Using whole *Plasmodium falciparum* genomes for monitoring of escape from diagnosis, drug treatment and vaccine-induced immunity in Central-West Africa

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Thomas Carl Stabler

2024

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Erstbetreuerin: Prof. Dr. Claudia A Daubenberger, PhD

Zweitbetreuer: Prof. Dr. Pascal Mäser, PhD

Externer Experte: Prof. Dr. Michael von Fricken, PhD

Basel, den 17.09.2024

Prof. Dr. Marcel Mayor

Dekan

Table of contents

ACKNOWLEDGEMENTS

The presented doctoral work would not have been possible without the unwavering support of my co-supervisors, Prof. Claudia Daubenberger and Prof. Joana C. Silva. I am incredibly grateful for your guidance, patience, and encouragement throughout this process. You've made these past few years enjoyable and productive, even in the face of a global pandemic.

I want to thank Prof. Pascal Mäser, who first introduced me to bioinformatics and further supported my doctoral progress as my second supervisor. Additionally, I am very grateful to Prof. Michael E. Von Fricken for providing essential feedback as my external expert.

From SwissTPH, I want to thank Dr. Salome Hosch and Johanna Giger for addressing all my laboratory-related questions, reviewing my manuscript drafts, and updating me on the local women's rugby stats. I am also indebted to Dr. Tobias Schindler, who continued to mentor and support me after moving to another organization. From the Institute of Genome Sciences, Kemi Ifeonu, Wilson Tavares, Dr. Gillian Mbambo, Yiwei Yang, and Ryan Scalsky were instrumental in my development, working with me and providing valuable feedback. A special thank you goes to Dr. Ankit Dwivedi at IGS, who provided mentorship, guidance, and support for many of my computational analyses. I cannot overstate how fortunate I feel to have such supportive colleagues and friends.

I would also like to acknowledge the contributions from my colleagues at the Bioko Island Malaria Elimination Project (BIMEP) and the reference laboratory in Baney. Specifically, the entire BIMEP field team for their efforts in conducting the annual MIS, and the laboratory staff who eagerly worked with me to process a mountain of samples.

I would like to express deep gratitude to my other collaborators: Dr. Christopher Schwabe, Wonder Philip Phiri, Dr. Carlos Guerra, Elizabeth Nyakarungu, Dr. Max Mpina, Biraj Shrestha, and Sudhaunshu Joshi, who all provided much-needed support and helped make this PhD project a success. A special thank you to Guillermo Garcia, who has been my colleague and great friend since I first landed on Bioko Island in 2014. Working with such dedicated, passionate people makes the work we do so rewarding.

To my parents and brothers, this thesis is a testament to the unwavering support and love you've given me throughout my life. Love you.

Most of all, I would like to express my heartfelt appreciation for my wife, Olivia, whose love and care get me up each morning. Also Mochi, our canine daughter.

SUMMARY

In the first section of the thesis, I aimed to assess available malaria diagnostic tools and improve the visualization and identification of target genes involved in rapid diagnostic tests (RDTs) using whole genome sequences (WGS). Part 1 is comprised of 2 manuscripts.

Manuscript 1: Diagnostic performance and comparison of ultrasensitive and conventional rapid diagnostic test, thick blood smear and quantitative PCR for detection of low-density *Plasmodium falciparum* infections during a controlled human malaria infection study in Equatorial Guinea.

A novel ultrasensitive RDT (uRDT) has been made available to identify low parasite density infections while maintaining conventional RDT's versatility. This manuscript evaluates the efficacy of different diagnostic tools and their ability to detect low density *Plasmodium falciparum* infections. The research compares a uRDT, a conventional rapid diagnostic test (cRDT), thick blood smear (TBS), and quantitative PCR (qPCR) using samples from malaria-exposed individuals participating in a controlled human malaria infection (CHMI) study. The study found that qPCR was the most sensitive method, detecting infections earlier and at lower parasite densities compared to TBS, uRDT, and cRDT. TBS and uRDT demonstrated similar sensitivities and were more effective than cRDT, but none matched the sensitivity of qPCR. The study concludes that while qPCR remains the most sensitive diagnostic tool, uRDT could serve as a viable alternative to TBS in specific applications due to its comparable sensitivity and operational simplicity.

