. Malaria in pregnancy causes several adverse pregnancy and birth outcomes including maternal anemia, severe malaria, spontaneous abortions, stillbirth, infant deaths, preterm birth and low birth weight (LBW) and small-for-gestational age (SGA) [2-12]. In Africa, pregnancy-associated malaria (PAM) is responsible for approximately one million cases of infant LBW and of maternal anemia, plus 10,000 maternal and 70,000 to 200,000 infant deaths annually [13, 14].

The adverse consequences of *P. falciparum* infections on pregnancy are due to the ability of infected erythrocytes (IE) to sequester in the placenta. This sequestration of IE in the placenta is facilitated by VAR2CSA, a specific variant of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1). VAR2CSA, a large polymorphic antigen, is expressed on the surface of the IE and binds to the placental chondroitin sulfate A (CSA) on the syncytiotrophoblast [15-19]. Antibodies against VAR2CSA prevent the cytoadhesion and the adverse effects of PAM [20-23]. The N-terminus region of VAR2CSA up to the DBL2x domain is critical to the binding process and able to induce an antibody response with similar inhibitory capacity as that elicited against full-length VAR2CSA [24-26].

Recently, the ID1-DBL2x region of VAR2CSA was identified as the minimal binding epitope and has emerged as one of the lead vaccine candidates [27-31]. The overall goal of this

dissertation is to characterize the genetic diversity of ID1-DBL2x vaccine target region and determine its association with adverse birth outcomes in order to identify pathogenic clades and inform vaccine development efforts.

Specifically, we will:

AIM 1: Characterize the genetic diversity of the 1.6 kb ID1-DBL2x region of *var2csa* in Beninese and Malawian pregnant women at delivery according to gravidity and country.

Rationale: As efforts are underway to develop an effective vaccine against malaria in pregnancy, a recombinant ID1-DBL2x of VAR2CSA is one of the lead vaccine candidates and is currently in phase I testing [32, 33]. However, we don't completely know yet how much diversity is there in ID1-DBL2x. Cloning and sequencing of short regions of ID1-DBL2x as well as other regions of *var2csa* have shown significant diversity [28, 29, 34, 35]. The gravidity-dependent nature of acquired immunity in pregnant women has shown that multigravid women elicit a stronger humoral response to *var2csa* than primi- and secundigravid women. It is this acquired immunity that is hypothesized to protect them and the fetus from the harmful effects of placental malaria [2, 3, 10-12, 20, 23, 36, 37]. A more comprehensive understanding on the extent of *var2csa* diversity will aid in vaccine development.

AIM 2: Identify pathogenic clades of ID1-DBL2x by determining the association between specific clades and adverse birth outcomes.

Rationale: We do not know whether specific variants of ID1-DBL2x are more virulent than others. Deep sequencing the ID1-DBL2x region of *var2csa* in a clinically well-defined cohort of malaria-infected pregnant women will provide an opportunity to look for pathogenic clades that are responsible for the adverse birth outcomes. Due to acquired immunity, different ID1-DBL2x

variants will adhere to the CSA with different avidities and parasite will sequester differentially in the placenta and this will ultimately affect fetal development [21, 36-38].

CHAPTER TWO: BACKGROUND AND SIGNIFICANCE

MALARIA DISEASE BURDEN

Malaria is a disease caused by infection from the *Plasmodium* parasite transmitted by the female Anopheles mosquito. There are five different species of the *Plasmodium* parasite that infect humans, namely, *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Of these, *P. falciparum* and *P. vivax* are the most prevalent and responsible for majority of the global malaria disease burden. While *P. vivax* has a wider geographic distribution because of its ability to develop in the Anopheles mosquito vector at lower temperatures and survive at higher altitudes, *P. falciparum* is the deadliest of the two *Plasmodium* species [39-42]. Following an infective bite from the Anopheles mosquito, the incubation period varies from 7 to 30 days before development of clinical symptoms. Incubation period also varies by type of infecting *Plasmodium* species with longer periods observed with *P. malariae* and shorter periods observed with *P. falciparum* infections. Typically, clinical manifestation of malaria infection is characterized by paroxysms, which include cyclical fevers, chills, and flu-like illness. The spectrum of illness due to malaria infection ranges from asymptomatic infections with no clinical symptoms to severe malaria and death. In general, malaria is a completely curable disease. However, if misdiagnosed or left untreated, it can lead to severe and complicated manifestations of the disease. Severe and complicated malaria typical result in vital organ dysfunction (i.e. anemia, jaundice, kidney failure, and coma) and can lead to death [39-41].

Over the last two decades, the world has seen a significant decline in malaria related morbidity and mortality. Globally, since 2000, malaria incidence and mortality has declined by 18% and 48% respectively[42]. Much of the progress made in this time period can be attributed to a concerted effort to scale up effective interventions such as indoor residual spraying (IRS), insecticide treated bed nets (ITNs), prompt treatment with artemisinin-based combination therapy (ACTs), and improvements in diagnosis and case management. Despite the encouraging gains, malaria remains a significant public health problem. According the World Health Organization (WHO), malaria is endemic in 94 countries with approximately 40% of the world's population (~3.2 billion people) currently at risk. In 2015, the WHO estimated 214 million cases (uncertainty range: 149 – 303 million) and 438,000 deaths (uncertainty range: 236,000 – 635,000) were due to malaria worldwide [39, 42].

As is the case with many tropical diseases, malaria disproportionately affects the poor and under-developed regions of the world [43]. The heaviest burden of disease is experienced by individuals in Africa (Figure 2.1). In 2015, 88% of all malaria cases and 90% of all deaths due to malaria occurred in Africa alone [39, 42]. Malaria in Africa has not only impacted health of its people but also affected economic and social development in that part of the world. The link between malaria and poverty is a cyclical one. Poverty promotes conditions where malaria transmission is sustained, and in turn malaria hinders economic growth and development, propagating the cyclical relationship between the two [44]. Studies have shown that after controlling for geographic factors, malaria has a strong negative association with country income levels and development indices [45-47]. In 2009, among three countries in sub-Saharan Africa (Kenya, Tanzania, and Ghana), total economic costs for treatment and prevention of malaria in children under the age of 5 were between \$37.8-131.98 million. Additionally, malaria was

responsible for \$67-290 million in productivity loss due to premature death in these countries [48, 49].

MALARIA IN PREGNANCY

While malaria infects all people where the parasite is in transmission, specific populations are at a higher risk of infection, clinical manifestation of the disease, and death due to malaria than others. Vulnerable populations include infants, children under five years of age, individuals with HIV/AIDS, migrants, mobile populations, and pregnant women [39, 40]. Pregnant women are at a higher risk for malaria infections and its adverse effects than non-pregnant adults. This increased susceptibility for malaria infection is likely due to their suppressed immune system among other factors [2, 3, 10, 11, 39].

Each year, an estimated 125 million pregnant women are at risk of malaria with 56 million of those living in high transmission areas [1, 13, 14]. In areas of stable endemic malaria transmission in Africa, the prevalence of maternal malaria infection (peripheral or placental infection) at delivery is estimated to be around 25% (range: 5-52%) [3]. These burden estimates are likely underestimated as they are point-prevalence numbers collected at delivery. Prevalence estimates at delivery do not account for malaria infections before or after the time of ascertaining point-prevalence and would miss infections during pregnancy. Further, these estimates are based on microscopy, which fail to diagnose submicroscopic infections that can be detected readily by more sensitive methods such as polymerase chain reaction (PCR) and placental histology [3, 8, 11, 14].

Malaria in pregnancy (MiP) or pregnancy associated malaria (PAM) is associated with several adverse outcomes including maternal anemia, severe malaria, spontaneous abortions, stillbirth, infant deaths, preterm birth, low birth weight (LBW), and small-for-gestational age

(SGA) [2-14]. Of these, maternal anemia and LBW are the most frequently observed ill effects of PAM. In sub-Saharan Africa, malaria is estimated to be responsible for approximately one million cases of infant LBW and of maternal anemia, plus 10,000 maternal and 70,000 to 200,000 infant deaths annually [13, 14].

The increased pregnancy-related risk to malaria infection is apparent and higher among pregnant women compared to non-pregnant adults despite acquired immunity due to past exposure to malaria. The risk of malaria infection during pregnancy is highest during second trimester. Little is known about the risk during first trimester; however, susceptibility in first trimester must also increase in order to explain the peak prevalence of malaria observed during the second trimester [2, 3, 9, 50]. There are several factors that affect the risk of PAM. These include age, co-infection with HIV, transmission intensity, use of intermittent preventive treatment in pregnancy (IPTp) with sulphadoxine–pyrimethamine (SP), ITNs, and gravidity. Age is an important risk factor for malaria in pregnancy. Young pregnant women (especially adolescent women) have a higher risk of malaria and its adverse effects than older pregnant women. This relationship is observed irrespective of gravidity, implying that age-associated immunity influences a woman's ability to control an infection during pregnancy [3]. Another factor that modifies the effect of malaria during pregnancy is co-infection with HIV/AIDS. HIV-infected pregnant women have shown to have higher prevalence of parasites than HIV-uninfected pregnant women. Transmission intensity also appears to modify susceptibility to malaria infection during pregnancy. In low-transmission areas, women of all gravidities are at risk for severe disease and adverse pregnancy outcome. However, that is not the case with women in high-transmission areas where parity dependent immunity to ill effects of malaria is observed [3, 10, 11, 14].

The adverse effects of malaria in pregnancy can be substantially reduced or avoided all together using proven interventions such as IPTp- SP and ITNs along with effective case management that involves prompt and accurate diagnosis of malaria and anemia [5, 51-53]. Currently, the world health organization (WHO) recommends that all women living in stable transmission areas such as sub-Saharan Africa receive one dose of IPTp with SP every month starting in their second trimester [54]. IPTp and ITNs have shown to be significantly associated with reduced odds of delivering a LBW infant (protective efficacy: 21-23%) [5, 55]. One of the most important risk factors for PAM is gravidity. The risk of malaria infection and severity of disease during pregnancy is modified by gravidity. Particularly, primigravid and secundgravid women are at higher risk for adverse effect from malaria infections than multigravid women [2, 3, 10-12, 20, 23, 36, 37]. This indicates that parity-dependent acquired immunity due to past exposures and co-infections with diseases that affect the immune system plays an important role in protection against malaria.

PATHOGENESIS OF MALARIA IN PREGNANCY: ROLE OF VAR2CSA

The key mechanism by which malaria infects pregnant women is through the accumulation of parasite infected erythrocytes (IE) in the intervillous (vascular) space of the placenta [2]. Studies have shown that malaria infection during pregnancy also leads to increased frequency of maternal phagocytic cells, especially monocytes in the intervillous space, and deposition of hemozoin in phagocytic leukocytes and within fibrin deposits in the intervillous space [2, 10, 23, 56, 57]. These changes in the placenta are associated with adverse outcomes such as preterm birth, maternal anemia, and decreased birth weight due to fetal growth restriction.

The sequestration of malaria parasites in the placenta is facilitated by the parasite antigens expressed on the surface of the infected erythrocytes that act as ligands to bind to the placental chondroitin sulfate A (CSA). The CSA is thought to act as a reversible immobilizer of several molecules including hormones and cytokines [10, 58, 59]. The surface antigens involved in placental infection are distinct from other surface antigens in their form and function. The expression of the surface antigens present in the IE that bind to CSA is responsible for eliciting a specific IgG response from the host. The IgG response developed after repeated infections or exposures to the malaria parasites mediates the protective immunity that is seen in multigravid women. The distinction in form and function of antigens expressed during placental infection and the immune response elicited provides an explanation as to why previously clinically immune women become susceptible to malaria infection during pregnancy, especially women of low gravidities [10, 22, 23, 36].

The major surface antigenic protein in the parasite that acts as a ligand responsible for the cytoadherence of the infected erythrocytes to the CSA belong to a group of proteins called *P. falciparum* Erythrocyte Membrane Protein-1 (PfEMP1). The PfEMP1s are large proteins, ranging from 200-350 kDa in size. All PfEMP1s share structural similarities: an N terminal segment (NTS); varying numbers of Duffy Binding Like Domains (DBLs); cysteine-rich interdomain regions (CIDR or ID); transmembrane domain (TM); C2 domain; and an intra-cellular acidic terminal segment (ATS). PfEMP1s are expressed during the late erythrocytic stage, form a knob like structure on the infected erythrocytes and bind to different adhesins. In the case of malaria in pregnancy, the PfEMP1s bind to the CSA [59-61].

PfEMP1s are coded by a family of a multi-copy gene family named *var*. There are a total of about 60 *var* genes in the *P. falciparum* genome [62]. Studies have shown that the expression

of the *var* gene involves a set of regulation mechanisms including activation, switching, and silencing of localization sites. Only a single variant of the *var* gene family is expressed on the surface of the infected erythrocyte at any given time. This expression mechanism is mutually exclusive meaning that even though there are several *var* genes, during the course of an infection, *P. falciparum* massively expresses only one *var* gene at a time and then switches expression to another *var* gene, successively. This switch lets the parasite to undergo clonal antigenic variation, allowing the parasite to evade the host immune response and maintain a persistent infection. Genetic recombination between *var* paralogs is thought to be an important mechanism in the generation and maintenance of the extreme diversity in the antigen repertoire [10, 56, 60, 63-65].

The *var* genes are located mostly in the subtelomeric regions and share a similar structure [66]. PfEMP1 has a long exon encoding the variable extracellular domain, a conserved intron and a short second exon coding for the acid terminal segment [61]. Initially, transcription of two variants of the *var* gene was thought to be critical in the pathogenesis of malaria in pregnancy, namely, *var1csa* and *var2csa*. Both *var1csa* and *var2csa* showed affinity for CSA and interclonal conservation. Consequently, a number of studies showed that *var1csa* is not associated with malaria in pregnancy as it is not a target of protective immune response by the host (pregnant women) [21]. Further investigation revealed that the other variant, *var2csa*, was indeed present in all *P. falciparum* clones and was transcribed by CSA-selected and placental parasites [15-19]. Several domains of the *var2csa* gene have shown to have affinity for CSA and disruption of *var2csa* is correlated with inability of infected erythrocytes to bind to the CSA. Studies analyzing immune responses to PAM showed that *var2csa* specific IgG levels in *P. falciparum* exposed adults confirmed the female-restricted and parity dependent nature of

var2csa. Importantly, VAR2CSA was present on the surface of intact PAM infected erythrocytes but absent from the surface of other non-pregnant adults and children [7, 17, 18, 21, 56, 63].

High levels of VAR2CSA-specific IgG plasma levels have also been shown to be protective against adverse consequences of PAM. *In vitro* studies using sera from *P. falciparum*-exposed multigravid pregnant elicited specific immune responses to two or more domains of VAR2CSA. The IgG antibodies targeted the polymorphic rather than the conserved epitopes of VAR2CSA indicating that the domains that are accessible to protective antibodies are under selection pressure that favor polymorphism. This selection pressure is likely due to the host immune response and highlights the significance of VAR2CSA-specific IgG response on PAM outcomes [4, 20-22, 37, 38, 67, 68].

Similar to the other *var* genes, *var2csa* comprises of DBL domains and cysteine-rich interdomain regions. Specifically, *var2csa* has the NTS region along with ATS and TM regions flanking on the two ends of the gene, and in between it consists of six DBL domains and four interdomain regions (Figure 2.2) [16, 19, 21, 24, 63]. Significant effort has been invested to identify the regions of *var2csa* that play a role in the pathogenesis of malaria in pregnancy and potential targets for vaccine candidates. Several of the individual domains and interdomain regions have shown to bind to the CSA *in vitro*. DBL2x, DBL3, DBL4, DBL5, and DBL6 along with the NTS and interdomain regions 1 (ID1) and 2 (ID2) have shown to elicit immune response or be critical in the binding of the infected erythrocytes to the CSA [20, 22, 24, 31, 67-78].

However, the N-terminus region of VAR2CSA up to the DBL2x domain has been identified as critical to the binding process and able to induce antibody response with similar inhibitory capacity as that elicited against full-length VAR2CSA [26, 33]. Recently, studies have

shown that a shorter fragment in the N-terminus region, the ID1-DBL2x region plus 93 amino acids of the ID2 interdomain was critical to pathogenesis of malaria in pregnancy as it binds to the CSA with same avidity as the whole protein. The ID1-DBL2x region is the shortest VAR2CSA segment that was able to induce immune response which totally abrogated the adhesion of infected erythrocytes with same efficiency and specificity as the full length extracellular part of the protein. Further, in context of vaccine development, the ID1-DBL2x region also elicited cross-reactive immune response making it a promising vaccine target [27-31].

KEY GAPS IN OUR UNDERSTANDING OF VAR2CSA

In little over 10 years since *var2csa* was identified as being essential to the pathogenesis of PAM, significant progress has been made to characterize the form and function of this *var* gene. *Var2csa* has emerged as one of the primary targets for vaccine candidates against PAM [10, 23, 33, 79]. However, large gaps in our knowledge of *var2csa* remain which must be addressed before going further with vaccine development. First, even though the *var2csa* gene is more conserved than the other *var* genes, some studies have indicated that the binding sites which are exposed to the immune system are highly polymorphic [16, 21, 30, 58, 65, 78]. Knowledge on the extent of genetic diversity of *var2csa* is very important for vaccine development, but remains sparse. Second, given that gravidity modifies the effects of malaria infections in pregnancy, very little data is available for specific *var2csa* variants that are responsible for the naturally occurring gravidity-dependent protective immunity [23]. Third, it is known that PAM is associated with LBW and anemia among several other adverse outcomes. However, no one has identified pathogenic variants or clades of *var2csa* and the magnitude of its association with adverse outcomes. Looking forward, it is likely that any vaccine for pregnant

women will be a polyvalent vaccine; hence, it is critical for vaccine development that the most pathogenic variants are included in the vaccine to maximize the vaccine's efficacy and effectiveness. Additionally, studies have only looked at either genetic sequences of *var2csa* or analyzed immune responses at the protein level. Very few studies have done both. It is very important to identify pathogenic variants of *var2csa* but it is also essential to determine which of the potentially pathogenic variants are actually being transcribed and eliciting an immune response from the host. Fourth, the bulk the research so far on *var2csa* has been *in vitro*. To truly ascertain the pathogenic variants or clades of *var2csa* that can be potential important vaccine candidates, analyzing field isolates from different malaria-endemic areas will provide information on the extent of genetic variation and associations to adverse outcomes that laboratory isolates and *in vitro* are unable to.

RATIONALE

PAM is a significant public health problem that has a high disease morbidity and mortality. The current preventive measures, IPTp and ITNs are only effective if high levels of coverage are achieved and sustained. Recent monitoring and evaluation reports and studies have found that the protective effect of IPTp and ITNs for pregnant women in Africa is inadequate (24.5-35.3%) [5, 53, 55, 80, 81]. There were several reasons that could explain for the less than ideal coverage of IPTp and ITNs, some of which include unclear policies, healthcare infrastructure and system issues (i.e. stock outs), confusion over the timing of each dose of SP, women's poor antenatal attendance, and lack of knowledge about preventive measures [5, 51, 54, 80].

Additionally, parasite resistance to SP conferred by acquiring multiple mutations in *dhps* and *dhfr* genes of *P. falciparum* genome is a major threat to the effectiveness of IPTp. Resistance

to SP increases with accumulation of mutations with the highest level of SP resistance observed with the *dhfr/dhps* “quintuple mutant” - *dhfr* substitutions N51I, C59R, and S108N and the *dhps* substitutions A437G and K540E [82]. The risk of SP failure is higher and rates of parasite clearance are lower among pregnant women residing in areas of widespread SP resistance [83]. High-resistance areas also experience greater number of reinfections and shorter time to reinfections, further providing support of reduced efficacy of SP against malaria infections during pregnancy in face of increasing SP resistance. Despite high levels of SP resistance in several areas in sub-Saharan Africa, use of IPTp-SP remains associated with improved pregnancy and birth outcomes [84, 85]. However, the question is how long will this last before IPTp-SP becomes ineffective at preventing adverse outcomes and necessitate alternate strategies.

Indeed, there are reports that the sextuple mutant - quintuple mutant with an additional *dhps* mutation, A581G exacerbates malaria infections during pregnancy [86]. The sextuple mutant, which hasn't been reported frequently yet, is suggested to confer a significantly greater level of SP resistance than the quintuple mutant alone. The sextuple mutant is also linked with worsened pregnancy and birth outcomes, including low maternal hemoglobin concentrations at birth and birth weight [86, 87].

Spread of pyrethroid resistance caused by the *kdr* mechanism among *Anopheles gambiae s.s.* and *Anopheles funestus* vectors in sub-Saharan Africa has raised concerns regarding effectiveness of ITNs [88, 89]. In areas of moderate pyrethroid resistance, ITNs have remained effective [90], but significant loss of insecticidal effect of ITNs has been reported in areas of high pyrethroid resistance where *kdr* resistance levels are ubiquitous [91]. Subpar efficacy of ITNs can severely undermine prevention of PAM using bed nets [89].

Given the barriers of preventing PAM, a VAR2CSA-based vaccine would be an effective intervention that could achieve and maintain high coverage and supplement existing preventive measures easily. The RTS,S/AS01 vaccine designed for the general population showed suboptimal efficacy (vaccine efficacy = 25.9% - 36.3%) [92, 93], even after a booster dose. It is likely that the effectiveness of the vaccine will be even lower if integrated in routine vaccine campaigns and antenatal care. Indeed, the suboptimal efficacy of RTS,S/AS01 can partly be attributed to plasmodium antigenic diversity [94-97].

PAM is curable and preventable, and an effective vaccine for pregnant women at-risk living in malaria endemic areas can help substantially reduce the disease burden. Currently, phase 1 clinical trials are testing two vaccine candidates (PlacMaIVac and PriMaIVac) that target overlapping constructs in the N-terminus region of VAR2CSA. PlacMaIVac targets the ID1-ID2 region of the FCR3 variant [32] while PriMaIVac targets the DBL1x-DBL2x region of the 3D7 variant (Figure 2.2) [98].