Manuscript 2: Gene Coverage Count and Classification (GC3): a locus sequence coverage assessment tool using short-read whole genome sequencing data, and its application to identify and classify histidine-rich protein 2 and 3 deletions in *Plasmodium falciparum*.

A concerning trend among *P. falciparum* populations is the expansion of strains with non-functional HRP2, the target antigen of RDTs, leading to evasion of RDT-based detection of ongoing *P. falciparum* infections. This manuscript describes the development and validation of a computational tool termed GC_3. This tool is designed to analyze whole genome shotgun sequence (WGS) data to detect and classify deletions in the *hrp2* and *hrp3* (*hrp2/3*) genes of *P. falciparum*, which are critical for the efficacy of malaria RDTs. The study highlights the tool's ability to accurately identify these deletions by comparing its results with those of a

hrp2/3-specific qPCR assay, demonstrating high agreement. GC_3 was applied to various WGS datasets, identifying *hrp2/3* deletions in samples from Cambodia, Brazil, and other regions using publicly available WGS data. The tool's application is crucial for monitoring the spread of *hrp2/3* deletions among sequence data when access to a samples' physical material is limited. GC_3 facilitates the analysis of WGS data and aims to provide investigators the tools to detect gene deletions.

The second part of the thesis aims to use epidemiology, population genetics and molecular monitoring methods to describe transmission dynamics of *P. falciparum* on Bioko Island, Equatorial Guinea. Using data from malaria indicator surveys (MIS) conducted annually on the island, regression models and WGS data was used to characterize human and parasite clusters that provide insight into transmission pathways. To ascertain the impact of importation on Bioko Island, population genetics was used to estimate the impact of suspected incoming *P. falciparum* strains to Bioko Island from neighboring regions on the African continent. The WGS data generated is the first whole genome description of the *P. falciparum* population on Bioko Island. Additionally, the molecular monitoring of *P. falciparum* using the Diaxxo PlasmoPod device is the first instance of using the machine with an entirely local staff at the Baney reference laboratory in Equatorial Guinea. This section is based on three manuscripts.

Manuscript 3: Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions.

The Bioko Island Malaria Elimination Project (BIMEP) has conducted annual MIS on Bioko Island, Equatorial Guinea since 2004. This manuscript investigates the various factors contributing to malaria infection risks on the island utilizing data from the 2015 and 2018 MIS, along with environmental data. Multilevel logistic regression models are applied to the data to analyze the odds of malaria infection according to human behavior. Findings suggest that travel to mainland Equatorial Guinea is the most significant risk factor, increasing infection odds by four times. Additionally, using long-lasting insecticidal nets (LLINs) significantly reduces infection risk by about 30%. Malaria prevalence was highest among children aged 10-14 and was similar between genders. The study underscores the importance of tailored interventions based on local contexts to enhance malaria control efforts.

Manuscript 4: Whole genome sequences of *Plasmodium falciparum* on Bioko Island, Equatorial Guinea illustrate connectedness with geographical neighbors and island-wide partial population structure.

The *P. falciparum* population on Bioko Island, Equatorial Guinea has not been previous described at a genomic level. This manuscripts provides estimates of the genetic diversity and population structure of *P. falciparum* on Bioko Island. Despite significant malaria control efforts, malaria prevalence remains high at 10-12%. This study utilized WGS data from samples collected during the 2019 malaria indicator survey to analyze genetic diversity, differentiation, and relatedness among *P. falciparum* isolates from Bioko Island. Results were compared to publicly available WGS data of continental sub-Saharan African parasites, like those from neighboring regions such as Cameroon and Gabon. Results showed high genetic diversity and a substantial proportion of polyclonal infections among Bioko parasites, indicating sustained transmission intensity and influence from imported malaria cases. Bioko parasites appeared most closely related to Gabonese parasites, suggesting a significant proportion go imported strains come from Gabon or nearby regions, such as Rio Muni, the continental region of Equatorial Guinea. These findings suggest that parasite migration from the mainland poses a significant challenge to malaria elimination on Bioko Island. Continuous molecular surveillance and targeted interventions at high traffic border crossings may be required to decrease imported malaria cases and to sustain effectiveness of malaria control activities implemented on Bioko Island.