As ID1-DBL2x region of *var2csa* is one of the leading vaccine candidate against PAM, the gaps in our knowledge highlighted above illustrate the paucity of data and underscored how far we are from developing a viable vaccine. Key questions and challenges remain in developing an effective vaccine for PAM which include characterization of sequence polymorphism of *var2csa*, identifying the pathogenic variants or clades and determining number of variants that should be included in the vaccine for optimal efficacy.

INNOVATION

The study is innovative in that it is the one of the first studies that uses long-read next-generation sequencing technologies to characterize a large functional fragment of the *var2csa* gene (~1600 bp) that has shown to be critical in the pathogenesis of PAM. It is known that the

var2csa gene is highly polymorphic and diverse, yet there is very little information on the extent of genetic diversity found in field isolates. This study aims to quantify the genetic diversity of the ID1-DBL2x region of *var2csa* in two clinically characterized cohorts of *P. falciparum*-infected pregnant women in Benin and Malawi. Deep sequencing of placental malaria parasites offer a significant advantage over conventional Sanger sequencing methods as it allows for sequencing of mixed and multiple infections, which are prevalent in high malaria transmission areas such in sub-Saharan Africa. Additionally, deep sequencing allows for identification and quantification of variants occurring at different frequencies, especially low frequencies (<10%) [99, 100].

Placental blood samples from delivering women will be deep-sequenced using the PacBio circular consensus sequencing (CCS) platform. Using PacBio CCS over other commonly used deep-sequencing platforms such as Roche-454, Illumina, or Ion Torrent provide the advantage of being able to sequencing longer fragments (~5kb vs ~800bp). PacBio would allow for sequencing of complete 1.6 kb haplotypes of the ID1-DBL2x region as compared breaking up the fragment in shorter pieces and reassembling haplotypes *in silico*, a method which may complicate haplotype reconstruction.

There has been no study to date that shows an association between specific *var2csa* variants or clades and pathogenicity in pregnancy and adverse birth outcomes. This is the first study that aims to use deep sequencing data in an epidemiological study aimed to identify pathogenic variants or clades of *var2csa* associated with adverse outcomes of malaria in pregnancy. The study is unique in that it incorporates state-of-the-art deep-sequencing, population genetics and epidemiological methods to comprehensively understand the role *var2csa* plays in the pathogenesis of malaria in pregnancy.

Figure 2.1: Global distribution of malaria deaths – 2010. Size of each country is proportional to the malaria deaths in the respective country.

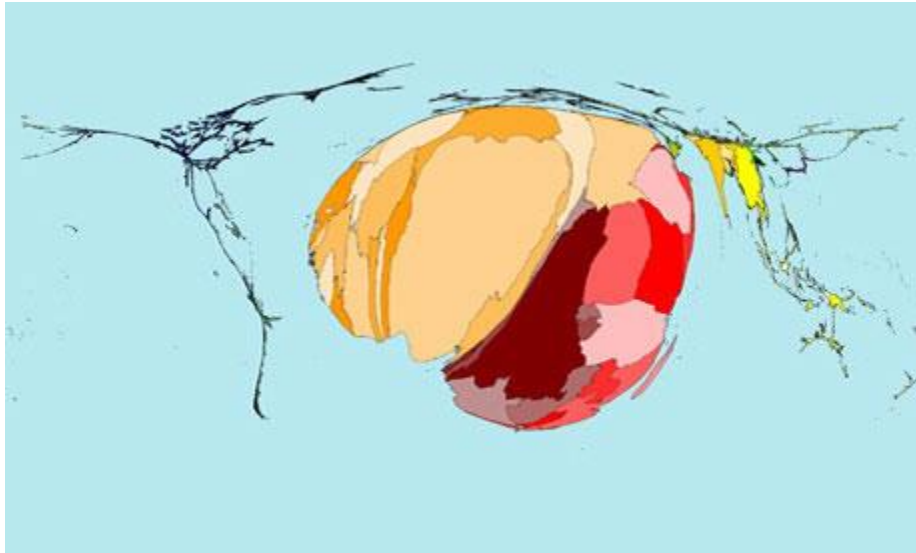
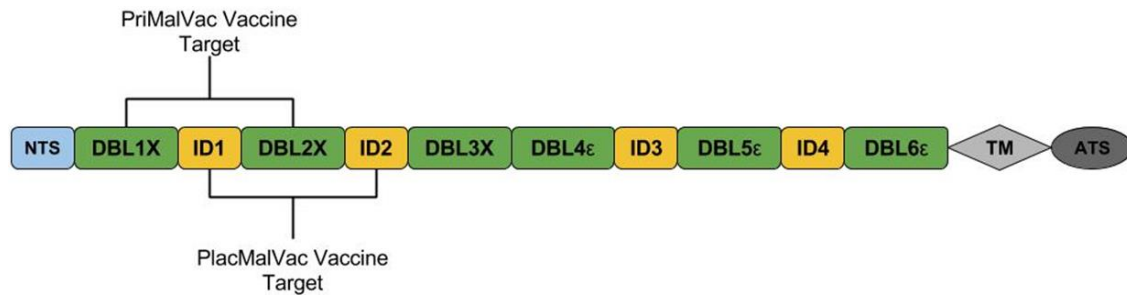


Figure 2.2: Schematic representation of the var2csa gene. VAR2CSA consists of an N-terminal segment (NTS), six Duffy Binding Like Domains (DBL), four cysteine-rich interdomain regions (ID), a transmembrane domain (TM), and an intra-cellular acidic terminal segment (ATS). Current vaccine development efforts (PlacMalVac and PriMalVac) are targeting the NTS-DBL2x region. The ID1-DBL2x region has recently been identified as the minimal CSA-binding region that contains major protective epitopes and elicits a strong host immune response



CHAPTER THREE: DESCRIPTION OF DATA SOURCES

STUDY DESIGN

The study used existing data and samples collected from two previously completed studies in Malawi and Benin. In Malawi, samples were obtained from a randomized controlled trial aimed to assess the efficacy of intermittent screening and treatment in pregnancy (ISTp) with Sulfadoxine-Pyrimethamine (SP) (Madanitsa et al. 2016). The trial was conducted between 2010 and 2013 where 1,873 HIV-negative pregnant women were randomized to receive either at least three doses of IPTp-SP or at least three screenings with a rapid diagnostic test (RDT) or subsequent treatment of RDT-positive cases with Dihydroartemisinin-Piperaquine (DP). In Benin, samples were acquired from a prospective cohort study conducted to quantify the effects of PAM and investigate immunological responses to malaria infection during pregnancy [50]. 1,037 pregnant women were enrolled starting in June 2008 and the last delivery occurred in September 2010. Follow-up in both studies was done during routine antenatal care visits.

Women in both studies were enrolled after obtaining signed informed consent. Both studies were carefully designed and conducted using expert staff and field workers. Several data quality checks were placed in the protocol to ensure minimal errors associated with data collection and error. Data from both studies were of good quality and reliable. The current and parent studies were approved by institutional review boards at Institut de recherche pour le développement in France, Science and Health Faculty (University of Abomey Calavi) in Benin,

Liverpool School of Tropical Medicine, Malawian National Health Science Research Committee, and University of North Carolina at Chapel Hill.

Aim 1 of this study used samples collected from pregnant women enrolled in these studies in Malawi and Benin at delivery. Specifically, blood samples collected from the placental tissue of delivering women were used to sequence and characterize the 1.6 kb fragment of the ID1-DBL2x region of *var2csa*. Aim 2 used genetic sequences generated from Aim 1 as well as clinical and demographic data collected from the pregnant women and infants to identify pathogenic variants and clades of the ID1-DBL2x region that are associated with adverse birth outcomes.

Given, that the sequestration of plasmodium parasites in the placenta is believed to cause the adverse birth outcomes, this study design which uses samples collected at delivery is ideal as it gives confirmatory results of the presence of parasites in the placenta at the time of birth. Additionally, presence of parasites in the placenta is also an indication of malaria infection in the last month of pregnancy, which is a critical period in fetal development.

An alternative study design would be to use placental as well as multiple peripheral blood samples collected throughout the course of pregnancy, specifically early in pregnancy. This would enable characterization of *var2csa* strains in both peripheral and placental infections and track the changes in frequencies in the most pathogenic variants over time and between the two compartments. However, this alternative study design will be limited in scale given the resources and effort required in deep sequencing each sample. The current study design is a critical first step in identifying pathogenic variants of *var2csa* in the placenta and its impact on adverse outcomes

SOURCE POPULATION AND STUDY AREA

The source population included all pregnant women and infants born in the districts of Blantyre and Chikwawa in the southern region on Malawi between September 2010 and October 2013 and in the district of Comé in the Mono province in Benin between June 2008 and September 2010.

Blantyre and Chikwawa are districts located in the southern region of Malawi with a population of approximately 950,000. The districts have 2 tertiary hospitals (one in each district) where women with high risk pregnancies are referred to for care and delivery. The principal malaria vectors in the study area are *Anopheles gambiae s.s.*, and *Anopheles funestus*. Malaria transmission is perennial in these districts with peaks observed during the monsoon season (November-March). *P. falciparum* is the dominant species in circulation as it causes over >90% of all malaria infections (Figure 3.1) [101, 102].

Comé district, with a population of about 58,396 is a semi-rural to rural district located about 70 km west from the largest city and economical capital of Benin, Cotonou. The primary occupations of inhabitants of the study area are farming, fishing, and trading. Health care is provided through three health dispensaries, 11 private clinics, and a district hospital. The area experiences two rainy seasons: from April to July and from September to November. The annual rainfall is over 1300 mm. Like Malawi, the principal malaria vectors in the study area are *Anopheles gambiae s.s.*, and *Anopheles funestus*. Comé district experiences high malaria transmission throughout the year with peak transmission occurring during rainy season. The predominant parasite species in circulation in the region is *P. falciparum* (97%) and the entomological inoculation rate ranges from 35 to 60 infective bites per person per year (Figure 3.2) [103].

This identified source population is ideal for the proposed analyses as the burden of malaria in pregnancy is highest in sub-Saharan African countries such as Benin and Malawi. Further, malaria in pregnancy has been well characterized in the study area through numerous studies [50, 85, 104-111]. Both studies from which the data for this project will be acquired have prospectively followed pregnant women through delivery and have high quality data. Further, we believe the results from this study are generalizable to pregnant women living in other malaria endemic sub-Saharan African countries and potentially to countries outside the African sub-continent.

STUDY POPULATION

The study population included pregnant women who were residing in Blantyre and Chikwawa districts between 2010 and 2013 in Malawi and in the Comé district between 2008 and 2010 in Benin.

In Malawi, HIV-negative pregnant women were enrolled in the trial starting September 2010 in the antenatal clinics in Mpemba and Madziabango in Blantyre and Chikwawa districts, respectively (Figure 3.1). The inclusion criteria included singleton pregnancy, gestational age of 16 to 28 weeks, no history of IPTp use during current pregnancy, resident of the study area, and willing to deliver at the study clinic or hospital. Women who met the inclusion criteria were enrolled in the trial after obtaining informed consent and were randomized to receive either at least three doses of IPTp-SP or at least three screenings with an RDT and subsequent treatment of RDT-positive cases with DP. Newborn infants to the women enrolled in the trial were clinically examined and anthropometric measurements were recorded. Women delivering outside the study clinic or hospital were identified by a network of community nurses and traditional

birth attendants and birth outcomes along with other relevant information were recorded within two days after the delivery (Madanitsa 2016).

In Benin, women were enrolled starting in June of 2008 in the three dispensaries: Comé, Akodeha, and Ouedeme Pedah (Figure 3.2). The inclusion criteria consisted of gestational age under 24 weeks, living within 15km from the dispensary for >6 months, and having planned to deliver at the hospital. Local midwives performed gynecological examinations and collected clinical information on the pregnant women. Pregnant women who met the inclusion criteria were then enrolled after study objectives were explained twice and informed consent was obtained. Nurses from the community were recruited and trained as project assistants to complete questionnaires and collect blood samples from the study participants. In addition to the pregnant women, the newborn infants were clinically examined and anthropometric measures were collected. For women delivering outside the study time frame, birth outcomes were collected from antenatal care book or if the women presented at the study center within two days after delivery [50].

For both aims of this study, we included women testing positive for malaria at delivery in their placental blood samples via microscopy or PCR for whom placental blood samples were available for ID1-DBL2x genotyping.

DEMOGRAPHIC AND CLINICAL DATA

In Malawi, a baseline assessment was conducted at enrollment which included collection of demographic and socio-economic information, use of ITNs, IRS, medical and reproductive history. Anthropometric measures collected on mothers included height, weight, fundal height, last menstrual period, and an ultrasound to determine gestational age. The participants were then

randomized to receive the intervention (ISTp-DP) or standard-of-care (IPTp-SP) (Madanitsa et al. 2016).

During routine antenatal visits, clinical symptoms and illnesses experienced in the last month were recorded. The same clinical and biological information as initial visit was collected and changes in the demographic data were noted. As part of routine antenatal care in Malawi, women received ITNs, iron and folic acid supplements, tetanus vaccination if applicable, praziquantel (40 mg/kg single dose) for schistosomiasis, albendazole (400 mg/kg single dose) for hookworm, trichuriasis and ascariasis, and albendazole (400 mg/kg per day for 3 days) for strongyloidiasis (Madanitsa et al. 2016).

At delivery, information on the newborn infants included weight, vital signs, and congenital anomalies. Gestational age was assessed clinically by using the new Ballard score. After 24 hours, the baby was re-examined for jaundice (Madanitsa et al. 2016).

In Benin, at the initial visit, for each pregnant woman, information regarding reproductive history, current pregnancy, medical history, socio-economic indicators, and use of malaria preventive measures such as insecticide treated bed nets was collected. Anthropometric measures collected on mothers included axillary temperature, blood pressure, and the mid-upper arm circumference (MUAC). Four ultrasound scans were also performed using a portable ultrasound system to determine exact term of pregnancy and to evaluate intrauterine growth and fetal morphology [50].

During monthly follow-up antenatal care visits, clinical symptoms and illnesses experienced since last follow-up visits were recorded. The same clinical and biological information as initial visit was collected and changes in the demographic data were noted. At

delivery, measurements of axillary temperature, weight, blood pressure of the mother were collected [50].

According to national guidelines, pregnant women were given monthly kits that included tablets of iron (200 mg to be taken daily), folic acid (5 mg to be taken daily), and mebendazole (500 mg for 3 days) for deworming along with ITNs. Additionally two doses of sulfadoxine pyrimethamine (SP) (1500 mg of sulfadoxine and 75 mg of pyrimethamine) were administered as IPTp at least one month apart starting in the second trimester of pregnancy [50].

Information collected on the newborn infants included the APGAR score, a method that rapidly assess the health of the newborn immediately after birth [112], indications for icterus and malformations, anthropological measurements that included weight, height, MUAC, head circumference, abdominal circumference, and foot length. Gestational age was assessed clinically by using the new Ballard score [113].

OUTCOME ASSESSMENT

It is known that PAM is associated with poor infant health and survival. It is believed that PAM affects infant's health through preterm delivery and low birth weight through fetal growth restriction among other pathways. The outcomes of interest in this study were adverse birth outcomes. Birth weight was the primary outcome of interest in this study. Birth weight was measured by an electronic scale at the time of the infant's birth in both studies in Benin and Malawi. If a birth occurred outside the study frame, birth outcomes were assessed from antenatal cards including birth weight in Benin or within 48 hours of birth in person in Malawi. Birth weight was recorded as a continuous variable in grams. Gestational age was determined by four ultrasound measurements during the course of the pregnancy in Benin and one ultrasound in Malawi.

Birth weight was included in the study as a categorical as well as a continuous outcome. Low birth weight (LBW) was defined according to the WHO's criteria as infant with a birth weight of less than 2500 grams [114]. Additionally, to account for confounding in birth weight due to gestational age [115-119], small-for-gestational age (SGA) was calculated for each infant using nomograms from previous studies from the study areas. SGA, which takes into account gestational age as well as birth weight give a more accurate representation of fetal growth in the study population [120, 121]. SGA was defined as infant with a birth weight below the 10th percentile for babies of the same gestational age at delivery

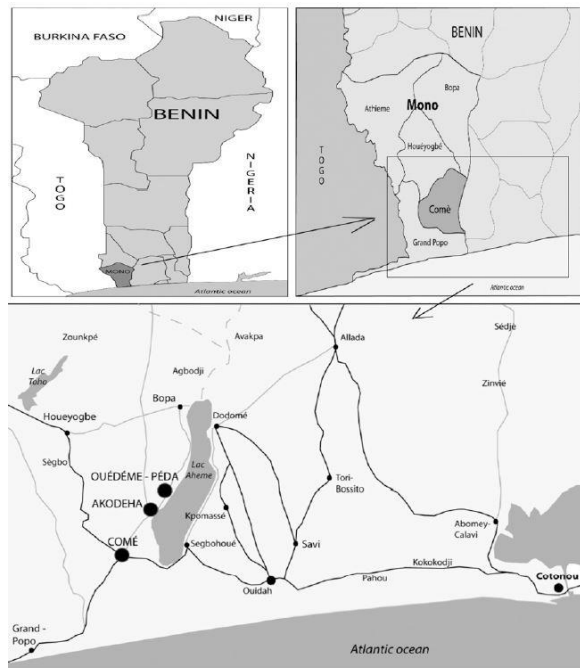
EXPOSURE ASSESSMENT

The main exposure of interest for this study was the genetic sequence of the ID1-DBL2x region of *var2csa*. The ID1-DBL2x region was characterized by targeted deep-sequencing using the PacBio consensus deep-sequencing (CCS) platform. Sequences from PacBio were clustered to using a k-mer clustering algorithm to generate a list of unique ID1-DBL2x variants and their relative frequencies infecting the pregnant women in our study population. Phased consensus genetic sequences from each woman were used in the analyses of both aims. Additionally, results of the phylogenetic clustering performed in aim 1 were used to help inform epidemiologic modeling approaches in aim 2.

Figure 3.1: Malawi Study Site



Figure 3.2: Benin Study Site



CHAPTER FOUR: METHODS

DNA EXTRACTION

Among samples from the ISTp trial in Malawi, genomic DNA was extracted from dried blood spots (DBS) using 20% Chelex-100 (Bio-Rad, Richmond, CA) [122]. DBS were hole-punched into 96-well plates and incubated overnight at 4 °C in 1ml of 1x PBS and 50 uL of 10% saponin. After a wash with 1x PBS, 50 µl of 20% Chelex was added to each sample and incubated at 95°C for 12 minutes. Genomic DNA in the solution was aspirated to separate it from the Chelex beads and stored at -20°C for downstream molecular testing and PCR amplifications.

For Benin, genomic DNA was extracted from RBC pellets using Thermo Scientific GeneJET DNA extraction kit (Catalog #K0781) as recommended by the manufacturer. Briefly, packed RBCs were digested with Proteinase K and incubated with ethanol at 56°C for ten minutes. DNA was separated from the lysate via purification column, and eluted in water, and stored at -20°C.

ID1-DBL2X PCR ASSAY

We developed a hemi-nested PCR amplification strategy to amplify the 1.6 kb ID1-DBL2x region of *var2csa* from genomic DNA. First all publicly-available genetic sequences of *var2csa* from GenBank and PlasmoDB were downloaded and aligned in MEGA 6.0 [123]. Published primers targeting the 1.6 kb region [28] were superimposed on the alignment to ensure

that the primers were indeed in the conserved region. An outer reverse primer was designed for the first round of the hemi-nested PCR (Table 4.1). Barcodes were attached to the forward and reverse primers for the second round PCR (Appendix A). The second round of hemi-nested PCR was designed to be a touchdown PCR in order to increase the specificity of the DNA amplification.

The PCR reaction mix for the first round contained: 2.5µl of Roche Hi-Fi buffer, 0.5 µl of 10 mM dNTPs, 20 µM of each primer, 1.2 µl of 25 mM MgCl₂, 0.25 µl of Roche Hi-Fi DNA polymerase, 1.25 µl of DMSO, and 5 µl of DNA template in a 25 µl reaction. The first round PCR cycling conditions were as follows: 95°C for 2m, 35 cycles of 95°C for 30s, 52°C for 30s, and 72°C for 3min, and a final extension of 72°C for 7m. PCR products from the primary round were then used as DNA template for the second round PCR. The PCR reaction mix for the second round contained: 2.5µl of Roche Hi-Fi buffer, 0.5 µl of 10 mM dNTPs, 20 µM of each barcoded primer, 0.25 µl of Roche Hi-Fi DNA polymerase, 1.25 µl of DMSO, and 2 µl of primary PCR product in a 25 µl reaction. The second round PCR cycling conditions were as follows: 95°C for 2m, 15 cycles of 95°C for 30s, 67-52°C for 30s (-1°C per cycle during first 15 cycles, then 52°C for 25 cycles), and 72°C for 3m, and a final extension of 72°C for 7m. All second round PCR products were visualized on a 1% agarose gel.

PACBIO CIRCULAR CONSENSUS SEQUENCING (CCS)

Genomic DNA from each clinical sample was amplified in technical duplicate PCR reactions using unique forward and reverse barcoded primers for each replicate. Concentrations of successful amplifications were quantified using the Genomic DNA ScreenTape assay on the Agilent 2200 TapeStation system. PCR products were then pooled together for deep sequencing on PacBio CCS platform using the P5-C4 chemistry. PCR products with unique barcodes were

mixed in an equimolar fashion for each PacBio Single Molecule, Real-Time (SMRT) sequencing cell. This was done to ensure all PCR products would be sequenced with equal probability. Each PacBio SMRT cell was gel extracted and purified using Qiagen gel extraction kit as per manufacturer's protocol. The SMRT cells were quantified again to ensure adequate concentration and quantity of DNA was present before sequencing.

CLUSTERING OF NEXT-GENERATION SEQUENCING DATA

Raw sequences from PacBio were cleaned and clustered using a novel clustering algorithm based on k-mer distances [124]. First, sequences sharing the same barcodes were clustered using the k-mer clustering algorithm. For each sequence, all its k-mers were indexed and k-mer distances were created by counting the k-mers shared between two sequences and dividing by the maximum number of k-mers the two sequences could share. The range of values for k-mer distance is between 0 and 1 with 0 being no k-mers shared and 1 being all k-mers shared. K-mer distances were calculated at several k-mer lengths and the rate at which the k-mer distance decreased with increasing length was compared to monoclonal PacBio samples rates to cluster similar sequences. Second, after within-PCR replicate (same barcode) clustering, between-PCR replicate (same samples) or population clustering was performed to determine the similarity between technical replicates from each sample, identify different variants of the ID1-DBL2x region, and estimate the frequencies of each variant. Identification of variants and estimation of their frequencies was within each sample as well as between samples at a population level. The end result of k-mer clustering yielded one or more phased consensus sequences for each woman present in the study population (Figure 4.1).

VALIDATION STUDY

As the PCR assay and bioinformatics clustering of a long fragment of *var2csa* has not been previously reported, we conducted a validation study using pools of reference genomic parasite lines to ensure that our assay and clustering algorithm was appropriate and accurate. We mixed seven reference parasite lines (3D7, FCR3, 7G8, DD2, K1, RO33, and V1/S) in varying frequencies (1-50%) in five pools and deep-sequenced it on the PacBio CCS platform. The sequences derived from deep-sequencing were then clustered using the k-mer clustering algorithm. We estimated the component parasite lines for each pool and their relative frequency. We then compared the results of de novo clustering to the expected and reference-guided observed frequencies.

SPECIFIC AIM 1

STUDY SAMPLE

The study sample for Specific Aim 1 included pregnant women enrolled in the parent studies in Benin and Malawi whose placental samples collected at delivery were amplified and deep-sequenced successfully and for whom we had clinical and demographic data. To characterize the genetic diversity of the ID1-DBL2x region, we calculated measures of alpha (within-group) and beta (between-group) diversity. We then compared these measures between the two countries and between women of differing gravidities.

CHARACTERIZING ALPHA (α) DIVERSITY

To estimate the alpha diversity, we calculated rarefaction curves, expected heterozygosity (H_e), Shannon index (H'), abundance coverage estimator (ACE), incidence coverage estimator (ICE), Chao richness estimators, and maximum composite likelihood (MLE) phylogenetic trees.

To estimate species richness of the ID1-DBL2x region of *var2csa*, we calculated rarefaction curves [125] in Malawi and Benin as well as by gravidity. As observed species richness is highly sensitive to sample size and diversity of the marker being used, a simple ratio of species per sampling unit may distort richness values. Hence rarefaction curves were calculated taking into account how sampling was conducted. Rarefaction curves of haplotypes were calculated in EstimateS [126] as:

$$S_{sample}(t) = S_{obs} - \sum_{Y_i > 0} \left[\frac{\binom{T - Y_i}{t}}{\binom{T}{t}} \right]$$

where, $S_{sample}(t)$ is expected number of species, S_{obs} is the actual number of species observed, T is the total number of sampling units, t is the incident sample, and Y_i is the observed species incidence frequencies. Rarefaction curves were bootstrapped 1000 times with replacement to generate 95% confidence intervals. Since *var2csa* is highly diverse and we were likely to find many more variants of *var2csa* through next-generation sequencing, we extrapolated the rarefaction curves by 100 additional samples to predict how species richness would change if our sample size was larger.

Expected heterozygosity (H_e) was calculated at each locus within each country and within women of differing gravidities as [127]:

$$H_e = \left[\frac{n}{n-1} \right] \left[1 - \sum_{i=1}^n p_i^2 \right]$$

where n is the number of isolates sampled and p_i is the frequency of the i th allele. All calculated estimates of H_e were then bootstrapped 1000 times to estimate the precision of the estimates and test if H_e was statistically different between the two countries and different by gravidity.

Additionally, to examine how genetic diversity in the variants affected protein sequences, we

calculated H_e at each amino acid position among variants in Malawi and Benin from translated protein sequences.

Shannon index (H') was calculated for the ID1-DBL2x populations within each country and within women of differing gravidities as [128]:

$$H' = - \sum_{i=1}^R p_i \ln p_i$$

where p_i is the proportion of individuals found in the i th species. All estimates of H' were bootstrapped 1000 times with replacement to generate precision estimates.

Abundance coverage estimator (ACE) was calculated for the ID1-DBL2x populations within each country and within women of differing gravidities as [129]:

$$S_{ace} = S_{abund} + \frac{S_{rare}}{C_{ace}} + \frac{F_1}{C_{ace}} \gamma_{ace}^2$$

where S_{abund} is the number of abundant species when all samples are pooled, S_{rare} is the number of rare species (each with 10 or fewer individuals) when all samples are pooled, C_{ace} is the sample abundance coverage estimator, F_1 is the frequency of singletons, and γ_{ace}^2 is the estimated coefficient of variation of singletons. All estimates of ACE were bootstrapped 1000 times with replacement to generate precision estimates.

Incidence coverage estimator (ICE) was calculated for the ID1-DBL2x populations within each country and within women of differing gravidities as [129]:

$$S_{ice} = S_{freq} + \frac{S_{infr}}{C_{ice}} + \frac{Q_1}{C_{ice}} \gamma_{ice}^2$$

where S_{freq} is the number of frequent species (found in >10 samples), S_{infr} is the number of infrequent species (found in <10 samples), C_{ice} is sample incidence coverage estimator, Q_1 is the frequencies of unique species, and γ_{ice}^2 is the estimated coefficient of variation of unique species.

All estimates of ICE were bootstrapped 1000 times with replacement to generate precision estimates.

The Chao estimates were calculated for the ID1-DBL2x populations within each country and within women of differing gravidities as [130]:

$$S_{Chao1} = S_{obs} + \left(\frac{n-1}{n} \right) \frac{F_1(F_1-1)}{2(F_2+1)}$$

where S_{obs} is the total number of species observed in all samples pooled, n is the number of samples, F_1 is the frequency of singletons, and F_2 is the frequency of doubletons.

$$S_{Chao2} = S_{obs} + \left(\frac{m-1}{m} \right) \left(\frac{Q_1(Q_1-1)}{2(Q_2+1)} \right)$$

where S_{obs} is the total number of species observed in all samples pooled, m is the number of samples, Q_1 is the frequencies of unique species, and Q_2 is the frequencies of duplicate species. Statistically significant differences in alpha diversity metrics between different groups were tested using Kruskal–Wallis test. An α of 0.05 was determined a priori to test for significant differences.

Using genetic sequences from the haplotypes, we constructed phylogenetic trees using the maximum composite likelihood method in MEGA6. Genetic distances in the maximum composite likelihood method were calculated using the Tamura-Nei model [131]:

$$d = -b \log_e(1 - p/b)$$

where p is the proportion of sites with different nucleotides and

$$b = \frac{1}{2} \left[1 - \sum_{i=1}^4 g_i^2 + p^2/c \right]$$

$$c = \sum_{i=1}^3 \sum_{j=i+1}^4 \frac{x_{ij}^2}{2g_i g_j}$$

where x_{ij} is the relative frequency of the nucleotide pair i and j , g_{is} are the nucleotide frequencies. The maximum composite likelihood phylogenetic trees were bootstrapped 1000 times to compute branch support and provide precision in the differences observed in the trees. Bootstrap values of greater than 80% were used as cutoff point to identify significant differentiation between variants and clusters. All phylogenetic trees were visualized using the APE package for R [132].

CHARACTERIZING BETA (β) DIVERSITY

To characterize genetic relatedness among the ID1-DBL2x variant populations between Malawi and Benin as well as between the different clades, we calculated Wright's fixation index (F_{ST}) [133] as.

$$F_{ST} = \frac{\pi_{Between} - \pi_{Within}}{\pi_{Between}}$$

where $\pi_{Between}$ is the average number of pairwise nucleotide differences between two variants from two different sub-populations and π_{Within} is the average number of pairwise nucleotide differences between two variants from the same sub-population. The F_{ST} index quantifies genetic relatedness based on allele frequencies among population and ranges from 0 to 1, where 0 signifies a genetically identical population and 1 signifies completely differentiated populations. The F_{ST} analyses were performed using a sliding window to identify regions along the vaccine target that would account for the genetic relatedness between countries and clades.

We performed a principal coordinate analysis (PCoA) to assess the genetic relatedness between population of variants from different countries, gravities, and clades identified by

phylogenetic clustering. F_{ST} values were used as the genetic distance matrix for the PCoA. The top coordinates explaining the most amount of variation were used to visualize the genetic relatedness between the clades.

We calculated nucleotide diversity (π) and Tajima's D test [134] on the entire ID1-DBL2x region as well as with a sliding window approach to assess the selection pressures acting upon vaccine target[135]. Tajima's D was calculated on ID1-DBL2x variant populations by country and major clades as follows:

$$D = \frac{\hat{\pi} - \hat{\theta}_w}{\sqrt{\text{Var}(\hat{\pi} - \hat{\theta}_w)}}$$

where π is the average number of pairwise differences and Θ_w is the Watterson's estimator of Θ calculated as:

$$\hat{\theta}_w = \frac{S_n}{a_n}$$

where S_n is the number segregating sites in a population and a_n is:

$$a_n = \sum_{i=1}^{n-1} \frac{1}{i}$$

All measures of genetic relatedness and selection (F_{ST} , PCoA, π , Tajima's D) were calculated in R using the `adegenet` [49] and `popgenome` [56] packages.

SPECIFIC AIM 2

STUDY SAMPLE

The study sample for Specific Aim 2 included pregnant women enrolled in the parent studies in Benin and Malawi whose placental samples collected at delivery were amplified and deep-sequenced successfully and for whom we had clinical and demographic data.

PHYLOGENETIC AND STATISTICAL ANALYSES

We examined associations between clades and birth outcomes to identify pathogenic clades. Outcomes of interest were birth weight, low birth weight (LBW), and small-for-gestational age (SGA). At delivery, infant birth weight was recorded to the nearest gram within 24 hours using a calibrated digital scale. LBW was defined according to World Health Organization's criteria as infant with a birth weight of less than 2500 grams [114]. SGA was defined as infant with a birth weight below the 10th percentile for babies of the same gestational age at delivery [120, 121, 136].

Unique variants identified in our study population were analyzed phylogenetically to detect clustering. We constructed phylogenetic trees using the maximum composite likelihood (MLE) method in MEGA6 [123]. Genetic distances in the maximum composite likelihood method were calculated using the Tamura-Nei model [131]. The MLE phylogenetic trees were then bootstrapped 1000 times to compute branch support and provide precision in the clustering detected in the trees. All phylogenetic trees were visualized using the APE package for R [132]. Bootstrap values of greater than 80% were used as cutoff point to identify significant differentiation between variants and clusters.

Associations between specific clades and birth outcomes were first explored by examining the clustering of adverse birth outcomes (i.e. LBW and SGA) in the MLE

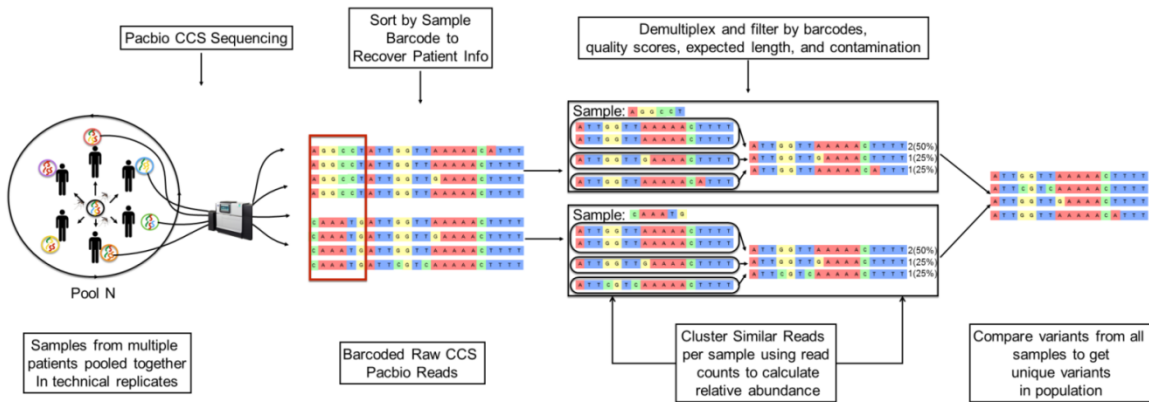
phylogenetic trees in Malawi or Benin, respectively. Pregnant women were categorized based on the respective clade in which the variants detected from their samples clustered. We then used linear and logistic regression to estimate the effect of clades on birth weight, LBW, and SGA, respectively. We controlled for potential confounders and standardized the groups defined by clades to the marginal distribution of potential confounders in the study population using inverse probability weights. The denominator of the weights for the different clades of ID1-DBL2x was estimated using polytomous logistic regression, conditional on parity. For the pooled analyses, we also included country in estimating the denominator of the weights. Weights were stabilized using the marginal distribution of the ID1-DBL2x clades in our study population. We estimated differences in birth weight and relative odds of LBW and SGA associated with infection with each clade, using FCR3 as the referent category. All analyses were conducted using SAS version 9.3 (SAS Institute, Cary, North Carolina).

Table 4.1: ID1-DBL2x hemi-nested PCR primer sequences

Primer Name	5'→3'
ID1-F*	GATCCTTATTCCGCAGAATA
CIDR-R_Heminested	TTTCTTTGTTCCACTGTTCAA
CIDR-R*	GTCGTGTATGTTGTCCA

*Bordbar et. al, 2014

Figure 4.1: Schematic representation of the bioinformatics clustering pipeline employed in this study. Pools of amplified ID1-DBL2x from individual women in technical replicates using barcoded primers were sequenced on PacBio CCS platform. Reads were clustering using a k-mer algorithm to identify unique ID1-DBL2x variants in the study population.



CHAPTER FIVE: RESULTS. GENETIC DIVERSITY AND POPULATION STRUCTURE OF VAR2CSA ID1-DBL2X IN MALAWI AND BENIN: IMPLICATIONS FOR VACCINE DEVELOPMENT

STUDY POPULATION

P. falciparum was detected by PCR in the placental blood samples of 281(18.8%) women in Malawi and 175(27.6%) women in Benin.; of these, 281 (100%) and 126 (72%) samples were available from Malawi and Benin, respectively. Of the samples available, we were able to successfully amplify the 1.6kb ID1-DBL2x region of *var2csa* in a total of 101 samples – 56 (19.9%) from Malawi and 45 (37.5%) from Benin, constituting the analytic population. The mean maternal age of the 101 women was 22.5 years (standard deviation (SD) ± 5.5 years) and the mean gestational age at delivery was 38.1 weeks (SD ± 2 weeks). Of the women included in our analyses 38.6% (n = 39) were primigravid, 28.7% (n = 29) were secundigravid, and 32.7% (n = 33) were multigravid (Table 5.1). Among the available *P. falciparum* positive specimens, there were no significant differences between women whose samples amplified successfully and those whose samples did not with respect to demographic characteristics (Table 5.2). Samples from Benin amplified more readily than those from Malawi; this variability was likely due to differences in sample storage (DBS versus RBCs) across study sites.

VALIDATION OF GENOTYPING APPROACH

We conducted a validation study to determine if our laboratory and bioinformatics approach was sensitive and precise at genotyping ID1-DBL2x. To do so, pools of seven genomic parasite lines in varying frequencies were amplified, sequenced, and clustered via our PCR assay

and k-mer clustering pipeline. In five distinct multiclonal template mixtures, we were able to accurately detect each of the seven parasite genomic lines. Additionally, we detected parasite haplotypes that were included in as little as 5% of the mixed template (Figure 5.1). Quantitatively the observed frequencies of each parasite line correlated well with the expected frequencies ($R^2=0.82$) (Figure 5.2). The results from k-mer clustering showed reliable variant calling and clustering of ID1-DBL2x variants. The clustering of variants using the k-mer algorithm produced similar results as the reference-guided clustering (Figure 5.1 and Figure 5.2).

ID1-DBL2X DEEP-SEQUENCING AMONG CLINICAL ISOLATES

For our clinical samples, a total 523,482 CCS reads were obtained from 14 PacBio SMRT cells. After extracting reads by barcoded primer sequences and filtering based on size and quality requirements, 256,377 (48.98%) reads were used for k-mer clustering to identify unique variants in our study population (Figure 5.3). The median sequence length from the extracted reads was 1594bp (range: 1400 – 1700bp). The median coverage from CCS was 1795 reads per sample (range: 16 – 9241).

Overall, we identified 152 unique ID1-DBL2x variants, 95 in Malawi and 57 in Benin. Of these 152 variants, 17 were observed in more than one woman. Interestingly, one variant was shared between Malawi and Benin (Table 5.1). The mean multiplicity of infection (MOI) was 1.88 (range: 1-7) in Malawi and 1.56 (range: 1-4) in Benin (Table 5.1). Overall the ID1-DBL2x fragment showed significant genetic variation with ~30% ($n = 470$) of the region containing polymorphic sites (Table 5.3).

WITHIN-GROUP (ALPHA) DIVERSITY OF ID1-DBL2X

The within-group (alpha) diversity of ID1-DBL2x at the two study sites was assessed by rarefaction, which compensates for the difference in depth of sampling (Figure 5.4). The two

curves overlap for most of the sampled region but only the extrapolated curve for Malawi approaches its asymptote (Figure 5.4). This suggests that ID1-DBL2x is more diverse in Benin as compared to Malawi. This finding is corroborated by other alpha diversity metrics (Table 5.4). The metrics which do not compensate for sampling (H_e and H') showed no difference between the sites while the metrics that do (ACE, ICE, and Chao), as predicted, demonstrate significantly greater diversity in Benin. H_e at the amino acid level along the ID1-DBL2x fragment showed regions of high diversity separated by conserved regions. Amino acid diversity was also notably higher in Benin, especially in the ID1 region (Figure 5.5).

PHYLOGENETIC CLUSTERING

We then compared all variants in Malawi and Benin phylogenetically to identify any clustering of ID1-DBL2x in our study population. The maximum-likelihood phylogenetic trees clustered the variants in Malawi in two significantly distinct clades (bootstrap values >80%). The two clades co-clustered with the two referent strains which were used to design the current vaccine candidates (3D7 and FCR3) (Figure 5.6). In contrast, the phylogenetic trees in Benin clustered the variants into five significantly distinct clades (bootstrap values >80%). In addition to the 3D7-like and FCR3-like clades identified in Malawi, three clades unique to Benin were detected (Figure 5.6).

The principal coordinate analysis (PCoA) showed that the vast majority of the difference in the populations was explained by clade, rather than by geography or country. The first two principal coordinates (PC1 & PC2) accounted for 21% of the variation. As expected, the PC1-PC2 plot showed a significant differentiation between the 3D7 and FCR3 clade. Interestingly, while clades 4 and 5 appear to be closely related to 3D7, variants in clade 3 displayed a genetic background resembling both 3D7 and FCR3 clades, indicating a possible recombination event

(Figure 5.7). The PC1-PC3 plot (Figure 5.7) showed similar genetic relatedness between clades as was observed in the PC1-PC2 plot.

BETWEEN-GROUP (BETA) DIVERSITY OF ID1-DBL2X

To study the genetic relatedness between ID1-DBL2x populations in our study, we calculated Wright's Fixation index (F_{ST}) with a sliding window over the entire 1.6kb region for pairwise comparison between countries, women with differing gravidities, and the two vaccine clades (3D7 and FCR3). The F_{ST} values comparing variants from women with differing gravidities approached zero (range: 0-0.11), indicating high degree of genetic similarity (data not shown). The F_{ST} values between Malawi and Benin were again close to zero indicating that the ID1-DBL2x populations in the two countries did not differ significantly as we detected variants from the two major clades in both countries. However, when comparing F_{ST} values between 3D7 and FCR3 clade, overall the two populations look similar except for a ~100bp region where the two populations differ significantly ($F_{ST} = 0.26-1.0$) (Figure 5.8).

DETECTING SIGNATURES OF SELECTION WITHIN ID1-DBL2X

Detecting signatures of selection on the vaccine target was done through calculating Tajima's D across the entire ID1-DBL2x region. The Tajima's D test statistic can indicate whether a nucleotide sequence is under directional selection ($D < 0$), genetic drift ($D = 0$), or balancing selection ($D > 0$). Among all groups (i.e. country and gravidity), the ID1-DBL2x variants in our study population demonstrated significant signatures of balancing selection (range: 1.183 – 1.976) (Table 5.3). To determine whether the entire region was uniformly under balancing selection, we applied Tajima's D test using a sliding window approach. Scanning across the 1.6kb fragment, we observed regions under high balancing selection along with regions of directional selection (Figure 5.9). The dimorphic region between the 3D7 and FCR3

clades displayed the strongest signal for balancing selection indicating that the region interacts with human immune system and may play an important role in vaccine efficacy.

DISCUSSION

The ID1-DBL2x region of VAR2CSA is an important PAM vaccine candidate, as it is the minimal CSA-binding region that elicits an immune response. Phase 1 clinical trials are testing two vaccines (PlacMaIVac and PriMaIVac) that target overlapping constructs of distinct variants of this region [32, 98]. PlacMaIVac targets the ID1-ID2 region of the FCR3 variant while PriMaIVac targets the DBL1x-DBL2x region of the 3D7 variant (Figure 2.2). Our study on the genetic diversity of ID1-DBL2x can directly inform these ongoing vaccine development efforts.