Manuscript 5: Integrating local malaria molecular monitoring into regular malaria indicator surveys on Bioko Island: high association between urban communities and low density infections.

This manuscript evaluates the effectiveness of the Diaxxo PlasmoPod, a portable qPCR device, in detecting malaria infections evading diagnosis by conventional rapid diagnostic tests (RDTs), such as *hrp2*-deletion strains and low density infections. Conducted on Bioko Island, Equatorial Guinea, the research involved 1,500 dried blood spot (DBS) samples from the 2023 malaria indicator survey (MIS). An 18S rDNA/rRNA RT-qPCR-based assay was used with PlasmoPod, identifying *Plasmodium* spp. infection within 30 minutes. The study found 26.5% of RDT-negative individuals were qPCR positive, identifying false negatives by RDTs. Further, 74.7% of RDT-positive individuals were qPCR-positive, suggesting 25.3% were false positive RDT results. Among age groups, children under 5 years had higher parasite densities on average. Multivariate regression models revealed older age groups, males, and individuals from the Baney district had a higher probability of malaria infection. In terms of low density infections, MIS participants from urban communities, and individuals who recently reported fever, and households with higher density (people/room) had higher odds of having a low-density infection below an RDT's limit of detection. The study highlights the importance of integrating molecular monitoring, like PlasmoPod, into malaria surveillance; such monitoring improves infection detection, particularly for cases missed by

RDTs, and thereby providing a sensitive and accurate assessment of malaria prevalence and aiding better-informed intervention strategies.

The third section of this thesis applies population genetics and bioinformatics to NGS data to inform the next generation of malaria vaccines. Gene-specific long read sequence data from public repositories and WGS from clinical trials was used to assess diversity and review the potential of a single-locus, whole protein-based target. The application of a whole genome sieve analysis (wgSA) using WGS for *P. falciparum* was also reviewed. The subsequent work is the first instance of applying wgSA to *P. falciparum* to identify differential genomic characteristics between infected vaccinated and infected control groups in clinical trials conducted in Africa.

Manuscript 6: *Plasmodium falciparum* merozoite surface protein 1 as asexual blood stage malaria vaccine candidate.

Two malaria vaccines, RTS,S/AS01 and R21/Matrix-M have been approved by the WHO for use as preventative vaccines, but their reported efficacy remains below ideal standards. This manuscript reviews the potential of merozoite surface protein 1 (MSP1) as a vaccine target for malaria. MSP1 has been identified as a promising candidate due to its critical role in the asexual blood stage of the parasite. The review covers the biological functions of MSP1, its genetic diversity, and the outcomes of clinical trials that have tested various MSP1 constructs and formulations in both animal models and humans. Despite initial promise, human trials using MSP1 subunits or combinations with other antigens have shown limited efficacy. Recent advancements include testing a full-length MSP1 vaccine (MSP1FL) in malaria-naive volunteers, showing potential in generating multifunctional antibodies. The review emphasizes the need for further research to understand the relationship between antibody responses and protective immunity against asexual blood stages in malaria.

Manuscript 7: Whole-genome sieve analysis: identification of protective malaria antigens by leveraging allele-specific vaccine efficacy.

This manuscript explores the application of whole-genome sieve analysis (wgSA) in malaria research to identify protective antigens against *P. falciparum*. Traditional sieve analysis demonstrated the allele-specific vaccine efficacy for RTS,S/AS01. Higher protection was observed against *P. falciparum* strains with closer genetic variation in the target protein, circumsporozoite protein (CSP), to the vaccines'. The wgSA scans the entire genome to pinpoint differences between the infections of vaccinated and individuals who received placebo, and thus identifying the genetic variations associated with vaccine-induced

protection. This manuscript underscores wgSA's efficacy in high-recombination pathogens, reducing genetic linkage and noise, and offering a comprehensive approach to discovering protective loci. This novel approach holds significant promise for advancing malaria vaccine development and efficacy assessment.

Manuscript 8: Protective targets of PfSPZ-based whole organism vaccines identified from genome-wide sieve analyses of isolates from field efficacy trials.

This manuscript describes to study investigating the identification of protective antigens in malaria vaccines. Utilizing whole-genome sieve analysis (wgSA), the study analyzes parasite genomic data from two field trials of PfSPZ vaccines in malaria-exposed adults in Mali and Burkina Faso. In Burkina Faso, a radiation-attenuated whole sporozoite vaccine (PfSPZ Vaccine) was administer and compared to a saline placebo control group. In Mali, a chemo-attenuated version of the vaccine was used (PfSPZ-CVac), also compared to a control group. These trials, conducted in high-transmission areas, reveal that certain genomic sites and protein-coding loci are significantly underrepresented among breakthrough infections in vaccinated individuals compared to placebo recipients. Specifically, the study identifies 165 and 93 target loci in the Burkina Faso and Mali trials, respectively. Of which, there were 18 loci common to both, including the well-known pre-erythrocytic antigen PfSSP2/TRAP. The research demonstrates that wgSA can be an effective tool for identifying protective vaccine antigens, supporting the development of multivalent or multi-strain vaccines to combat the genetic diversity of *Plasmodium falciparum*.

LIST OF ABBREVIATIONS

ACT	Artemisinin combination therapy
AL	Artemether and lumefantrine
ASAQ	Artesunate/amodiaquine
ASMQ	Artesunate and mefloquine
ASTMH	American Society of Tropical Medicine and Hygiene
BIMEP	Bioko Island Malaria Elimination Program
bp	Base pairs
CHMI	Controlled Human Malaria Infection
CIA	Central Intelligence Agency
Cq	Quantification cycle
CSP	Circumsporozoite protein
CRT	Chloroquine resistance transporter
DBS	Dried blood spots
DHA-PPQ	Artesunate/sulfadoxine/pyrimethamine/dihydroartemisinin-piperaquine
DNA	Deoxyribonucleic acid
DRC	Democratic Republic of Congo
EG	Equatorial Guinea
EKNZ	Ethikkommission Nordwest- und Zentralschweiz
Fst	Fixation index
Fws	Complexity of Infection
GAC	Glideosome-associated connector
GATK	Genome Analysis Toolkit
GC_3	Gene Coverage Count and Classification
GWAS	Genome wide association study
HRP2	Histidine-rich protein 2
IBD	Identity-by-descent
IGS	Institute of Genome Sciences
ITN	Insecticide treated net
IRS	Indoor residual spraying
LD	Linkage disequilibrium
LLIN	Long-lasting insecticidal net
LMIC	Low-middle income countries
MCDG	Medical Care Development Global
MIS	Malaria Indicator Survey
MSP	Merozoite surface protein
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing

NIH	National Institutes of Health
NMP	National Malaria Program
ONT	Oxford Nanopore Technology
PBMC	Peripheral Blood Mononuclear Cell
PCA	Principal Components Analysis
qPCR	Quantitative polymerase chain reaction
Pf	*Plasmodium falciparum*
pLDH	*Plasmodium* lactate dehydrogenase
PPHS	Program Health Sciences
RAS	Radiation-attenuated sporozoites
RDT	Rapid Diagnostic Test
RNA	Ribonucleic acid
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
SNP	Single nucleotide polymorphism
SP	Sulfadoxine and pyrimethamine
SRA	Sequence Read Archive
sWGA	Selective whole genome amplification
TRAP	Thrombospondin-related adhesion protein
USD	United States dollar
WGS	Whole genome shotgun sequencing
wgSA	Whole genome sieve analysis
WHO	World Health Organization