The results from our study demonstrate that ID1-DBL2x is indeed highly diverse in Malawi and Benin. The population genetic analyses indicate high species richness and within group diversity of the ID1-DBL2x population in both countries and among women with differing gravidities. Furthermore, signatures of selection indicate that the region is primarily under balancing selection. Most importantly, we also report significant clustering of variants into distinct clades that will likely warrant a polyvalent vaccine and an approach that can take the relevant genetic diversity into account.

To our knowledge this is the first study to employ long-read next-generation sequencing of the 1.6kb ID1-DBL2x region from clinical isolates to characterize the genetic diversity of this vaccine target. Previous studies have relied on Sanger sequencing or cloning, both of which limit depth and sensitivity, or on breaking up the fragment into smaller fragment to deep-sequence and reassembling haplotypes *in silico*, a method which may complicate haplotype reconstruction [16, 17, 24, 28, 38, 78, 137]. The PacBio CCS platform allowed us to sequence a long fragment of *var2csa* at great depth without requiring assembly of haplotypes. Additionally, barcoding

individual samples allowed us to link variants detected through deep-sequencing to individual women and their clinical and demographic data. Due to the high rates of polymorphism, polyA repeats, and PacBio error rate, variant calling through conventional software was not adequate. As a result, we implemented a k-mer clustering algorithm which provided a more appropriate way to determine unique variants of *var2csa*.

Our population genetic analyses revealed high species richness of the sequenced fragment of *var2csa* in both Malawi and Benin. This is in concordance with previously reported overall species richness of *var* genes [138]. Using F_{ST} as a measure of genetic relatedness, we observed that overall, the ID1-DBL2x populations in Malawi and Benin were very similar and the genetic differences observed were not explained by geography. Likewise, the ID1-DBL2x variant populations were similar between primi-, secundi-, and multigravid women. This suggests that VAR2CSA-based vaccines targeting the ID1-DBL2x region can be effective across geographic and demographic subpopulations.

We found that the ID1-DBL2x variants clustered in distinct phylogenetic clades. Two of the clades which contained the vaccine referent strains, 3D7-like and FCR3-like variants, were found in both countries. In addition to those two clades, we detected three other clades in Benin that were not present in Malawi. A similar clustering has been reported previously [139]. Variants from one of the clades unique to Benin (clade 3) displayed genetic similarities to both 3D7 and FCR3 clades in the PCoA analyses. This suggests that the variants in clade 3 are a product of potential recombination as they contained a 3D7 as well as a FCR3 genetic backbone. The variants from the two major (3D7 and FCR3) clades were largely similar to each other across the 1.6kb region except for a ~100bp region that was very different between the two clades. This dimorphic region reported before in West Africa is important as this difference

likely explains the suboptimal cross-protective effect (between 3D7 and FCR3 clades) observed among the current vaccine candidates [31, 137].

The ID1-DBL2x region is under balancing selection as indicated by overall positive Tajima D's values. This corroborates its role as an important epitope. Scanning across the 1.6kb region, we detected signatures of both balancing as well as directional selection. While the ID1-DBL2x populations in Malawi and Benin displayed similar signatures of selection, the variants in 3D7 and FCR3 clades displayed contrasting signatures of selection in the ID1 region. Additionally, the dimorphic region between the 3D7 and FCR3 clades is under strong balancing selection, suggesting that it is a target of host immunity. Taking the relevant genetic diversity into account when designing a vaccine will be critical in producing a candidate that has excellent cross-protective effect. Efforts need to continue further characterization of the genetic diversity of *var2csa* vaccine targets among clinical isolates and identify pathogenic variants and clades that are associated with adverse pregnancy and birth outcomes. Including these pathogenic variants and clades in a polyvalent vaccine can help maximize the vaccine's efficacy and effectiveness.

We were limited in the number of samples we were able to include in our study. This was due to the difficulty in amplifying a long fragment of a hypervariable gene from DNA extracted from dried blood spots in Malawi. The generalizability of our results is also limited as we had access to samples from only two countries in Africa. However, the results from our study agree with other studies that were able to deep-sequence smaller fragment of *var2csa* from multiple countries in Africa as well as from Asia and South America [28].

We employed an integrative approach, using molecular epidemiology, next-generation sequencing, and population genetics to examine the antigenic diversity of *var2csa*, identify

immunologically relevant selection pressures and signatures of genetic relatedness to inform vaccine development efforts. Previous attempts to develop a general vaccine against malaria have not taken antigenic diversity into account; this is one possible explanation for suboptimal efficacy and effectiveness [92, 95]. The results from our study provide further support for the development of polyvalent vaccines against malaria. As clinical trials get underway for testing vaccines against PAM, characterizing genetic diversity and monitoring parasite populations will help explain changes that affect vaccine efficacy and inform future vaccine improvement and development efforts.

Table 5.1: Description of the study population

Group	
Samples, n (%)	101 (100)
Benin	45 (44.6)
Akodeha	19 (18.8)
Comé	12 (11.9)
Ouedeme Pedah	14 (13.9)
Malawi	56 (55.4)
Mpemba	24 (23.8)
Madziabango	21 (20.8)
Chikwawa	11 (10.9)
Maternal Age	
Mean (SD)	22.5 (5.5)
Gestational Age	
Mean (SD)	38.1(2)
Gravidity, n (%)	
Primigravid	39(38.6)
Secundigravid	29(28.7)
Multigravid	33(32.7)
MOI, mean (range)	
Benin	1.56 (1-4)
Malawi	1.88 (1-7)
Unique var2csa variants, n (%)	152
Benin	57 (37.7)
Malawi	95 (62.9)
Variants shared between samples	17
Variants shared between countries	1
Variants shared between sites	11

Table 5.2: Characteristics of PCR amplified and unamplified samples

Group	Amplified (n=101)	Not Amplified (n = 315)	P values
Maternal Age			
Mean (SD)	22.5 (5.5)	24.5 (6.2)	0.82
Gestational Age			0.83
Mean (SD)	38.1(2)	38.87(1.9)	
Gravidity, n (%)			
Primigravid	39(38.6)	97 (30.8)	0.15
Secundigravid	29(28.7)	89 (28.3)	0.99
Multigravid	31(32.7)	129 (41.0)	0.16

Table 5.3: Nucleotide diversity (π) and Tajima's D

Group	n*	π **	Segregating sites	Tajima's D
All	152	0.105	470	1.781
Malawi	57	0.096	415	1.976
Benin	95	0.113	460	1.699
Primigravid	63	0.099	459	1.183
Secundigravid	49	0.11	451	1.244
Multigravid	53	0.115	458	1.467
*number of haplotypes				
**nucleotide diversity				

Table 5.4: Within-group (alpha) diversity of ID1-DBL2x populations

Group	n*	H_E (SD)	Shannon (SD)	ACE (SD)	ICE (SD)	Chao1 (SD)	Chao2 (SD)
All	152	0.266 (0.027)	4.5 (0.08)	157.12 (21.96)	157.36 (22.08)	149.51 (16.65)	149.31 (16.59)
Malawi	95	0.307 (0.019)	4.04 (0.13)	104.19 (19.32)	104.51 (19.5)	100.21 (15.31)	99.91 (15.19)
Benin	57	0.298 (0.037)	3.98 (0.02)	257.3 (23.87)	259.7 (24.09)	201.81 (64.7)	200.8 (64.16)
Primigravid	63	0.274 (0.027)	3.56 (0.09)	65.06 (16.13)	65.29 (16.33)	61.84 (12.2)	61.61 (12.09)
Secundigravid	49	0.299 (0.032)	3.34 (0.13)	55.97 (18.49)	56.29 (18.8)	52.72 (12.82)	52.42 (12.65)
Multigravid	53	0.300 (0.035)	3.41 (0.12)	57.91 (15.59)	57.71 (15.85)	57.25 (13.65)	56.94 (13.47)

*number of haplotypes

Figure 5.1: Results from the validation study - expected and observed frequencies of the seven genomic parasite lines in five pools using referent-based and k-mer clustering.

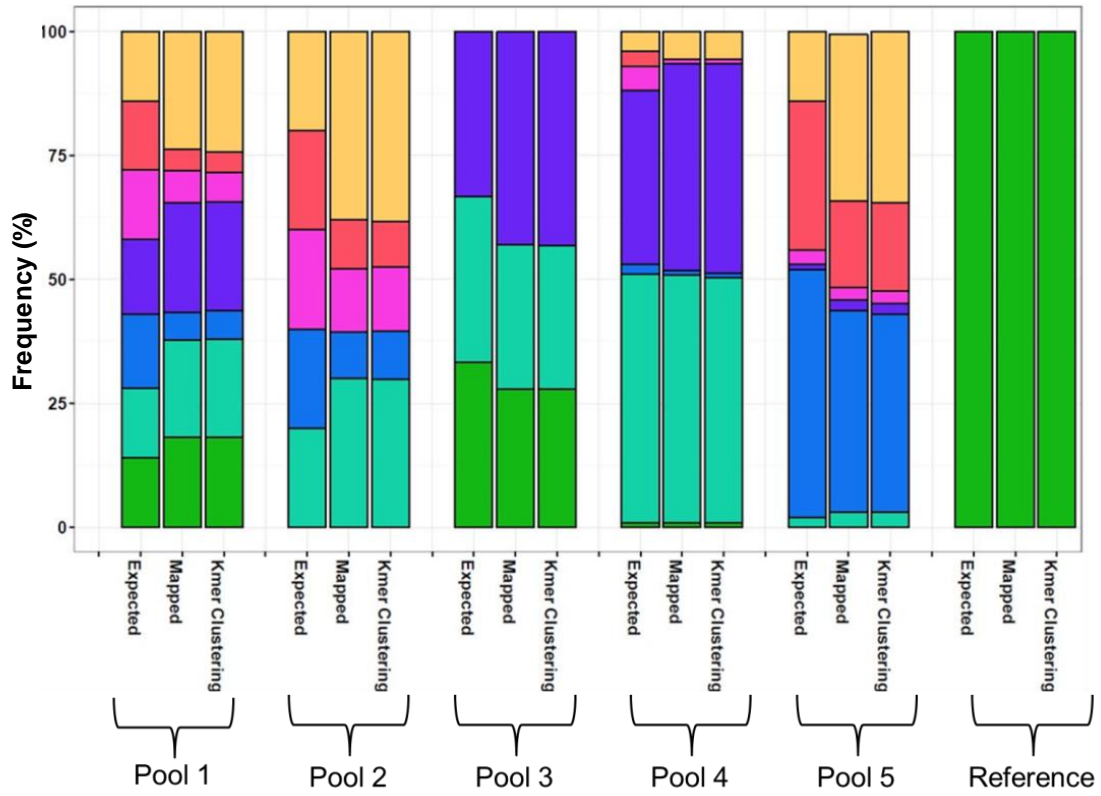


Figure 5.2: Scatterplot of expected and observed frequencies of ID1-ID2x haplotypes in mixtures of reference parasite lines.

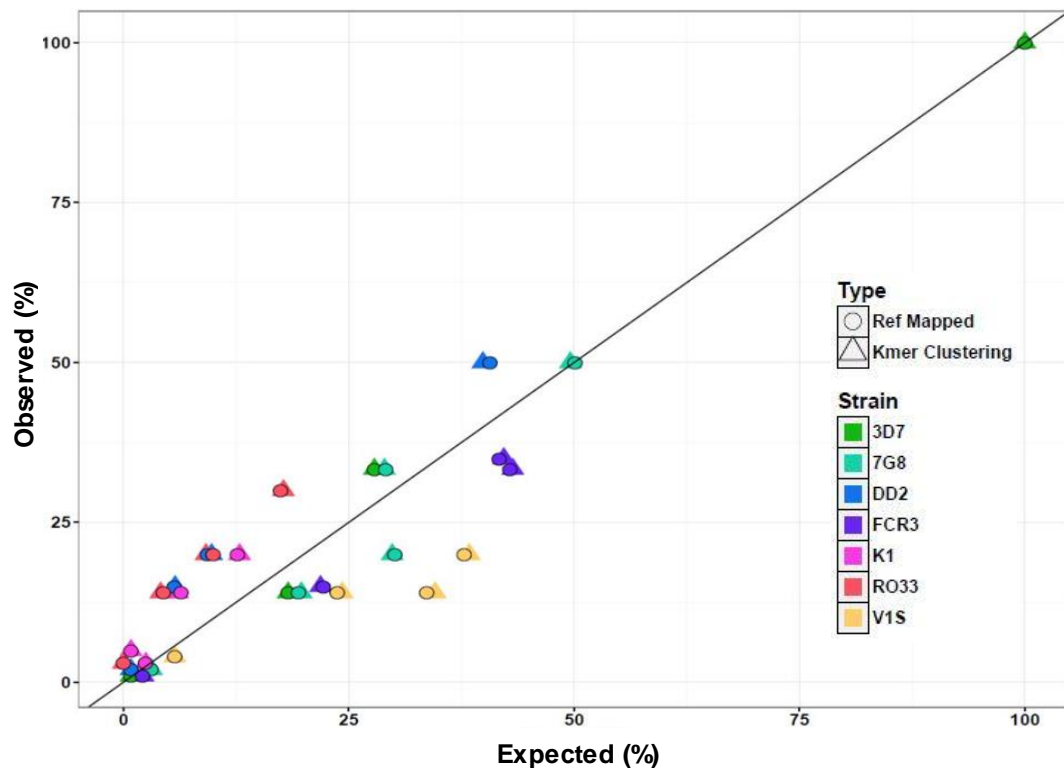


Figure 5.3: Flow chart of the processing and filtering raw reads by barcodes and quality.

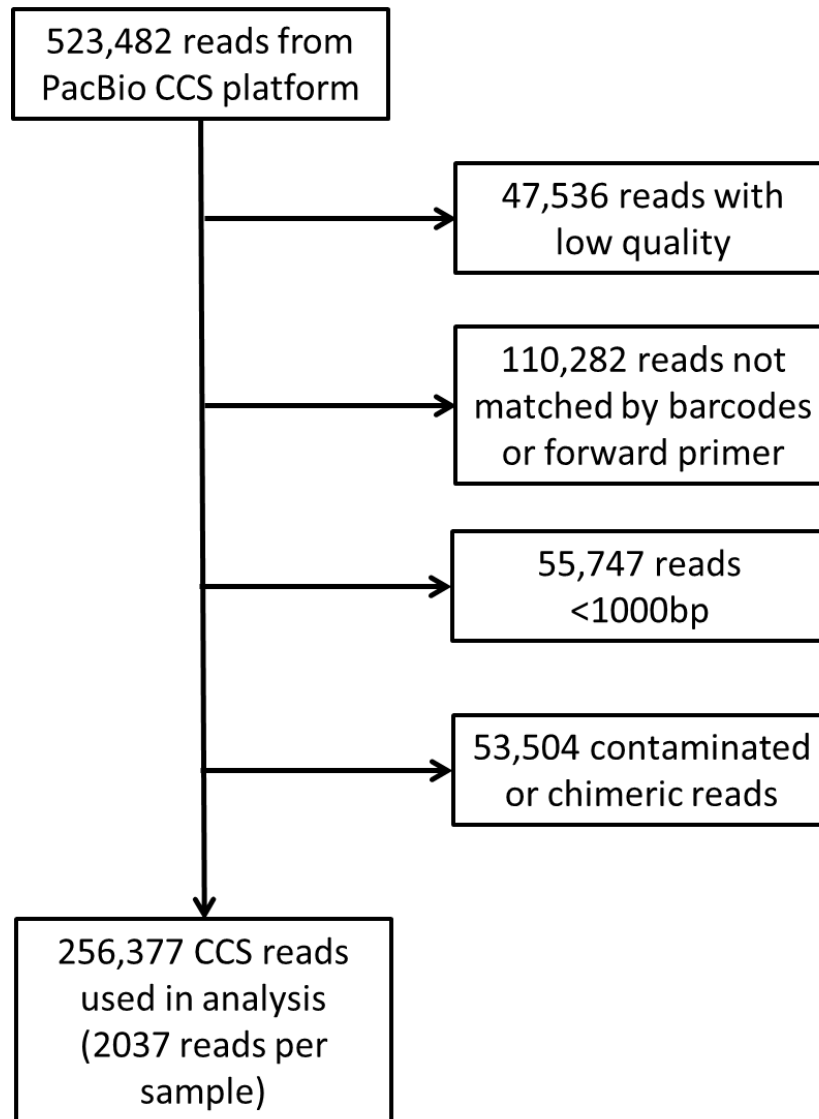


Figure 5.4: Rarefaction curves of ID1-DBL2x variants in Malawi (blue) and Benin (red). Calculated rarefaction curves are represented by solid lines. Extrapolated curves are depicted by dotted lines. The rarefaction curves were bootstrapped 1000 times to generate the 95% confidence intervals (CIs), which are represented by lightly shaded red and blue.

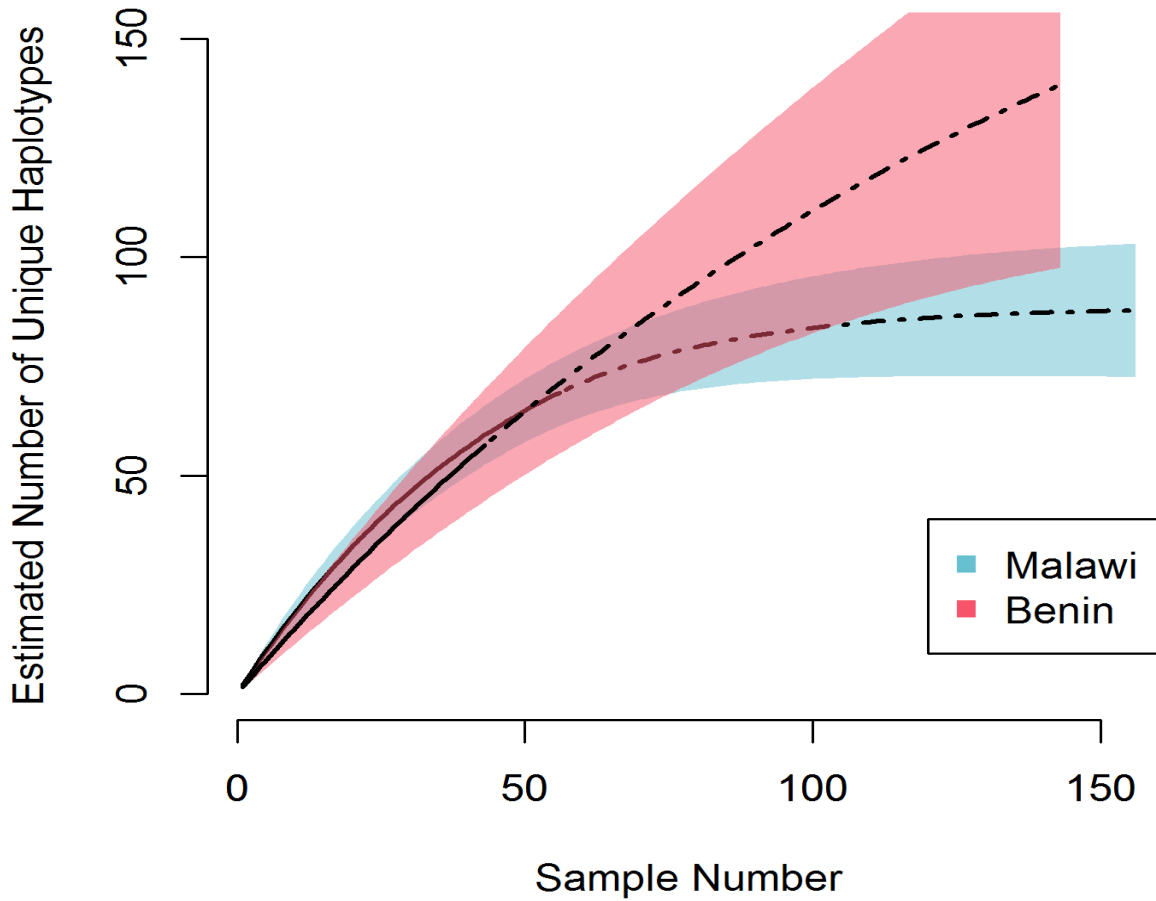


Figure 5.5: Expected heterozygosity (H_e) at each amino acid position along the ID1-DBL2x region among variants populations in Malawi (blue) and Benin (red).

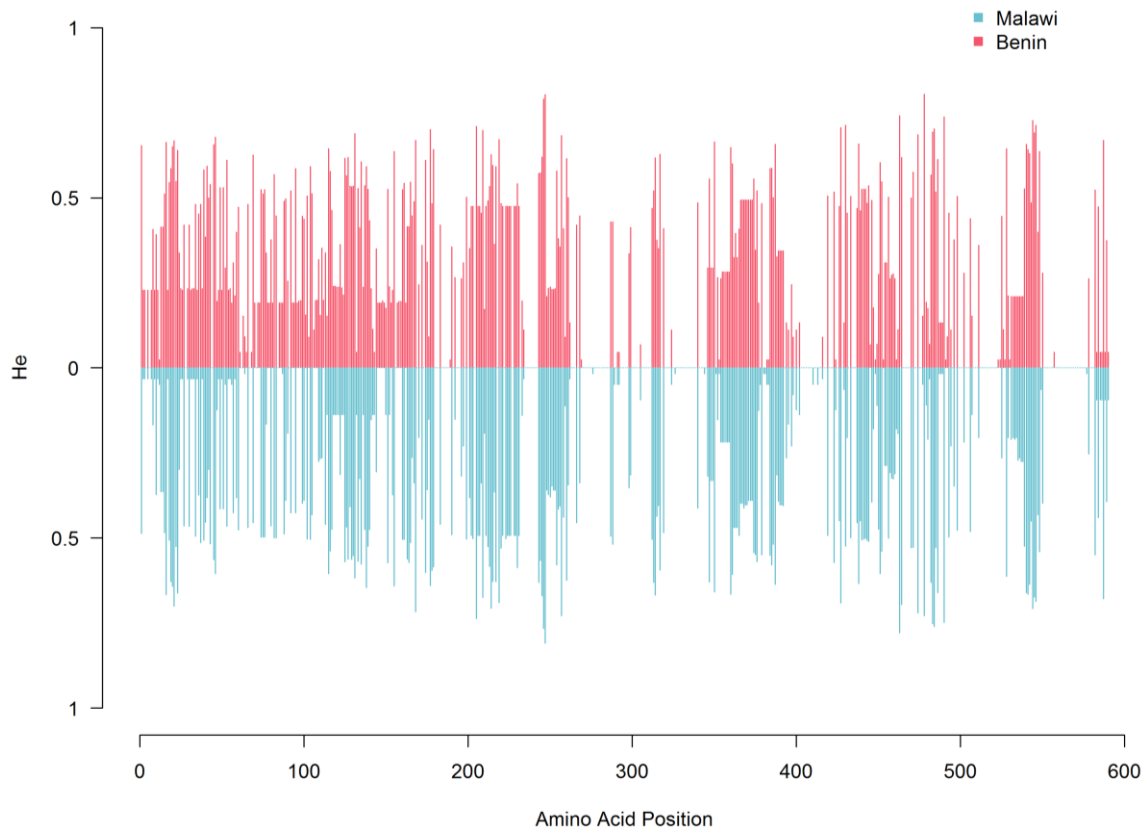
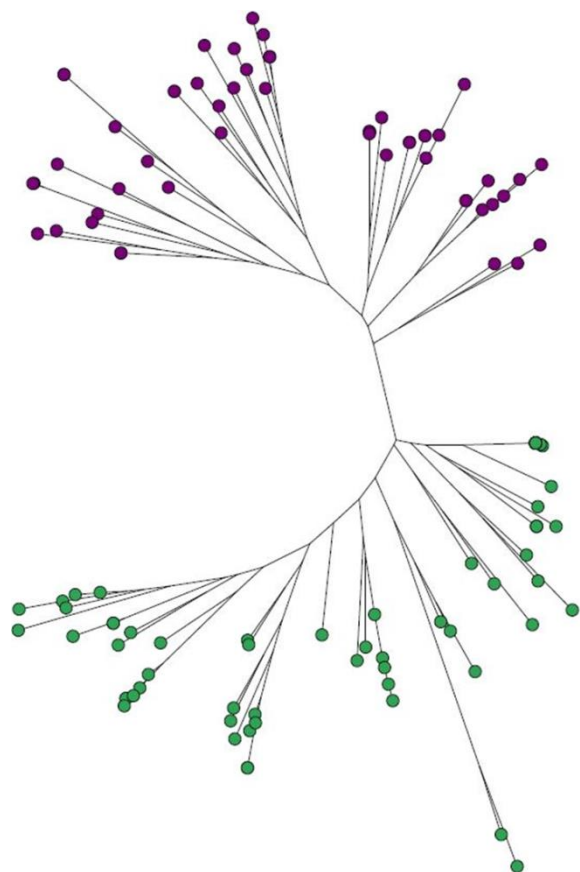


Figure 5.6: Maximum-likelihood phylogenetic trees of ID1-DBL2x variants in (A) Malawi and (B) Benin. All trees were bootstrapped 1000x and bootstrap values >80 were used as cutoff points for significant branch differentiation. In Malawi, variants clustered in two distinct clades with the two referent strains (3D7 and FCR3). In Benin, in addition to the 3D7 and FCR3 clades, three other clades were detected.

A



B

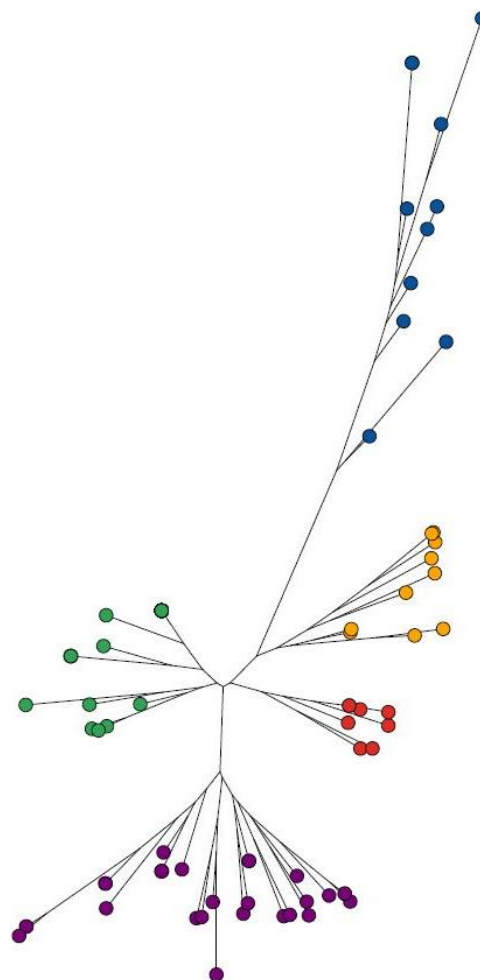


Figure 5.7: Principal Coordinate Analysis (PCoA) of ID1-DBL2x variants in the study population. (A) PCoA plot using principal coordinates 1 and 2 (PC1 & PC2) (B) PCoA plot using principal coordinates 1 and 3 (PC1 & PC3). Each dot represents the unique ID1-DBL2x variants colored according to the clade. The axes indicate which coordinates are being plotted and the percentage of variation explained by that particular axis.

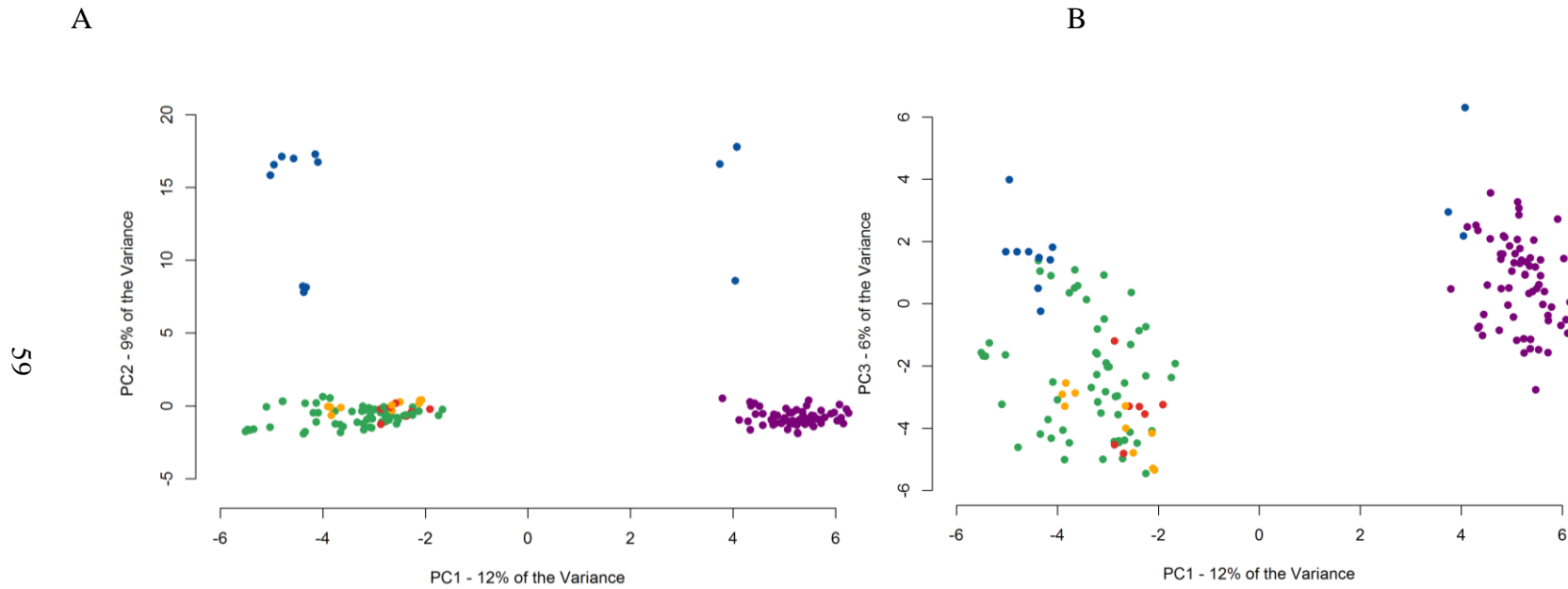


Figure 5.8: F_{ST} values using a sliding window approach (window size = 10bp, step size = 10bp) across the 1.6kb ID1-DBL2x region of *var2csa*. F_{ST} values were calculated between the two major vaccine clades (3D7 and FCR3). F_{ST} values range from 0 to 1, where 0 signifies a genetically identical population and 1 signifies completely differentiated populations. Sliding window F_{ST} analyses show that scanning across the region, the parasite populations in the two clades are very similar except for a ~100 bp region (highlighted in gray) where the two populations differ substantially.

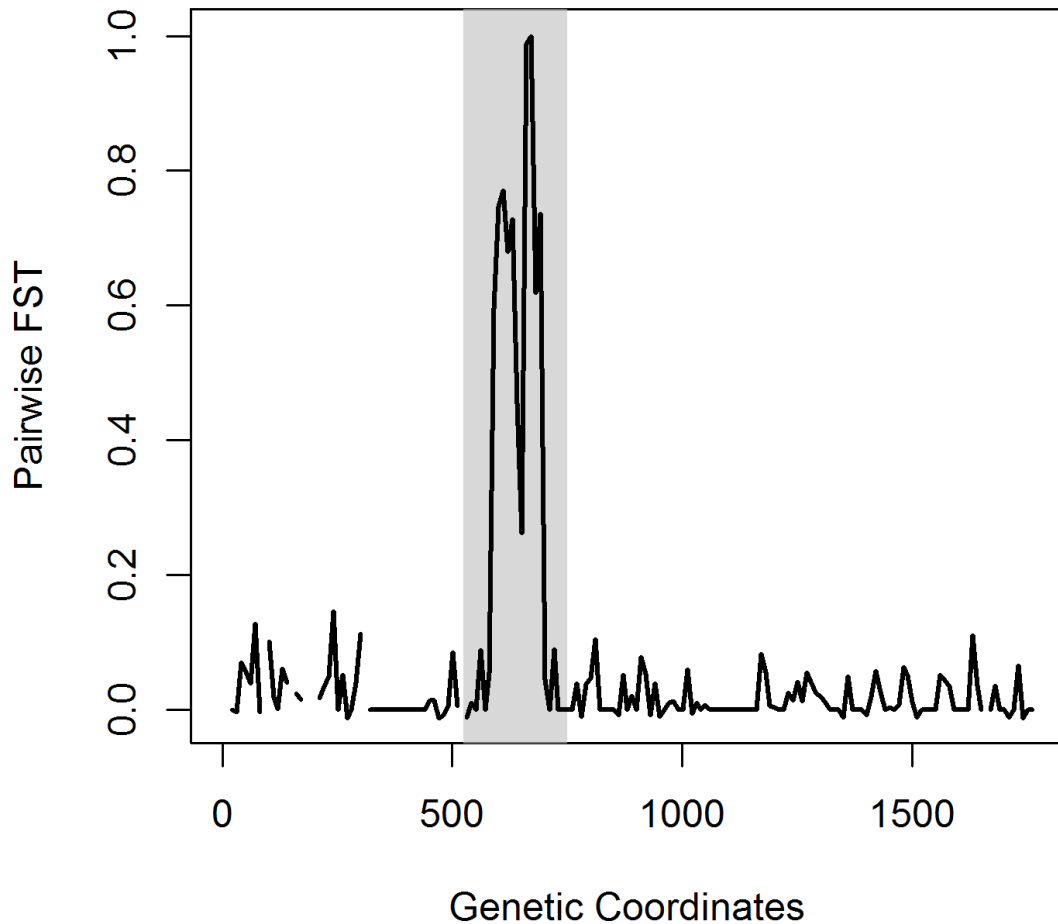
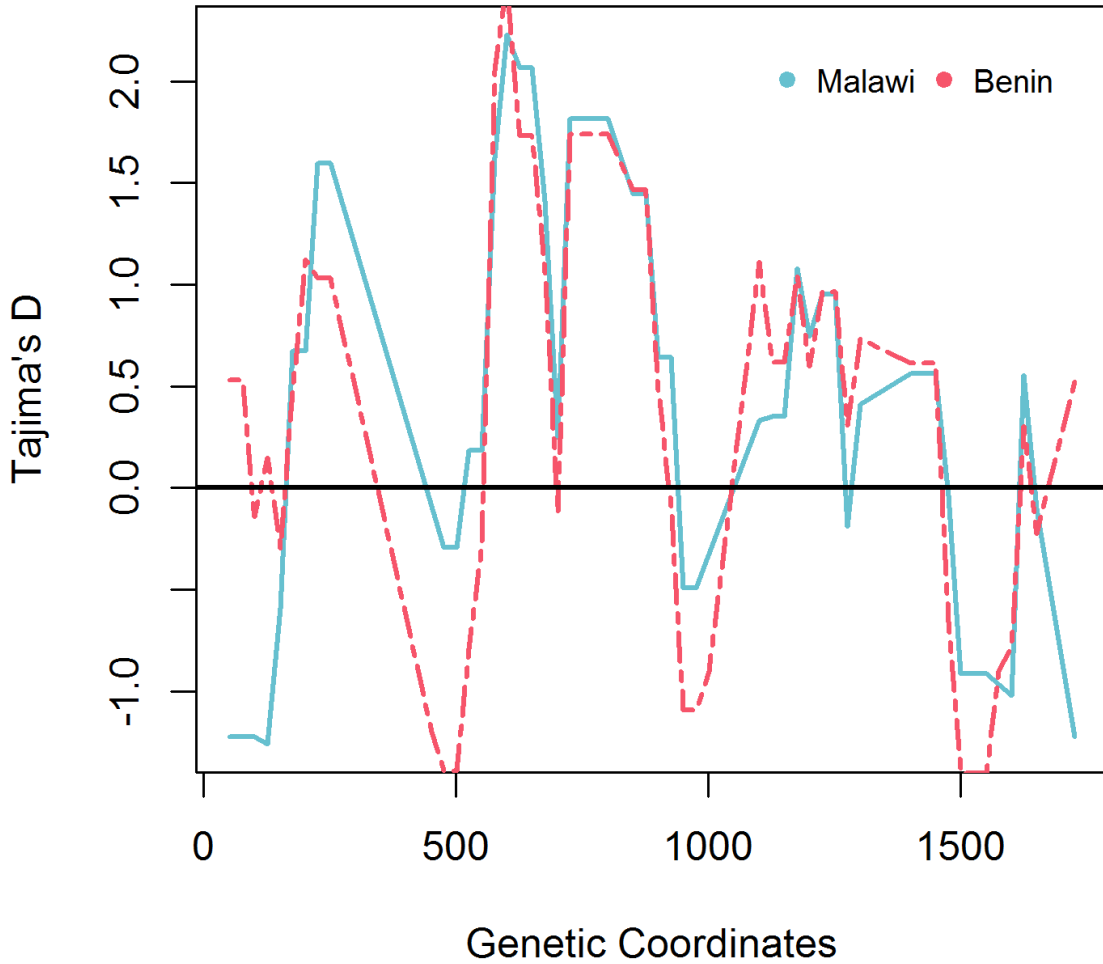
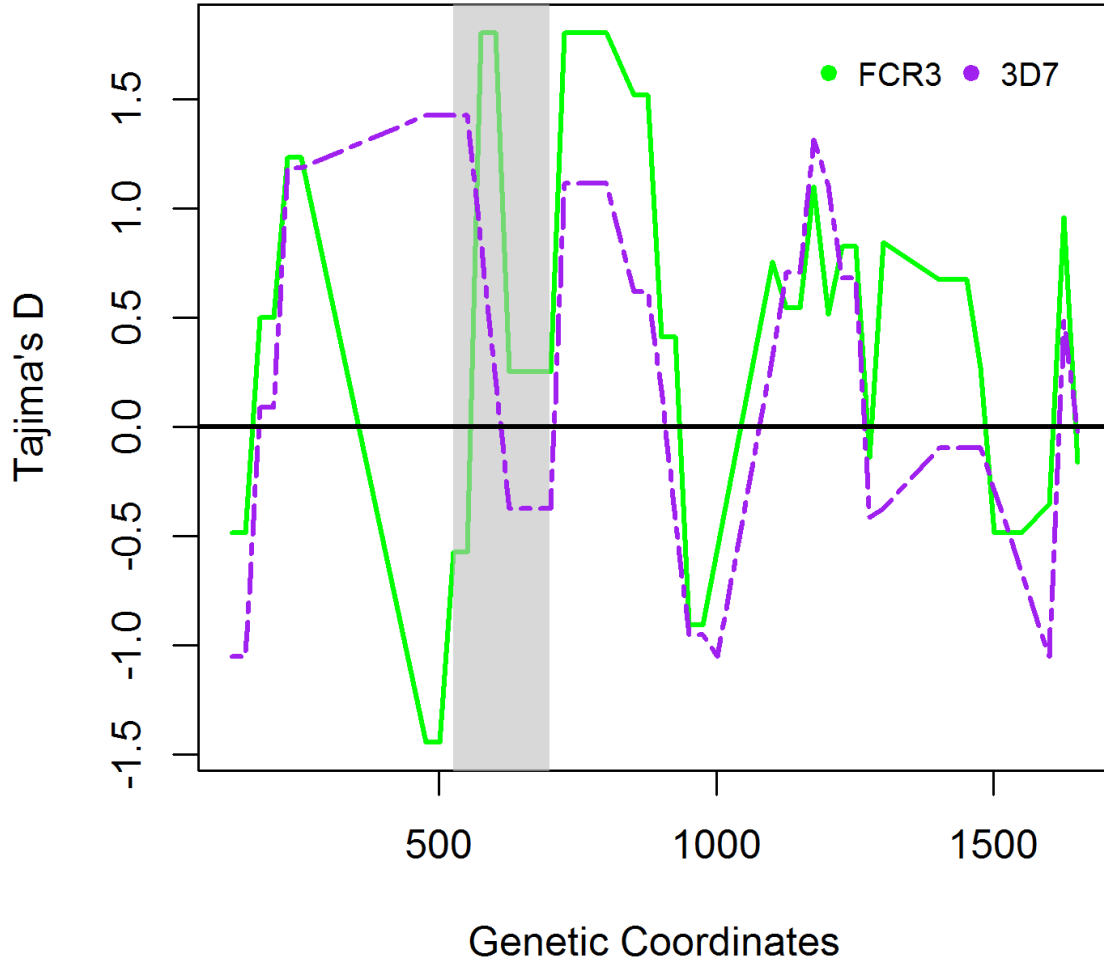


Figure 5.9: Tajima's D using a sliding window approach (window size = 100bp, step size = 25bp) across the 1.6kb ID1-DBL2x region of *var2csa*. (A) Tajima's D calculated for ID1-DBL2x populations in Malawi and Benin and (B) 3D7 and FCR3 clades (dimorphic region highlighted in gray). Tajima D values less than 0 indicate directional selection, equal to 0 indicate genetic drift, and greater than 0 indicate balancing selection.

A)



B)



CHAPTER SIX: RESULTS. IDENTIFICATION OF PATHOGENIC VAR2CSA CLADES: ASSOCIATIONS BETWEEN ID1-DBL2X CLADES AND BIRTH OUTCOMES

STUDY POPULATION

A total of 281 (18.8%) women in Malawi and 175 (27.6%) women in Benin tested positive for *P. falciparum* malaria by PCR in placental blood samples at delivery. Of the women that tested positive, samples from 281 (100%) and 126 (72%) women were available from Malawi and Benin, respectively. We were able to amplify 101 (24.8%) of the available samples – 56 (19.9%) from Malawi and 45 (37.5%) in Benin, constituting our analytic population. Women included in our analytic population were similar to the overall cohort with respect to demographic characteristics (Table 5.2). The variability in amplification rate was likely due to differences in sample storage across study sites as samples extracted from DBS (Malawi) amplified less readily than samples stored as RBC (Benin). Between countries, there were significant differences in maternal age (mean (standard deviation [SD]) 20.5 years (4.6) in Malawi versus 24 years (6.3) in Benin; p value = 0.013) and gestational age at delivery (mean (SD) 37.1 (3.4) weeks in Malawi versus 39.2 weeks (39.2) in Benin; p value <0.001). There were also a significantly higher proportion of primiparous women in Malawi than Benin (51.8% versus 22.2%; p value = 0.004) (Table 6.1).

In our analytic population, the mean (SD) infant birthweight in both countries was similar (Malawi – 2677.7grams [539.6g] versus 2840g [379.7g]; p value = 0.146). The prevalence of LBW was 19.6% (n=11) in Malawi and 13.3% (n=6) in Benin and prevalence of SGA was

16.1% (n=9) in Malawi and 24.4% (n=11) in Benin. There were no significant differences in the adverse birth outcomes, LBW or SGA by country (Table 6.1).

ID1-DBL2X DISTRIBUTION – PHYLOGENETIC ANALYSES

From the 101 pregnant women included in our study, we identified 152 unique ID1-DBL2x variants – 95 in Malawi and 57 in Benin. Of the 152 unique variants, 17 variants were found in multiple women, and one variant was shared between Malawi and Benin. The mean multiplicity of infection (MOI) was 1.88 (range: 1-7) in Malawi and 1.56 (range: 1-4) in Benin. The phylogenetic analysis revealed significant clustering of ID1-DBL2x variants in our study population. The MLE phylogenetic trees in Malawi clustered the variants in two significantly distinct clades (bootstrap values >80%), which also contained the two referent strains which were used to design the current vaccine candidates (3D7 and FCR3) (Figure 6.1 and Figure 6.2). In Malawi, 21 (37.5%) and 22 (39.3%) women were infected with only 3D7-like and FCR3-like variants, respectively. Variants from both clades were detected in 13 (23.2%) women (Table 6.2).