1 INTRODUCTION

1.1 Malaria

1.1.1 Life Cycle of causative agent

Malaria is caused by a single-cell, eukaryotic parasite that belongs to the Apicomplexa phylum and the *Plasmodium* genus. Malaria parasites exist in various continents from South America to Africa, the Indian subcontinent and Southeast Asia (WHO, 2023). There are seven *Plasmodium* species infecting humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium simium*, *Plasmodium malariae*, *Plasmodium brasilianum*, *Plasmodium ovale*, and *Plasmodium knowlesi* (Sato, 2021, WHO, 2023). The *P. ovale* species is further divided into two sympatric subspecies: *P. ovale curtisi* and *P. ovale wallikeri*, only dissimilar genetically and these lineages split an estimated 1 - 3.5 million years ago (Sutherland et al., 2010). *P. knowlesi*, *P. brasilianum*, *P. simium* are regarded as a zoonotic malaria parasites. *P. knowlesi* is more commonly observed in macaques (Lee et al., 2011), and *P. brasilianum* and *P. simium* in South American primates (Fuehrer et al., 2022, Rougeron et al., 2022). *P. knowlesi* infections in humans have been steadily increasing in Southeast Asia causing clinical disease (Sato, 2021, Chin et al., 1968). *P. brasilianum* is genetically similar to *P. malariae*, to the point of being nearly indistinguishable (Tazi and Ayala, 2011, Fuentes-Ramírez et al., 2017, Fuehrer et al., 2022). Concurrently, *P. simium* is very similar genetically to *P. vivax*, believed to have undergone a host switch from human *P. vivax* into non-human primates (Mourier et al., 2021). Globally, there were an estimated 249 million clinical malaria cases and 608,000 deaths in the year 2022, primarily driven by *P. falciparum* infections in sub-Saharan Africa (WHO, 2023).

Malaria is an insect-borne infectious disease; the causative parasites, with a complex life cycle (Figure 1), are transmitted via female *Anopheles* mosquitos during blood feeding on a human host. At the time of feeding, sporozoites located in the mosquito's salivary glands are expelled into the human dermal layer. From there, within a matter of minutes, they migrate to nearby capillaries, entering the bloodstream and then reach the liver, where they invade hepatocytes (Ryan et al., 2019). Following an incubation period of 1-2 weeks, thousands of liver stage merozoites are asexually produced and released into the blood stream to invade erythrocytes. The duration of the liver-stage incubation period, which is clinically silent, varies according to *Plasmodium* species (Ashley et al., 2018). During the liver stages of *P. vivax* and *P. ovale*, hypnozoites are also produced, thought to be lying dormant in hepatocytes and reactivate days, months or even years after the initial infection took place (Jong and Stevens, 2021). Once parasites invade erythrocytes, they go through repeated asexual blood stage cycles, occurring every 48 hours for *P. falciparum*, *P. vivax*, *P. simium*, and *P. ovale*, 72 hours for *P.*

malariae and *P. brasilianum*, and 24 hours for *P. knowlesi* (White et al., 2014, Ryan et al., 2019, Fuehrer et al., 2022, Rougeron et al., 2022). After each asexual blood-stage cycle, merozoites are released into the blood stream from ruptured schizonts, invade healthy erythrocytes and repeat the asexual reproduction cycle (Ryan et al., 2019). It is during this blood stage of repeated invasion and lysing of erythrocytes that human hosts may experience symptoms due to erythrocyte destruction and accompanying immune responses, whereby the type and outcome of the immunological defense depends on the virulence of the infecting *Plasmodium* species/strain and level of previous malaria exposures (i.e. naïve *vs*. semi-immune individuals) (Stanisic et al., 2009, Walker and Rogerson, 2023). When there are diminished signals from a lack of essential host nutrients required for continued asexual progress (Brancucci et al., 2017) a subset of merozoites undergo sexual differentiation into male and female gametocytes. After gametocytes mature and migrate to host capillaries (stage V), some are eventually taken up by feeding mosquitos, where male gametocytes undergo mitosis into microgametes and fertilize female gametocytes in the mosquito's midgut (Josling and Llinás, 2015). The resulting zygote, the only diploid stage of the *Plasmodium* lifecycle, develops into haploid ookinetes that invade the mosquito midgut wall, develop further into oocysts, and after another incubation period (1-2 weeks), finally reach the sporozoite stage that migrates to the mosquito salivary gland to repeat the life cycle (Bennink et al., 2016).