In Benin, the MLE phylogenetic trees identified five distinct clades (bootstrap values >80%). In addition to the 3D7-like and FCR3-like clades identified in Malawi, three clades unique to Benin were detected (Figure 6.1 and Figure 6.2). While 4 (8.9%) and 13 (28.9%) women were infected with only 3D7-like and FCR3-like variants, respectively; 17 women (37.8%) were infected with a variant from one of the three clades unique to Benin. We also observed 6 (13.3%) women infected with both 3D7-like and FCR3-like variants, 4 (8.9%) women infected with FCR3-like variants and variants from the three Benin unique clades, and one (2.2%) woman who was infected with variants from the two vaccine clades as well as the other clades (Table 6.2).

ASSOCIATION BETWEEN ID1-DBL2X CLADES AND BIRTH OUTCOMES

BIRTHWEIGHT

In unadjusted descriptive analyses, the median (IQR) infant birthweight among mothers in Malawi infected with only 3D7-like variants was 2450g (2150g – 2850g) whereas mothers infected with only FCR3-like variants had a median birthweight of 2800g (2600g – 2960g). In Benin, the median (IQR) infant birthweight of mothers infected with only 3D7-like variants was 2710g (2490g – 2825g) and 2720g (2580g – 3250g) for those women infected with only FCR3-like variants (Table 6.2).

In the adjusted analyses, after accounting for parity and country through IPW, women infected with only 3D7-like variants observed a lower infant birthweight (-267.99g; 95% CI: -466.43g – -69.55g) (Table 3). This reduction in infant birthweight was present when stratified by country, but attenuated substantially in Benin, potentially in part due to lower prevalence of 3D7-like variants in this population. In Malawi, the infant birthweight was reduced by 330.97g (95% CI: -628.44g – -33.49) and in Benin, infant birthweight was reduced by 147.88g (95% CI: -530.06 – 234.g) after adjusting for parity (Table 6.3).

In Benin, women infected with variants from clades other than the vaccine clades were found to have somewhat increased infant birthweight (185.63g; 95% CI: -67.80– 439.06g) as compared to women infected with only FCR3-like variants. Similarly, women infected with both 3D7-like and FCR3-like variants reported an increase in infant birthweight across all populations (Table 6.3). However, women infected with variants from both vaccine and non-vaccine clades reported lower infant birthweight across both countries (range: -280g – -133.41g) (Table 6.3).

SMALL-FOR-GESTATIONAL AGE (SGA) AND LOW BIRTHWEIGHT (LBW)

Given the sparse distribution of categorical adverse birth outcomes, SGA and LBW in our study population (Table 6.2), we modeled the effect of only the vaccine clades (3D7-like and FCR3-like) on SGA and LBW. The unadjusted prevalence of LBW was higher in women with only 3D7-like variants than women with only FCR3-like variants in Malawi (47.6% versus 4.6%) and in Benin (25% versus 15.4%) (Table 6.2). Phylogenetically, while there were a few LBW variants clustered in the FCR3-like clade, the majority of the LBW variants in Malawi were present in the 3D7-like clade (Figure 6.1). In Benin, the distribution of LBW variants was more evenly distributed between 3D7-like and FCR3-like clade (Figure 6.1). In pooled analysis, women infected with only 3D7-like variants had higher odds of LBW than women infected with only FCR3-like variants (OR: 8.19; 95% CI: 1.65 – 40.57), after adjusting for parity and country.

Unadjusted prevalence of SGA was higher among women infected with only 3D7-like variants than women infected with only FCR3-like variants in Malawi (28.1% versus 9.1%) and in Benin (50% versus 23.1%) (Table 6.2). Variants present in women delivering SGA infants clustered more readily in 3D7-like clade in Malawi (Figure 6.2) whereas in Benin, the distribution again was more even between 3D7-like and FCR3-like clades (Figure 6.2). We also identified four SGA variants clustered in clade 3 (Figure 6.2). As observed with LBW, in pooled analysis, compared to women infected with only FCR3-like variants, women infected with only 3D7-like variants had higher odds of delivering a SGA infant (OR 2.45; 95% CI 0.67 – 8.94). After adjusting for parity and country, being infected with only 3D7-like variants was still associated with higher odds of delivering a SGA infant (OR 3.65; 95% CI 1.00 – 13.38) (Table 6.4). When stratifying by country, women infected with only 3D7-like variants consistently had

higher odds of delivering a SGA infant in Malawi (OR: 5.21; 95% CI: 0.77 – 35.41) and in Benin (OR: 2.98; 95% CI: 0.27 – 32.48) (Table 6.4.)

DISCUSSION

A VAR2CSA-based vaccine against PAM will be the first syndrome-specific malaria vaccine and may substantially reduce one the major causes of adverse pregnancy and birth outcomes in tropical countries, especially in sub-Saharan Africa. Current vaccine candidates (PlacMalVac and PriMalVac) against PAM are targeting overlapping constructs of the N-terminal region of VAR2CSA. PlacMalVac targets the ID1-ID2 region of the FCR3 variant while PriMalVac targets the DBL1x-DBL2x region of the 3D7 variant (Figure 2.2) [32, 98]. Both vaccine candidates, now in Phase 1 testing in West Africa, are monovalent vaccines designed using two different referent strains. Given that VAR2CSA is highly polymorphic, any effective VAR2CSA-based vaccine aimed at reducing the burden of PAM will have to be polyvalent. Additionally, it is important to identify and include variants that are associated with adverse pregnancy and birth outcomes. Hence, it is critical for vaccine development to take relevant genetic diversity into account to maximize vaccine efficacy. We have previously characterized genetic diversity of ID1-DBL2x from clinical isolates and demonstrated that the ID1-DBL2x region is indeed highly diverse in sub-Saharan Africa. In this study, we identified pathogenic clades of ID1-DBL2x that are associated with adverse birth outcomes among *P. falciparum* infected pregnant women in Malawi and Benin.

The ID1-DBL2x variants in our study population clustered in multiple phylogenetic clades. In Malawi, the variants clustered in two distinct clades, while variants in Benin clustered in five distinct clades. In both countries, we detected clades which contained the vaccine referent strains (3D7 and FCR3) as well as other clades only unique to Benin, however, the 3D7 clade

was much less prevalent in Benin overall, found in only 22% of Benin participants in comparison to 61% of Malawi participants. The LBW and SGA variants clustered more readily in 3D7-like clade in Malawi but were evenly distributed between 3D7-like and FCR3-like clades in Benin. Across multiple birth outcomes, we detected a trend that consistently identified 3D7-like clade as pathogenic in comparison to FCR3-like clade. Compared with variants from FCR3-like clade, variants from 3D7-like clade were associated with lower infant birth weight, LBW, and SGA in our study population. Women infected with variants only from 3D7-like clade saw reductions in infant birth weight ranging from 147.7g to 331.0g as compared to variants from only FCR3-like clade in both Malawi and Benin respectively (Table 6.3). This reduction in birth weight is clinically significant and comparable to the reduction in birth weight observed due to active maternal smoking – approximately a 200 gram deficit [140]. Women infected with variants from only 3D7-like clade also had higher odds of delivering a LBW and SGA infant. This persistent trend along with the detection of 3D7-like clade in multiple countries supports including variants from 3D7-like clade in a polyvalent VAR2CSA-based vaccine.

Even though we used FCR3-like clade as a referent group to estimate relative pathogenicity, we observed that LBW and SGA variants indeed clustered in FCR3-like clade in both Malawi and Benin. Albeit small numbers of LBW and SGA, of interest was the observation that variants from the FCR3-like clade appear to be more pathogenic in Benin than Malawi. While only 4.6% of the variants in FCR3-like clade in Malawi had LBW, 15.4% of the variants in Benin had LBW. Similarly, only 9.1% of the variants in FCR3-like clade in Malawi had SGA compared to 23.1% in Benin. These site-specific differences of the effect of FCR3-like clade on adverse birth outcomes again emphasize the need to develop a polyvalent vaccine which contains variants from the most common clades that are associated with adverse birth outcomes.

This is the first study investigating effects of specific genetic variants of *var2csa* on clinical outcomes. Employing long-read next-generation sequencing of ID1-DBL2x region of *var2csa* in a clinically defined cohort of malaria-infected pregnant women provided an excellent opportunity to look for pathogenic clades of *var2csa* most responsible for adverse birth outcomes. This study complements previous efforts to characterize the role of ID1-DBL2x region of *var2csa* in PAM pathogenesis and associations between immunological responses to placental malaria and adverse birth outcomes. Additionally, using clinical isolates from multiple malaria-endemic countries with differing transmission intensities allowed us to better characterize the extent of ID1-DBL2x genetic diversity and associations with adverse outcomes that were not possible from *in vitro* studies using laboratory isolates [20, 21, 29, 31].

Our study has its limitations. We were limited by the number of *P. falciparum* infected placental samples we were able to include in our analyses. This was in part due to low yield in amplifying ID1-DBL2x, a long hypervariable region of *var2csa* from DNA extracted from dried blood spots in Malawi. We also had sparse distribution of our categorical adverse birth outcomes that precluded us from detecting significant associations. However, including birth weight as a continuous outcome and controlling for confounders using inverse probability weights increased our ability estimate effect of vaccine clades on infant birth weight. Given that gestational age at delivery is a strong predictor of infant birth weight and is itself affected by malaria exposure during pregnancy, gestational age at delivery may be a mediator on the causal pathway of ID1-DBL2x clade and birth weight. We were therefore, limited in our ability to estimate the direct effect of ID1-DBL2x clade on birth weight. We were able to only estimate the total effect of ID1-DBL2x clade on birth weight, including the effect of gestational age on birth weight. This limited our ability to draw inferences on the effect of ID1-DBL2x clades on fetal growth

specifically. Nonetheless, our study approach integrating next-generation sequencing with population genetics and epidemiological modeling provided high resolution data on placental plasmodium parasites and identify trends of pathogenicity.

Recently, RTS,S a general vaccine for malaria completed phase III testing and was recommended by WHO for pilot implementation in 3-5 sub-Saharan African countries [141]. However, it is still not an ideal vaccine as efficacy was suboptimal (25.9-36.3%) [92] and effectiveness will likely be lower. The suboptimal efficacy could partially be explained by parasite antigen diversity that was not taken into account by the vaccine [95, 96]. Therefore, learning from past vaccine efforts, relevant genetic diversity must be taken into account in the early stages of development of a VAR2CSA-vaccine against PAM. The results from our study further provide support for a polyvalent VAR2CSA-based vaccine against PAM as adverse birth outcomes were detected in multiple clades including two clades which contained the referent strains used to design the current vaccine candidates. An effective VAR2CSA vaccine will add to the arsenal of IPTp and ITN to fight against PAM.

Table 6.1: Study population characteristics

	Malawi (n=56)	Benin (n=45)	p value
Maternal Age (mean, (SD))	20.5 (4.6)	24.0 (6.3)	0.013
Gestational Age (mean, (SD))	37.1 (3.4)	39.2 (1.6)	<0.001
Parity (n, (%))			
Primiparous	29 (51.8)	10 (22.2)	0.004
Multiparous	27 (48.2)	35 (77.8)	
Birthweight (mean, (SD))	2677.7 (539.6)	2840 (379.7)	0.146
Low birthweight (LBW), n(%)	11 (19.6)	6 (13.3)	0.459
Small-for-Gestational-Age (SGA), n(%)	9 (16.1)	11 (24.4)	0.535

Table 6.2: Distribution of infant birth weight and adverse birth outcomes by ID1-DBL2x clades in Malawi and Benin

Clades	Malawi (n=56)				Benin (n=45)			
	N	SGA (%)	LBW (%)	Birthweight (median (IQR))	N	SGA (%)	LBW (%)	Birthweight (median (IQR))
3D7	21	6 (28.1)	10 (47.6)	2450 (2150 - 2850)	4	2 (50.0)	1 (25.0)	2710 (2490 - 2825)
FCR3	22	2 (9.1)	1 (4.6)	2800 (2600 - 2960)	13	3 (23.1)	2 (15.4)	2720 (2580 - 3250)
3D7+FCR3	13	1 (7.7)		2850 (2700 - 3000)	6	2 (33.3)	1 (16.7)	2900 (2550 - 3200)
Other*					17	3 (17.6)	1 (5.9)	3100 (2790 - 3160)
FCR3+Other*					4		1 (25.0)	2550 (2365 - 2610)
3D7+FCR3+Other*					1	1 (100)		2700

*Other clade includes variants from clades 3, 4, and 5

Table 6.3: Associations between ID1-DBL2x clades and infant birthweight (grams)

Birthweight	Crude			Adjusted*		
	Effect**	95% CI		Effect**	95% CI	
Pooled						
FCR3	REF	-	-	REF	-	-
3D7	-251.49	-474.28	-28.69	-267.99	-466.43	-69.55
3D7+FCR3	99.55	-151.45	350.56	22.36	-199.23	243.94
Malawi						
FCR3	REF	-	-	REF	-	-
3D7	-262.24	-557.46	33.00	-330.97	-628.44	-33.49
3D7+FCR3	151.76	-204.90	508.43	100.75	-260.62	462.12
Benin						
FCR3	REF	-	-	REF	-	-
3D7	-153.41	-548.08	241.26	-147.66	-530.06	234.75
3D7+FCR3	14.10	-328.97	357.15	21.99	-310.39	354.39
Other†	187.33	-74.23	448.89	185.63	-67.80	439.06
FCR3 + other†	-323.41	-718.08	71.26	-280.21	-662.62	102.19
3D7+FCR3+other†	-110.91	-816.91	595.09	-133.24	-906.49	640.02

*Adjusted for country (pooled analysis only) and parity using inverse probability weights (IPW)

**Change in infant birth weight (grams)

† Other clade includes variants from clades 3, 4, and 5

Table 6.4: Associations between ID1-DBL2x vaccine clades (3D7 & FCR3) and small-for-gestational age (SGA)

	Crude				Adjusted**		
	OR*	95% CI			OR*	95% CI	
Pooled	2.45	0.67	8.94		3.65	1.00	13.38
Malawi	3.46	0.59	20.20		5.21	0.77	35.41
Benin	2.67	0.25	28.44		2.98	0.27	32.48

*Reference group = FCR3 clade

** Adjusted for country (pooled analysis only) and parity using inverse probability weights (IPW)

Figure 6.1: Maximum-likelihood phylogenetic trees of ID1-DBL2x variants in (A) Malawi and (B) Benin by low birth weight (LBW). All trees were bootstrapped 1000x and bootstrap values >80 were used as cutoff points for significant branch differentiation. LBW was defined according to World Health Organization's criteria as infant with a birth weight of less than 2500 grams. 3D7 (green) and FCR3 (purple) were detected in both countries. Majority of the LBW variants (red) in Malawi were present in the 3D7-like clade whereas in Benin, the LBW variants were more evenly distributed between 3D7-like and FCR3-like clades.

A

B

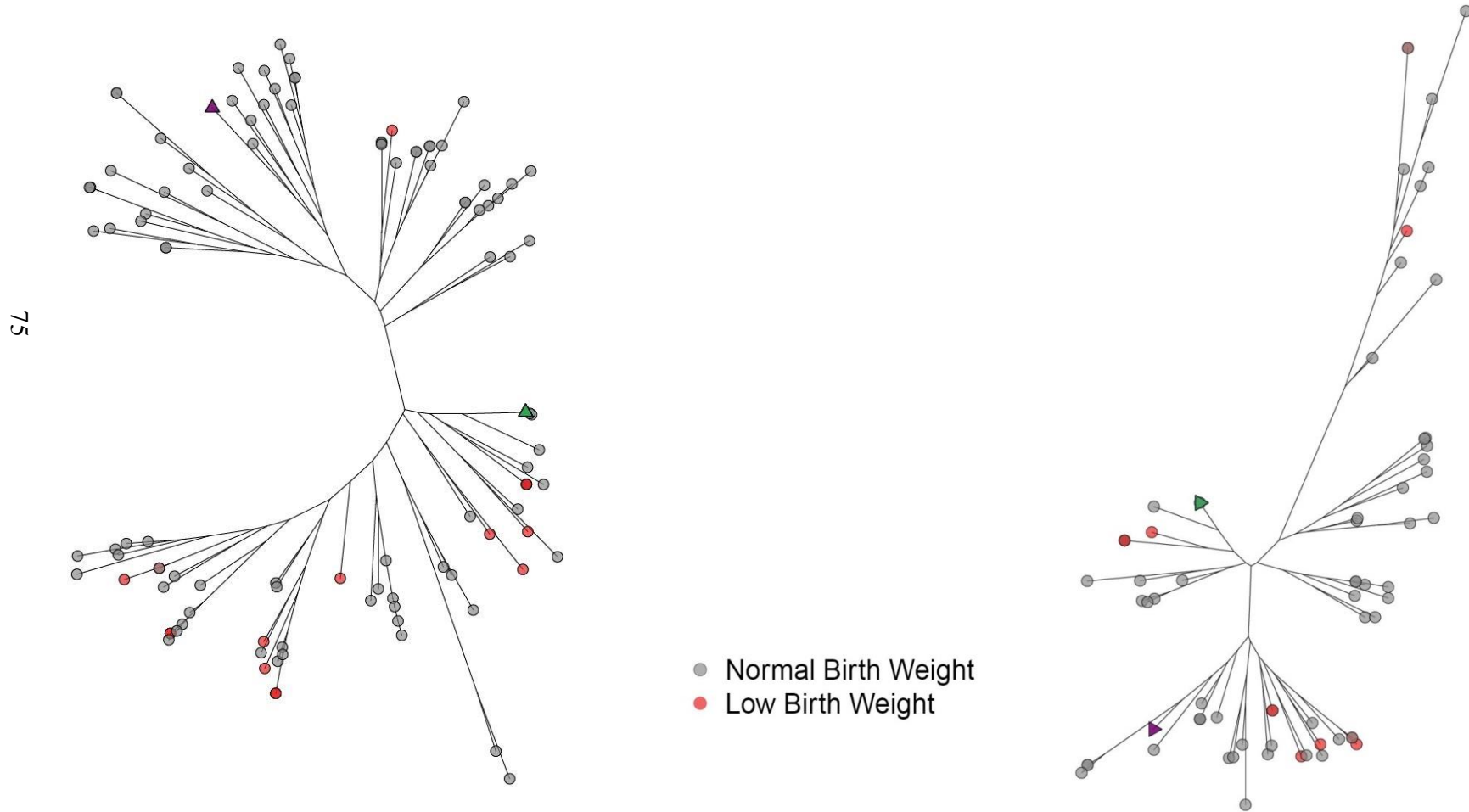
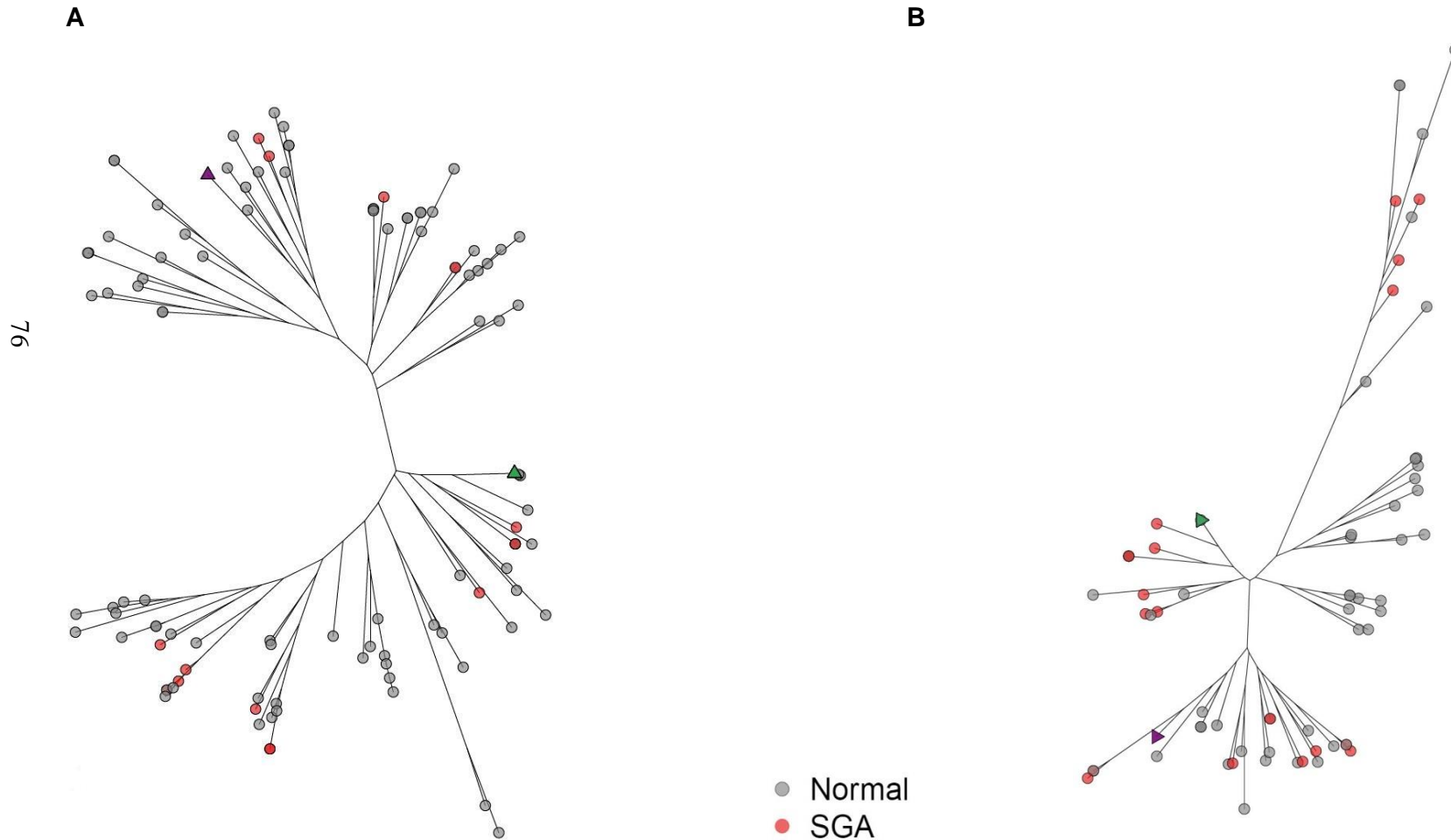


Figure 6.2: Maximum-likelihood phylogenetic trees of ID1-DBL2x variants in (A) Malawi and (B) Benin by small-for-gestational age (SGA). All trees were bootstrapped 1000x and bootstrap values >80 were used as cutoff points for significant branch differentiation. SGA was defined as infant with a birth weight below the 10th percentile for babies of the same gestational age at delivery. 3D7 (green) and FCR3 (purple) were detected in both countries. SGA variants (red) clustered more readily in 3D7-like clade in Malawi whereas in Benin, the distribution was more even between 3D7-like and FCR3-like clades. We also identified four SGA variants clustered in clade 3.



CHAPTER SEVEN: DISCUSSION

Pregnancy-associated malaria (PAM) is an important preventable cause of several pregnancy and birth outcomes including maternal and infant death, especially in sub-Saharan Africa [3, 5, 11-14, 117]. The role of VAR2CSA is critical in PAM pathogenesis and immune response [15-21]. VAR2CSA is a large polymorphic protein coded by one of the *var* genes, which are known to be highly diverse [34, 65, 78, 138]. While still in its infancy, efforts are underway to develop the first syndrome-specific malaria vaccine that can be administered to women of child-bearing age to reduce PAM-associated morbidity and mortality. Currently, two VAR2CSA-based vaccine candidates (PlacMaIVac and PrimMaIVac) targeting overlapping constructs of the N-terminal region are in phase I trials in Benin and Burkina Faso [32, 98].

Developing vaccines against malaria in the past have achieved limited success as only one candidate so far has completed phase III testing. The RTS,S/AS01 vaccine, recently approved by the European Medicines Agency (EMA) and supported by the WHO for pilot testing has shown suboptimal protective efficacy (vaccine efficacy = 25.9% - 36.3%) [92]. Why even after more than three decades of research and testing did RTS,S/AS01 provide only moderate protection? Part of the suboptimal efficacy can be explained by lack of parasite antigen diversity accounted for during the vaccine development stage. Like all other vaccines against *P. falciparum* malaria, RTS,S/AS01 too was designed using genetic sequences from a single, well-characterized reference strain – 3D7 [96]. As a result, strain-specific immunity was the culprit

and RTS,S/AS01 vaccine efficacy was higher against clinical malaria with infections with 3D7 strains than against clinical malaria from non-vaccine strains [95].

So what can we learn from past malaria vaccine efforts to help us develop an effective vaccine against PAM? The current VAR2CSA-based vaccine candidates are also designed using genetic sequences from a single reference parasite line - 3D7 or FCR3 [33]. Given the high diversity of the plasmodium parasite, the question remains whether the current approach for aVAR2CSA-based vaccine is adequate or do we need to fully assess the extent of genetic diversity of the parasite antigen to prevent vaccine escape due to strain-specific efficacy. The genetic diversity of the VAR2CSA-based vaccine candidates is not yet well characterized among clinical isolates. How diverse is the N-terminal region of *var2csa* and are there patterns of genetic clustering that we can leverage in designing a vaccine? Additionally, including specific variants from clades of *var2csa* that are most responsible for the adverse pregnancy and birth outcomes will boost vaccine efficacy and reduce PAM related morbidity and mortality, however, no pathogenic clades have been identified yet. Hence, the driving motivations of this dissertation were to characterize the extent of genetic diversity of the ID-DBL2x region of *var2csa*, one of the two VAR2CSA-based vaccine candidates and identify pathogenic variants or clades to directly inform vaccine development efforts.

SUMMARY OF FINDINGS

The studies completed as part of this dissertation report several findings that may help VAR2CSA-based vaccine development against PAM. The integrative study approach taken here provides a proof-of-concept to investigate other vaccine candidates and inform future malaria vaccine development efforts. In our first aim, we used long-read next-generation sequencing of clinical isolates collected from *P. falciparum*-infected pregnant women in Malawi and Benin to

characterize the genetic diversity of the 1.6kb fragment of ID1-DBL2x region of *var2csa*. We demonstrated that ID1-DBL2x is highly diverse in both Malawi and Benin as we detected immense species richness and within group diversity in both countries among women with differing gravidities. We found that the entire 1.6 kb region is primarily under balancing selection, confirming the role of ID1-DBL2x as an important epitope. Most importantly, our phylogenetic analyses showed that the ID1-DBL2x variants clustered significantly into multiple distinct clades in Malawi and Benin. We detected clades that contained the vaccine reference strains (3D7 and FCR3) in both Malawi and Benin. In addition to the vaccine clades, we found three clades unique to Benin. We also reported a ~100bp dimorphic region in the DBL2x domain which was present in both countries and has been reported previously in West Africa. Overall, the results from aim 1 provide strong evidence that will likely warrant a polyvalent vaccine that includes variants from multiple clades.

In our second aim, we identified pathogenic clades of ID1-DBL2x in our study population by estimating the effect of specific ID1-DBL2x clades on adverse birth outcomes. Across multiple birth outcomes (birth weight, LBW, and SGA), we detected a trend that identified variants from the 3D7-like clade as relatively pathogenic. Phylogenetically, the LBW and SGA variants clustered more readily in the 3D7-like clade in Malawi but were more evenly distributed in Benin among the two vaccine clades. In our study population, compared to variants from the FCR3-like clade, variants from 3D7-like clade were associated with LBW, SGA, and lower infant birth weight.

Even though we used FCR3-like clade as a reference to estimate relative pathogenicity, we did detect LBW and SGA variants clustering in FCR3-like clade. We also observed some site-specific differences of the effect of FCR3-like clade as variants from the FCR3-like clade in

Benin had higher prevalence of LBW and SGA than variants from the FCR3-like clade in Malawi. The results from aim 2 helped identify potential variants and clades that are most responsible for adverse birth outcomes and should be included in a VAR2CSA-based vaccine.

Overall, results from both aims taken together provide strong support for developing a polyvalent vaccine which includes variants from multiple genetically distinct clades as well as variants from clades associated with adverse birth outcomes to maximize the vaccine's efficacy.

CONCLUSIONS

The current preventive interventions, IPTp-SP and ITNs are only effective at high coverage levels and there has been a consistent low uptake of these interventions over the past few decades [5, 51, 53, 55, 80, 81]. Spread of pyrethroid resistance caused by the *kdr* mechanism among Anopheles vectors in sub-Saharan Africa has raised concerns regarding effectiveness of ITNs [88, 89]. In areas of moderate pyrethroid resistance, ITNs have remained effective [90], but significant loss of insecticidal effect of ITNs has been reported in areas of high pyrethroid resistance where *kdr* resistance levels are ubiquitous [91]. Subpar efficacy of ITNs can severely undermine prevention of PAM using bed nets.

Parasite resistance to SP conferred by acquiring successive multiple mutations in *Pfdhps* and *Pfdhfr* genes also threaten the effectiveness of IPTp-SP [53, 82]. The risk of SP failure is higher and rates of parasite clearance are lower among pregnant women living in malaria-endemic areas where SP resistance is widespread [84]. Despite high levels of SP resistance in several areas in sub-Saharan Africa, use of IPTp-SP remains associated with improved pregnancy and birth outcomes [84, 85]. However, the question is how long will this last before IPTp-SP becomes ineffective at preventing adverse outcomes and necessitate implantation of alternate strategies.

To that end, an effective vaccine against PAM can help substantially reduce one of the major causes of adverse pregnancy and birth outcomes in tropical countries, especially in sub-Saharan Africa. The overall goal of this dissertation was to directly inform efforts to develop a VAR2CSA-based vaccine against PAM. We showed that the current monovalent vaccine candidates designed using two different reference strains will likely not confer complete protection against PAM-related adverse outcomes, specifically the vaccine candidate that targets the ID-DBL2x region of *var2csa*. Indeed, the ID1-DBL2x region is highly diverse, variants cluster in multiple distinct clades, and some clades may be more pathogenic than others. Vaccine efforts will have to take this relevant genetic diversity into account in order to develop an efficacious vaccine. A polyvalent vaccine constituting variants from multiple pathogenic clades will be more effective than the current monovalent vaccine candidates.

Our study was unique in that we used an integrative approach to achieve our aims. We leveraged advanced molecular methods (next-generation sequencing), population genetics, and molecular epidemiology to examine antigenic diversity of *var2csa*, identify immunologically relevant selection pressures and signatures of genetic relatedness to inform vaccine development efforts. This approach also allowed us to identify potential pathogenic clades which must be included in an efficacious polyvalent vaccine. Furthermore, using this integrative approach on field isolates collected from clinically well-defined cohorts of malaria-infected pregnant women from multiple malaria-endemic countries with differing transmission intensities allowed us to better characterize the extent of ID1-DBL2x genetic diversity and associations with adverse outcomes that were not possible from *in vitro* studies using laboratory isolates.

In order to better understand the relationship between *var2csa* clade and pathogenicity, several questions and issues need to be addressed. The dimorphic region in the DBL2x domain

that separated the variants into the two vaccine clades should be investigated closely to define specific residues that are associated with adverse birth outcomes, especially at the amino acid level. These residues can be used in future studies to assess pathogenicity of specific *var2csa* clades around the world. Additionally, including growth percentile derived from nomograms as an outcome will allow for estimating the direct effect of specific clades on fetal growth. Measurement of immunity to *var2csa* will also help clarify the associations between *var2csa* variants and adverse birth outcomes. Thoroughly investigating the relationship between gravidity, a measure of past exposure and genetic diversity of *var2csa* in different transmission settings will help elucidate the role of acquired immunity on preventing the adverse effects of malaria infections during pregnancy.

FUTURE DIRECTIONS

Efforts must be taken to characterize the diversity of the ID1-DBL2x region on field isolates from other countries in Africa and around the world. The genetic and amino acid residues identified in the dimorphic region can serve as markers to investigate the diversity and pathogenicity of *var2csa* clades around the world. We also need to characterize the other vaccine candidate that targets the DBL1x-DBL2x region in the same manner we characterized the ID1-DBL2x region in this project. Extensive investigations into the N-terminal region (NTS-DBL2x) of *var2csa* should also be a priority in order to identify other target regions that can help improve upon current vaccine candidates.

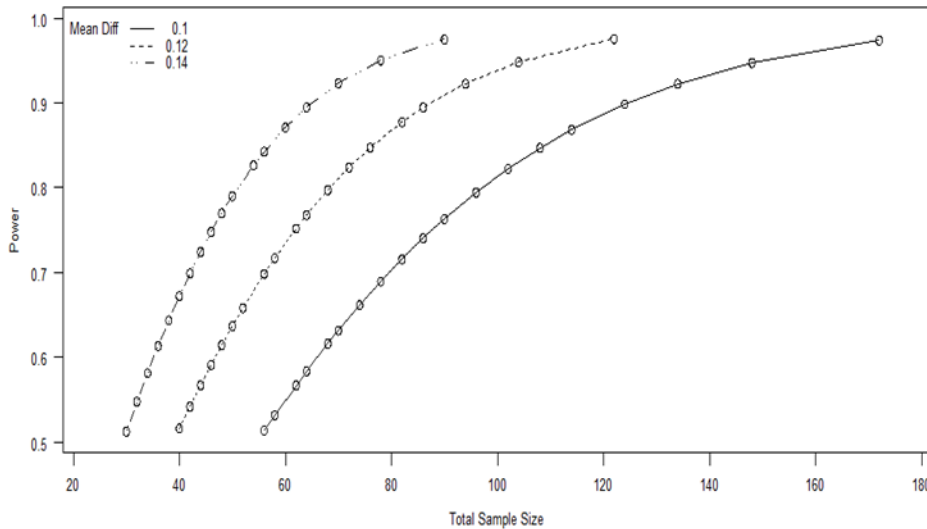
As clinical trials get underway for testing vaccines against PAM, characterizing genetic diversity and genetic surveillance of parasite populations will help monitor changes that affect vaccine efficacy and inform future vaccine improvement and development efforts. The

multidisciplinary approach we used here can also be used to inform the development of future malaria vaccine candidates targeting polymorphic antigens.

APPENDIX A: POWER CALCULATION

Aim1: Assuming mean expected heterozygosity of the ID1-DBL2x-ID2 region of 0.3 with a standard deviation of 0.15 [47], a sample size of 100 pregnant women will yield 81.4% power to detect a 10% change in mean expected heterozygosity between primi-, secundi- and multigravid women using a Bonferroni corrected alpha of 0.0167 for multiple comparisons. Additionally, as the diversity of the ID1-DBL2x-ID2 region has not been well characterized, figure 3 shows the change in power to detect a range of differences in mean expected heterozygosity in the study population.

Power curves for Aim 1



The power to detect change in mean expected heterozygosity among women who gave birth to normal birth weight infants and women who gave birth to low birth weight infants will be higher as the alpha will increase to 0.05. The power to detect a 10% change in mean expected heterozygosity in this comparison will be 91%.

Aim 2: Assuming a sample size of 100 pregnant women and a standard deviation of 10%, we will have 86.7% power to detect a 200 gram difference in birth weight using a Bonferroni corrected alpha of 0.005 for multiple comparisons.

APPENDIX C: DEVELOPMENT AND OPTIMIZATION OF ID1-DBL2X GENOTYPING APPROACH

VAR2CSA is coded by a *var* gene in the *P. falciparum* genome. *Var* genes are infamously known to be incredibly large and diverse. The 1.6 kb fragment of the ID1-DBL2x region of *var2csa* has not yet been well characterized and there is limited data on the genetic diversity of the region. Hence, we developed and optimized our own approach to genotype the entire ID1-DBL2x region using long-read next-generation sequencing technology.

We initially started with a 3 kb fragment spanning DBL1x, ID1, DBL2x, and ID2 regions. We aligned publicly available full length and partial genetic sequences of *var2csa* to identify conserved areas along the DBL1x-ID2 region for primer design. We found no regions flanking the DBL1x-ID2 fragment that were 100% conserved. As a result, we designed degenerate primer pairs in the most conserved flanking regions to account for the diversity.

The initial PCR assay for the 3 kb fragment included hi-fidelity *taq* polymerase and one round of amplification with a long elongation step (~3 minutes per cycle). We amplified seven parasite genomic lines (3D7, FCR3, 7G8, DD2, K1, RO33, and V1/S) and sequenced the PCR products using a primer walking approach. We confirmed the sequences from primer walking were indeed DBL1x-ID2 by comparing our results to publicly available sequences on GenBank and PlasmoDB. We then proceeded to test clinical placental samples from the ISTp trial. However, we were unsuccessful at amplifying the 3 kb fragment from genomic DNA extracted from dried blood spots from the ISTp placental samples collected at delivery. Increasing volume of DNA template to 5 μ l helped increase the PCR yield but the improvement was marginal.

We then modified the PCR assay to target a smaller 1.6 kb fragment spanning the ID1-DBL2x region which was recently recognized as the minimal binding epitope and a potential

vaccine candidate. Reducing to the 1.6 kb fragment improved our PCR amplification; however, we still had difficulty in consistently amplifying the fragment from clinical isolates. We then developed a hemi-nested PCR approach and designed an outer reverse primer that was used in the first round PCR. Primary PCR product was used DNA template for second round PCR which included the primers amplifying the 1.6 kb ID1-DBL2x region. Employing a hemi-nested strategy did significantly improve our PCR yields from clinical isolates; however, the success of PCR amplification from a clinical isolate depended on the quality of the extracted DNA. Additionally, when we added barcodes to both primer sequences for the second round PCR, we had to re-optimize the annealing temperature of the PCR assay to ensure consistent successful amplification of the ID1-DBL2x fragment.

We then compared our PCR amplification success rates using genomic DNA extracted from whole blood to the genomic DNA from DBS as template DNA for our hemi-nested PCR. Genomic DNA from whole blood performed significantly better even at lower concentrations of parasite DNA in amplifying the 1.6 kb fragment. Genomic DNA from whole blood also gave higher concentration and cleaner PCR products.

REFERENCES

1. Dellicour S, Tatem AJ, Guerra CA, Snow RW, ter Kuile FO. Quantifying the number of pregnancies at risk of malaria in 2007: a demographic study. *PLoS Med* **2010**; 7:e1000221.
2. Brabin BJ, Romagosa C, Abdelgalil S, et al. The Sick Placenta—The Role of Malaria. *Placenta* **2004**; 25:359-78.
3. Desai M, ter Kuile FO, Nosten F, et al. Epidemiology and burden of malaria in pregnancy. *The Lancet Infectious diseases* **2007**; 7:93-104.
4. Doritchamou J, Bertin G, Moussiliou A, et al. First-trimester *Plasmodium falciparum* infections display a typical "placental" phenotype. *The Journal of infectious diseases* **2012**; 206:1911-9.
5. Eisele TP, Larsen DA, Anglewicz PA, et al. Malaria prevention in pregnancy, birthweight, and neonatal mortality: a meta-analysis of 32 national cross-sectional datasets in Africa. *The Lancet Infectious diseases* **2012**; 12:942-9.
6. Griffin JB, Lokomba V, Landis SH, et al. *Plasmodium falciparum* parasitaemia in the first half of pregnancy, uterine and umbilical artery blood flow, and foetal growth: a longitudinal Doppler ultrasound study. *Malaria journal* **2012**; 11:319.
7. Mayor A, Kumar U, Bardaji A, et al. Improved pregnancy outcomes in women exposed to malaria with high antibody levels against *Plasmodium falciparum*. *The Journal of infectious diseases* **2013**; 207:1664-74.
8. Menendez C, D'Alessandro U, ter Kuile FO. Reducing the burden of malaria in pregnancy by preventive strategies. *The Lancet Infectious diseases* **2007**; 7:126-35.
9. Rijken MJ, McGready R, Boel ME, et al. Malaria in pregnancy in the Asia-Pacific region. *The Lancet Infectious diseases* **2012**; 12:75-88.
10. Rogerson SJ, Hviid L, Duffy PE, Leke RF, Taylor DW. Malaria in pregnancy: pathogenesis and immunity. *The Lancet Infectious diseases* **2007**; 7:105-17.
11. Takem EN, D'Alessandro U. Malaria in pregnancy. *Mediterranean journal of hematology and infectious diseases* **2013**; 5:e2013010.
12. Umbers AJ, Aitken EH, Rogerson SJ. Malaria in pregnancy: small babies, big problem. *Trends in parasitology* **2011**; 27:168-75.
13. Guyatt HL, Snow RW. Impact of malaria during pregnancy on low birth weight in sub-Saharan Africa. *Clinical microbiology reviews* **2004**; 17:760-9, table of contents.

14. Steketee RW, Nahlen BL, Parise ME, Menendez C. The burden of malaria in pregnancy in malaria-endemic areas. *The American journal of tropical medicine and hygiene* **2001**; 64:28-35.
15. Elliott SR, Duffy MF, Byrne TJ, et al. Cross-reactive surface epitopes on chondroitin sulfate A-adherent *Plasmodium falciparum*-infected erythrocytes are associated with transcription of var2csa. *Infection and immunity* **2005**; 73:2848-56.
16. Gamain B, Trimnell AR, Scheidig C, Scherf A, Miller LH, Smith JD. Identification of multiple chondroitin sulfate A (CSA)-binding domains in the var2CSA gene transcribed in CSA-binding parasites. *The Journal of infectious diseases* **2005**; 191:1010-3.
17. Rovira-Vallbona E, Monteiro I, Bardaji A, et al. VAR2CSA signatures of high *Plasmodium falciparum* parasitemia in the placenta. *PloS one* **2013**; 8:e69753.
18. Salanti A, Staaloe T, Lavstsen T, et al. Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. *Molecular microbiology* **2003**; 49:179-91.
19. Sander AF, Salanti A, Lavstsen T, et al. Positive selection of *Plasmodium falciparum* parasites with multiple var2csa-type PfEMP1 genes during the course of infection in pregnant women. *The Journal of infectious diseases* **2011**; 203:1679-85.
20. Brolin KJ, Persson KE, Wahlgren M, Rogerson SJ, Chen Q. Differential recognition of *P. falciparum* VAR2CSA domains by naturally acquired antibodies in pregnant women from a malaria endemic area. *PloS one* **2010**; 5:e9230.
21. Hviid L, Salanti A. VAR2CSA and protective immunity against pregnancy-associated *Plasmodium falciparum* malaria. *Parasitology* **2007**; 134:1871-6.
22. Lambert LH, Bullock JL, Cook ST, et al. Antigen reversal identifies targets of opsonizing IgGs against pregnancy-associated malaria. *Infection and immunity* **2014**; 82:4842-53.
23. Rogerson SJ, Mwapasa V, Meshnick SR. Malaria in pregnancy: linking immunity and pathogenesis to prevention. *The American journal of tropical medicine and hygiene* **2007**; 77:14-22.
24. Dahlback M, Jorgensen LM, Nielsen MA, et al. The chondroitin sulfate A-binding site of the VAR2CSA protein involves multiple N-terminal domains. *The Journal of biological chemistry* **2011**; 286:15908-17.
25. Srivastava A, Gangnard S, Dechavanne S, et al. Var2CSA minimal CSA binding region is located within the N-terminal region. *PloS one* **2011**; 6:e20270.
26. Bigey P, Gnidehou S, Doritchamou J, et al. The NTS-DBL2X region of VAR2CSA induces cross-reactive antibodies that inhibit adhesion of several *Plasmodium falciparum* isolates to chondroitin sulfate A. *The Journal of infectious diseases* **2011**; 204:1125-33.