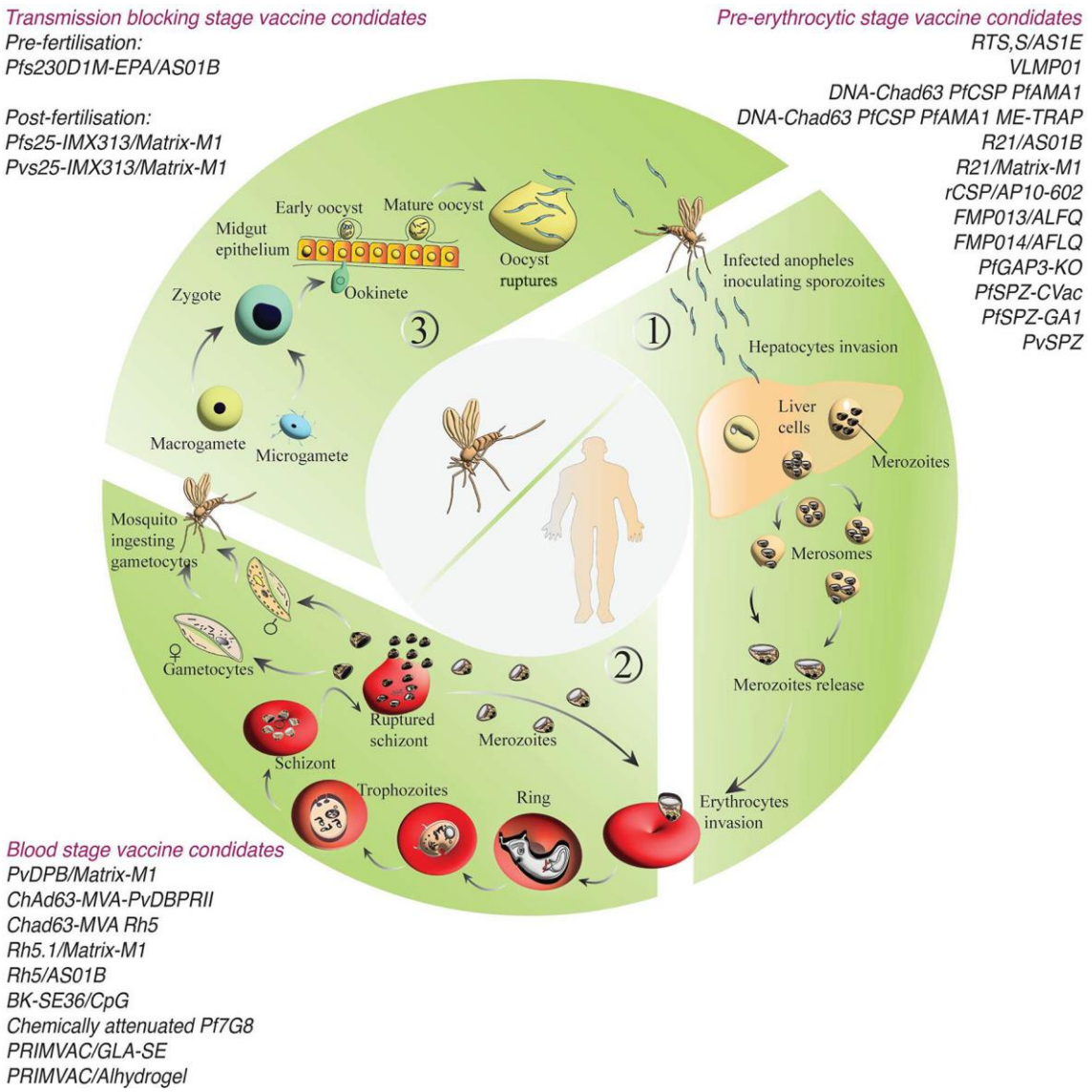


Figure 1. Complete life cycle of the *Plasmodium* parasite. **(1)** The pre-erythrocytic stage starts when sporozoites are injected into a human host by a female, infected *Anopheles* mosquito during a feeding event. Sporozoites migrate to the liver and infect hepatocytes, then asexually replicate and mature into merozoites after an incubation period. Once parasites are released from the liver, **(2)** the erythrocytic stage begins as merozoites escaping the liver to invade erythrocytes. The majority of parasites infecting erythrocytes undergo asexual reproduction, replicating at an exponential rate, destroying erythrocytes by lysing at the end of each asexual cycle, and eliciting the initial symptoms experienced by the human host. A small subset of infecting merozoites enter sexual development into male or female gametocytes. When a mosquito takes up gametocytes during a feeding event, **(3)** the male and female gametocyte fuse and undergo several steps in the midgut to eventually become sporozoites that migrate to the salivary glands. Image extracted from Rajneesh et al., 2023, and lists malaria vaccine candidates based on their vaccine strategy: pre-erythrocytic, erythrocytic, transmission-blocking (Rajneesh et al., 2023).

1.1.2 Epidemiology

Malaria is still one of the leading causes of mortality among children under 5 years of age and pregnant women (primarily primigravidae) in malaria endemic countries (Liu et al., 2021, Kogan, 2020). The majority of malaria-related morbidity and mortality cases occur in the sub-Saharan African region, accounting for 94% of global malaria cases (WHO, 2023). Nigeria, the

Democratic Republic of Congo (DRC), Uganda and Mozambique account for 27%, 12%, 5% and 4%, respectively, of all African malaria cases in 2022 (WHO, 2023). The majority of cases in Africa are due to *P. falciparum* infections, although non-*falciparum* species, such as *P. malariae* and *P. ovale*, are often observed as co-infections in some regions (WHO, 2023). Of concern are recent trends of increasing malaria prevalence, in part due to the interruptions in malaria control interventions due to the COVID-19 pandemic, growing from 231 million cases in 2015 to 249 million cases in 2022 (WHO, 2023). This global pattern illustrates the challenge of malaria elimination where delays or interruptions of control strategies can result in a rapid loss of progress gained over many years.

1.1.3 Vector Control and Prevention

There are several malaria control and prevention interventions that are implemented globally, designed mainly to reduce the probability of malaria transmission by its vectors. Distribution of insecticide-treated bed nets (ITNs) or long-lasting insecticide-treated bed nets (LLINs), and indoor residual spraying (IRS) are the more traditional interventions, which target feeding mosquitos within a household. ITNs have become a significant intervention strategy over the last decade, with almost 2.9 billion ITNs/LLINs distributed from 2004 - 2022 (WHO, 2023). Among malaria cases estimated to be averted between 2000 - 2015, ITNs/LLINs contributed 68% (Bhatt et al., 2015), and individuals who used ITNs had a 44% decrease in malaria incidence compared to non-ITN using individuals (Pryce et al., 2018). Although IRS is a more labor-intensive control intervention, requiring sprayers to cover the walls with insecticide in every room of a household, it has been estimated to reduce malaria prevalence by 68% (Gimnig et al., 2024, WHO, 2023). Another mosquito-targeting intervention, larvaciding, consists in the treatment of *Anopheles spp.* breeding sites with chemical or microbial insecticides to curtail mosquito hatching, decreasing malaria prevalence by 74% in treated communities *vs*. non-treated communities (Choi et al., 2019). Although initially beneficial, larvaciding effects appear to be short-lived and do not sufficiently reduce disease transmission by mosquitos (Dambach et al., 2021). Although effective so far, recent expansion of insecticide-resistant mosquitos has been observed, threatening the longer-term effectiveness of mosquito-targeting control interventions (Suh et al., 2023).

1.1.4 Drug Treatment

Malaria is a treatable disease, with artemisinin-based combination therapies (ACTs) being the current frontline drug treatment approach, globally. Symptomatic malaria typically manifests as flu-like symptoms such as fever, chills, headache, cough, diarrhea, malaise and achy joints (Warrell, 2017). Complicated malaria in children can manifest as severe anemia, respiratory distress and eventual coma, whereas in adults similar symptoms are present (except anemia),

and many cases also include renal injury and jaundice (White, 2022). If not treated rapidly and effectively, cases can progress to more severe malaria presentations where second-line treatments, such as Artesunate (a water soluble artemisinin-derivative) or Quinine combined with tetracycline or doxycycline, need to be administered intravenously for up to 7 days to effectively clear parasites (WHO, 2015). Cerebral malaria, the most severe malaria complication, is 100% fatal without treatment, but even with treatment mortality remains high, at 15-20%, accompanied with long-term health sequelae (Dvorin, 2017).