27. Babakhanyan A, Leke RG, Salanti A, et al. The antibody response of pregnant Cameroonian women to VAR2CSA ID1-ID2a, a small recombinant protein containing the CSA-binding site. *PloS one* **2014**; 9:e88173.
28. Bordbar B, Tuikue Ndam N, Renard E, et al. Genetic diversity of VAR2CSA ID1-DBL2Xb in worldwide *Plasmodium falciparum* populations: impact on vaccine design for placental malaria. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **2014**; 25:81-92.
29. Bordbar B, Tuikue-Ndam N, Bigey P, Doritchamou J, Scherman D, Deloron P. Identification of Id1-DBL2X of VAR2CSA as a key domain inducing highly inhibitory and cross-reactive antibodies. *Vaccine* **2012**; 30:1343-8.
30. Clausen TM, Christoffersen S, Dahlback M, et al. Structural and functional insight into how the *Plasmodium falciparum* VAR2CSA protein mediates binding to chondroitin sulfate A in placental malaria. *The Journal of biological chemistry* **2012**; 287:23332-45.
31. Doritchamou J, Bigey P, Nielsen MA, et al. Differential adhesion-inhibitory patterns of antibodies raised against two major variants of the NTS-DBL2X region of VAR2CSA. *Vaccine* **2013**; 31:4516-22.
32. Placental Malaria Vaccine (PlacMaVac)
- Available at: <http://cmp.ku.dk/collaborations/placmalvac/>. Accessed 9 February 2016.
33. Tuikue-Ndam N, Deloron P. Developing vaccines to prevent malaria in pregnant women. *Expert opinion on biological therapy* **2015**; 15:1173-82.
34. Duffy MF, Caragounis A, Noviyanti R, et al. Transcribed var genes associated with placental malaria in Malawian women. *Infection and immunity* **2006**; 74:4875-83.
35. Talundzic E, Shah S, Fawole O, Owino S, Moore JM, Peterson DS. Sequence polymorphism, segmental recombination and toggling amino acid residues within the DBL3X domain of the VAR2CSA placental malaria antigen. *PloS one* **2012**; 7:e31565.
36. Ataide R, Mayor A, Rogerson SJ. Malaria, primigravidae, and antibodies: knowledge gained and future perspectives. *Trends in parasitology* **2014**; 30:85-94.
37. Teo A, Hasang W, Randall LM, et al. Decreasing malaria prevalence and its potential consequences for immunity in pregnant women. *The Journal of infectious diseases* **2014**; 210:1444-55.
38. Tutterrow YL, Salanti A, Avril M, et al. High avidity antibodies to full-length VAR2CSA correlate with absence of placental malaria. *PloS one* **2012**; 7:e40049.
39. (WHO) WHO. Malaria-Fact Sheet. Available at: <http://www.who.int/mediacentre/factsheets/fs094/en/>. 2014.

40. Centers for Disease Control and Prevention (CDC). Malaria-Disease. Available at: <http://www.cdc.gov/malaria/about/disease.html>. 2014.
41. Centers for Disease Control and Prevention (CDC). Malaria-Biology. Available at: <http://www.cdc.gov/malaria/about/biology/index.html>. 2014.
42. WHO. World Malaria Report, **2015**.
43. Gwatkin DR, Guillot M, Heuveline P. The burden of disease among the global poor. *Lancet* **1999**; 354:586-9.
44. Somi MF, Butler JR, Vahid F, Njau J, Kachur SP, Abdulla S. Is there evidence for dual causation between malaria and socioeconomic status? Findings from rural Tanzania. *The American journal of tropical medicine and hygiene* **2007**; 77:1020-7.
45. Gallup JL, Sachs JD. The economic burden of malaria. *The American journal of tropical medicine and hygiene* **2001**; 64:85-96.
46. Teklehaimanot A, Mejia P. Malaria and poverty. *Annals of the New York Academy of Sciences* **2008**; 1136:32-7.
47. Worrall E, Morel C, Yeung S, et al. The economics of malaria in pregnancy--a review of the evidence and research priorities. *The Lancet Infectious diseases* **2007**; 7:156-68.
48. Orem JN, Kirigia JM, Azairwe R, Kasirye I, Walker O. Impact of malaria morbidity on gross domestic product in Uganda. *International archives of medicine* **2012**; 5:12.
49. Sicuri E, Vieta A, Lindner L, Constenla D, Sauboin C. The economic costs of malaria in children in three sub-Saharan countries: Ghana, Tanzania and Kenya. *Malaria journal* **2013**; 12:307.
50. Huynh BT, Fievet N, Gbaguidi G, et al. Influence of the timing of malaria infection during pregnancy on birth weight and on maternal anemia in Benin. *The American journal of tropical medicine and hygiene* **2011**; 85:214-20.
51. Huynh BT, Fievet N, Briand V, et al. Consequences of gestational malaria on birth weight: finding the best timeframe for intermittent preventive treatment administration. *PloS one* **2012**; 7:e35342.
52. McClure EM, Goldenberg RL, Dent AE, Meshnick SR. A systematic review of the impact of malaria prevention in pregnancy on low birth weight and maternal anemia. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* **2013**; 121:103-9.
53. ter Kuile FO, van Eijk AM, Filler SJ. Effect of sulfadoxine-pyrimethamine resistance on the efficacy of intermittent preventive therapy for malaria control during pregnancy: a systematic review. *Jama* **2007**; 297:2603-16.

54. World Health Organization (WHO). WHO policy brief for the implementation of intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP): WHO, **2014**.
55. Gamble C, Ekwaru PJ, Garner P, ter Kuile FO. Insecticide-treated nets for the prevention of malaria in pregnancy: a systematic review of randomised controlled trials. *PLoS Med* **2007**; 4:e107.
56. Ndam NT, Deloron P. Molecular aspects of *Plasmodium falciparum* Infection during pregnancy. *Journal of biomedicine & biotechnology* **2007**; 2007:43785.
57. Rowe JA, Claessens A, Corrigan RA, Arman M. Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert reviews in molecular medicine* **2009**; 11:e16.
58. Goel S, Gowda DC. How specific is *Plasmodium falciparum* adherence to chondroitin 4-sulfate? *Trends in parasitology* **2011**; 27:375-81.
59. Nunes MC, Scherf A. *Plasmodium falciparum* during pregnancy: a puzzling parasite tissue adhesion tropism. *Parasitology* **2007**; 134:1863-9.
60. Kraemer SM, Smith JD. A family affair: var genes, PfEMP1 binding, and malaria disease. *Current opinion in microbiology* **2006**; 9:374-80.
61. Pasternak ND, Dzikowski R. PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum*. *The international journal of biochemistry & cell biology* **2009**; 41:1463-6.
62. Chen DS, Barry AE, Leliwa-Sytek A, et al. A molecular epidemiological study of var gene diversity to characterize the reservoir of *Plasmodium falciparum* in humans in Africa. *PloS one* **2011**; 6:e16629.
63. Salanti A, Dahlback M, Turner L, et al. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *The Journal of experimental medicine* **2004**; 200:1197-203.
64. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in *Plasmodium falciparum*. *Annual review of microbiology* **2008**; 62:445-70.
65. Taylor HM, Kyes SA, Newbold CI. Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Molecular and biochemical parasitology* **2000**; 110:391-7.
66. Freitas-Junior LH, Bottius E, Pirrit LA, et al. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **2000**; 407:1018-22.
67. Nielsen MA, Pinto VV, Resende M, et al. Induction of adhesion-inhibitory antibodies against placental *Plasmodium falciparum* parasites by using single domains of VAR2CSA. *Infection and immunity* **2009**; 77:2482-7.

68. Obiakor H, Avril M, Macdonald NJ, et al. Identification of VAR2CSA domain-specific inhibitory antibodies of the Plasmodium falciparum erythrocyte membrane protein 1 using a novel flow cytometry assay. *Clinical and vaccine immunology : CVI* **2013**; 20:433-42.
69. Dahlbäck M, Nielsen MA, Salanti A. Can any lessons be learned from the ambiguous glycan binding of PfEMP1 domains? *Trends in Parasitology* **2010**; 26:230-5.
70. Dahlback M, Rask TS, Andersen PH, et al. Epitope mapping and topographic analysis of VAR2CSA DBL3X involved in P. falciparum placental sequestration. *PLoS pathogens* **2006**; 2:e124.
71. Deloron P, Milet J, Badaut C. Plasmodium falciparum variability and immune evasion proceed from antigenicity of consensus sequences from DBL6epsilon; generalization to all DBL from VAR2CSA. *PLoS one* **2013**; 8:e54882.
72. Ditlev SB, Nielsen MA, Resende M, et al. Identification and characterization of B-cell epitopes in the DBL4epsilon domain of VAR2CSA. *PLoS one* **2012**; 7:e43663.
73. Gangnard S, Badaut C, Ramboarina S, et al. Structural and immunological correlations between the variable blocks of the VAR2CSA domain DBL6epsilon from two Plasmodium falciparum parasite lines. *Journal of molecular biology* **2013**; 425:1697-711.
74. Salanti A, Resende M, Ditlev S, et al. Several domains from VAR2CSA can induce Plasmodium falciparum adhesion-blocking antibodies. *Malaria Journal* **2010**; 9:11.
75. Saveria T, Oleinikov AV, Williamson K, et al. Antibodies to Escherichia coli-expressed C-terminal domains of Plasmodium falciparum variant surface antigen 2-chondroitin sulfate A (VAR2CSA) inhibit binding of CSA-adherent parasites to placental tissue. *Infection and immunity* **2013**; 81:1031-9.
76. Singh K, Gitti RK, Diouf A, et al. Subdomain 3 of Plasmodium falciparum VAR2CSA DBL3x is identified as a minimal chondroitin sulfate A-binding region. *The Journal of biological chemistry* **2010**; 285:24855-62.
77. Singh K, Gittis AG, Nguyen P, Gowda DC, Miller LH, Garboczi DN. Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A. *Nature structural & molecular biology* **2008**; 15:932-8.
78. Trimnell AR, Kraemer SM, Mukherjee S, et al. Global genetic diversity and evolution of var genes associated with placental and severe childhood malaria. *Molecular and biochemical parasitology* **2006**; 148:169-80.
79. Tuikue Ndam N, Deloron P. Towards a vaccine against pregnancy-associated malaria. *Parasite (Paris, France)* **2008**; 15:515-21.
80. Hill J, Hoyt J, van Eijk AM, et al. Factors affecting the delivery, access, and use of interventions to prevent malaria in pregnancy in sub-Saharan Africa: a systematic review and meta-analysis. *PLoS Med* **2013**; 10:e1001488.

81. Menendez C, Ferencsik E, Roman E, Bardaji A, Mangiaterra V. Malaria in pregnancy: challenges for control and the need for urgent action. *The Lancet Global health* **2015**; 3:e433-4.
82. Kublin JG, Dzinjalama FK, Kamwendo DD, et al. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of *Plasmodium falciparum* malaria. *The Journal of infectious diseases* **2002**; 185:380-8.
83. Kayentao K, Garner P, van Eijk AM, et al. Intermittent preventive therapy for malaria during pregnancy using 2 vs 3 or more doses of sulfadoxine-pyrimethamine and risk of low birth weight in Africa: systematic review and meta-analysis. *Jama* **2013**; 309:594-604.
84. Desai M, Gutman J, Taylor SM, et al. Impact of Sulfadoxine-Pyrimethamine Resistance on Effectiveness of Intermittent Preventive Therapy for Malaria in Pregnancy at Clearing Infections and Preventing Low Birth Weight. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2016**; 62:323-33.
85. Taylor SM, Antonia AL, Chaluluka E, et al. Antenatal receipt of sulfadoxine-pyrimethamine does not exacerbate pregnancy-associated malaria despite the expansion of drug-resistant *Plasmodium falciparum*: clinical outcomes from the QuEERPAM study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2012**; 55:42-50.
86. Harrington WE, Mutabingwa TK, Muehlenbachs A, et al. Competitive facilitation of drug-resistant *Plasmodium falciparum* malaria parasites in pregnant women who receive preventive treatment. *Proceedings of the National Academy of Sciences of the United States of America* **2009**; 106:9027-32.
87. Minja DT, Schmiegelow C, Mmbando B, et al. *Plasmodium falciparum* mutant haplotype infection during pregnancy associated with reduced birthweight, Tanzania. *Emerging infectious diseases* **2013**; 19.
88. Ranson H, N'Guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends in parasitology* **2011**; 27:91-8.
89. Ranson H, Lissenden N. Insecticide Resistance in African Anopheles Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends in parasitology* **2016**; 32:187-96.
90. Lindblade KA, Mwandama D, Mzilahowa T, et al. A cohort study of the effectiveness of insecticide-treated bed nets to prevent malaria in an area of moderate pyrethroid resistance, Malawi. *Malaria journal* **2015**; 14:31.
91. N'Guessan R, Corbel V, Akogbeto M, Rowland M. Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, Benin. *Emerging infectious diseases* **2007**; 13:199-206.

92. 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.
93. The Rts S/AS01. Efficacy and Safety of the RTS,S/AS01 Malaria Vaccine during 18 Months after Vaccination: A Phase 3 Randomized, Controlled Trial in Children and Young Infants at 11 African Sites. *PLoS Med* **2014**; 11:e1001685.
94. Bowman NM, Congdon S, Mvalo T, et al. Comparative population structure of *Plasmodium falciparum* circumsporozoite protein NANP repeat lengths in Lilongwe, Malawi. *Scientific reports* **2013**; 3:1990.
95. Neafsey DE, Juraska M, Bedford T, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *The New England journal of medicine* **2015**; 373:2025-37.
96. Plowe CV. Vaccine-Resistant Malaria. *The New England journal of medicine* **2015**; 373:2082-3.
97. Takala SL, Coulibaly D, Thera MA, et al. Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. *Science translational medicine* **2009**; 1:2ra5.
98. PRIMALVAC. Available at: <http://www.euvaccine.eu/portfolio/project-index/primalvac>. Accessed 9 February 2016.
99. Bailey JA, Mvalo T, Aragam N, et al. 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* **2012**; 206:580-7.
100. Juliano JJ, Porter K, Mwapasa V, et al. Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proceedings of the National Academy of Sciences of the United States of America* **2010**; 107:20138-43.
101. Kalilani L, Mofolo I, Chaponda M, et al. A randomized controlled pilot trial of azithromycin or artesunate added to sulfadoxine-pyrimethamine as treatment for malaria in pregnant women. *PloS one* **2007**; 2:e1166.
102. Kalilani L, Mofolo I, Chaponda M, Rogerson SJ, Meshnick SR. The effect of timing and frequency of *Plasmodium falciparum* infection during pregnancy on the risk of low birth weight and maternal anemia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **2010**; 104:416-22.
103. Kelly-Hope LA, McKenzie FE. The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across sub-Saharan Africa. *Malaria journal* **2009**; 8:19.

104. Cottrell G, Moussiliou A, Luty AJ, et al. Submicroscopic *Plasmodium falciparum* infections are associated with maternal anemia, premature births and low birthweight. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2015**.
105. Fernandes S, Sicuri E, Kayentao K, et al. Cost-effectiveness of two versus three or more doses of intermittent preventive treatment for malaria during pregnancy in sub-Saharan Africa: a modelling study of meta-analysis and cost data. *The Lancet Global health* **2015**; 3:e143-53.
106. Gutman J, Kalilani L, Taylor S, et al. The A581G Mutation in the Gene Encoding *Plasmodium falciparum* Dihydropteroate Synthetase Reduces the Effectiveness of Sulfadoxine-Pyrimethamine Preventive Therapy in Malawian Pregnant Women. *The Journal of infectious diseases* **2015**.
107. Modeste Gouissi F, Salifou S, Patrick Etorh A, et al. Assessment of Long-Lasting Insecticidal Nets (LLINs) on Vectors and Malaria Transmission in the Commune of Aguegues, Benin. *BioImpacts : BI* **2012**; 2:159-66.
108. Ouedraogo S, Bodeau-Livinec F, Briand V, et al. Malaria and gravidity interact to modify maternal haemoglobin concentrations during pregnancy. *Malaria journal* **2012**; 11:348.
109. Taylor SM, Antonia A, Feng G, et al. Adaptive evolution and fixation of drug-resistant *Plasmodium falciparum* genotypes in pregnancy-associated malaria: 9-year results from the QuEERPAM study. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **2012**; 12:282-90.
110. Taylor SM, Parobek CM, DeConti DK, et al. Absence of Putative Artemisinin Resistance Mutations Among *Plasmodium falciparum* in Sub-Saharan Africa: A Molecular Epidemiologic Study. *The Journal of infectious diseases* **2015**; 211:680-8.
111. Vigan-Womas I, Lokossou A, Guillotte M, et al. The humoral response to *Plasmodium falciparum* VarO rosetting variant and its association with protection against malaria in Beninese children. *Malaria journal* **2010**; 9:267.
112. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *The New England journal of medicine* **2001**; 344:467-71.
113. Ballard JL, Khoury JC, Wedig K, Wang L, Eilers-Walsman BL, Lipp R. New Ballard Score, expanded to include extremely premature infants. *The Journal of pediatrics* **1991**; 119:417-23.
114. World Health Organization (WHO). Global Nutrition Targets 2025: Low birth weight policy brief, **2014**.
115. Luxemburger C, McGready R, Kham A, et al. Effects of malaria during pregnancy on infant mortality in an area of low malaria transmission. *American journal of epidemiology* **2001**; 154:459-65.

116. Menendez C, Ordi J, Ismail MR, et al. The impact of placental malaria on gestational age and birth weight. *The Journal of infectious diseases* **2000**; 181:1740-5.
117. Steketee RW, Wirima JJ, Hightower AW, Slutsker L, Heymann DL, Breman JG. The effect of malaria and malaria prevention in pregnancy on offspring birthweight, prematurity, and intrauterine growth retardation in rural Malawi. *The American journal of tropical medicine and hygiene* **1996**; 55:33-41.
118. Sullivan AD, Nyirenda T, Cullinan T, et al. Malaria infection during pregnancy: intrauterine growth retardation and preterm delivery in Malawi. *The Journal of infectious diseases* **1999**; 179:1580-3.
119. Verhoeff FH, Brabin BJ, van Buuren S, et al. An analysis of intra-uterine growth retardation in rural Malawi. *European journal of clinical nutrition* **2001**; 55:682-9.
120. Landis SH, Lokomba V, Ananth CV, et al. Impact of maternal malaria and under-nutrition on intrauterine growth restriction: a prospective ultrasound study in Democratic Republic of Congo. *Epidemiology and infection* **2009**; 137:294-304.
121. Wilcox AJ, Skjaerven R. Birth weight and perinatal mortality: the effect of gestational age. *American journal of public health* **1992**; 82:378-82.
122. Plowe CV, Djimde A, Bouare M, Doumbo O, Wellem TE. Pyrimethamine and proguanil resistance-conferring mutations in *Plasmodium falciparum* dihydrofolate reductase: polymerase chain reaction methods for surveillance in Africa. *The American journal of tropical medicine and hygiene* **1995**; 52:565-8.
123. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **2013**; 30:2725-9.
124. Hathaway NJ. SeekDeep: Targeted Amplicon Analysis. Available at: <http://baileylab.umassmed.edu/SeekDeep/index.html>.
125. Colwell RK, Chao A, Gotelli NJ, et al. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* **2012**; 5:3-21.
126. Colwell R. EstimateS: Statistical estimation of species richness and shared species from samples. 2005. Consultado en: <http://viceroy.eeb.uconn.edu/estimates> **2013**.
127. Nei M. *Molecular evolutionary genetics*. New York: Columbia University Press, **1987**.
128. Magurran AE. *Measuring Biological Diversity*. Wiley, **2003**.
129. Chao A. Estimating the number of shared species in two communities. *Statistica Sinica* **2000**; 10:227-46.

130. Chao A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **1987**; 43:783-91.
131. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution* **1993**; 10:512-26.
132. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **2004**; 20:289-90.
133. Wright S. Isolation by Distance. *Genetics* **1943**; 28:114-38.
134. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**; 123:585-95.
135. Weedall GD, Conway DJ. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends in parasitology* **2010**; 26:363-9.
136. Carlo WA. Prematurity and intrauterine growth restriction. In: Kliegman RM, Stanton, Bonita M.D., Geme, Joseph St., Schor, Nina F., ed. *Nelson Textbook of Pediatrics*. Philadelphia, PA: Elsevier Saunders, **2011**.
137. Sander AF, Salanti A, Lavstsen T, et al. Multiple var2csa-type PfEMP1 genes located at different chromosomal loci occur in many *Plasmodium falciparum* isolates. *PloS one* **2009**; 4:e6667.
138. Barry AE, Leliwa-Sytek A, Tavul L, et al. Population genomics of the immune evasion (var) genes of *Plasmodium falciparum*. *PLoS pathogens* **2007**; 3:e34.
139. Doritchamou J, Sabbagh A, Jespersen JS, et al. Identification of a Major Dimorphic Region in the Functionally Critical N-Terminal ID1 Domain of VAR2CSA. *PloS one* **2015**; 10:e0137695.
140. Wilcox AJ. On the importance--and the unimportance--of birthweight. *International journal of epidemiology* **2001**; 30:1233-41.
141. Strategic Advisory Group of Experts (SAGE) on Immunization: WHO, **2015**.