Unfortunately, as with increasing insecticide resistance, an expansion of drug-resistant *Plasmodium* strains has been observed historically. Initial treatment with chloroquine in the early 1900s eventually became ineffective after chloroquine-resistant strains emerged in South America and Southeast Asia independently, eventually spreading worldwide (Payne, 1987). Several single nucleotide polymorphisms (SNPs), or base-pair mutations, were observed on the *P. falciparum* chloroquine resistance transporter (PfCRT) gene that differentiated susceptible from resistant strains to chloroquine treatment, including those leading to amino acid residue changes: M74I, N75E, K76T, A220S, Q271E, N326S, I356T, and R371I (Wellems and Plowe, 2001). The PfCRT product is a membrane-bound protein associated with drug/metabolite transport and mediates the removal of chloroquine from the digestive vacuole (Bray et al., 2005). Due to chloroquine resistance, national malaria programs (NMPs) adopted a new frontline antimalarial drug combination: sulfadoxine and pyrimethamine (SP). SP was effective until multiple resistant mutations expanded out of Cambodia in the 1970s and reduced the antimalarial's effectiveness (Gatton et al., 2004, Vinayak et al., 2010). SP inhibits folate digestion targeting the protein products of dihydropteroate synthase (*dhps*) and dihydrofolate reductase (*dhfr*); however, the number and frequency of drug resistance-inducing SNPs in these genes increased in field isolates, with concomitant reduction of SP effectiveness (Sibley et al., 2001). SP wasn't adopted in Africa until the early 1990s, but rapid spread of resistance quickly occurred where, by the end of the decade, most countries had to adapt by switching to different frontline antimalarials (Conrad and Rosenthal, 2019).

ACTs currently remain effective frontline treatments, where artemisinin derivatives rapidly clear parasites in the blood. Due to the short half-lives of artemisinin derivatives, they need to be combined with a longer lasting partner drug to clear lingering parasites and prevent the development of resistant strains (Eastman and Fidock, 2009). Most ACTs used globally are artesunate/amodiaquine (ASAQ), artesunate and mefloquine (ASMQ), artemether and lumefantrine (AL), artesunate+sulfadoxine/pyrimethamine, dihydroartemisinin-piperaquine. (DHA-PPQ), artemisinin/piperaquine/primaquine, and pyroaridine/artesunate (WHO, 2023). However, despite ACTs' efficacy and combination with a range of partner drugs, there has been a rise in artemisinin-resistance strains, first reported in Cambodia in 2009 (Dondorp et al., 2009), subsequently spreading through Southeast Asia and just recently starting to be

reported in Africa (Hassett and Roepe, 2019, Balikagala et al., 2021). SNPs with the strongest association to artemisinin derivatives resistance are located on the *kelch13* gene, whose product is involved in various functions including the endocytosis process of host hemoglobin (Birnbaum et al., 2020). Verified SNPs in *kelch13* associated with artemisinin derivative resistance, including Y493H, R539 T, I543T, and C580Y (Ariey et al., 2014), appear to result in reduced endocytosis of host hemoglobin. Artemisinin activation requires degraded hemoglobin products, so this reduction of hemoglobin endocytosis in resistance strains diminishes the antimalarial effect (Birnbaum et al., 2020). At present, alternative drugs to ACTs are under development (Belete, 2020), but are not currently approved or widely available. The lack of a timely antimalarial alternative highlights the importance of molecular monitoring by NMPs. By measuring the occurrence and spread of *P. falciparum* strains carrying drug resistance-conferring markers allows timely adaption to control strategies, such as simultaneous or rotating deployment of differing ACT types (Audibert et al., 2024) or even deployment of triple combination therapies (Chen and Hsiang, 2022). Such strategies may reduce the emergence and expansion of drug-resistance strains, as has been observed with the return of chloroquine-susceptible parasites (Kublin et al., 2003, Laufer et al., 2006). However, as long as ACT's selective pressure on *Plasmodium* is present, artemisinin-resistant strains will continue to be selected for, and represent a threat to current antimalarial effectiveness.

1.1.5 Diagnosis

Essential to the successful drug treatment of malaria is accurate and prompt diagnosis of ongoing *Plasmodium* infection. The traditional method is microscopy-based, either through thick or thin blood smears. Thick blood smear allows for the detection and quantification of parasite density, while from thin blood smears *Plasmodium* species can be differentiated and blood stage progression of parasites identified (Gitta and Kilian, 2020). Although microscopy is considered the "gold standard" for malaria diagnosis, able to detect parasitemias (i.e. parasite densities) as low as 50-200 parasites/µl (Kolluri et al., 2018, Pham et al., 2018), it is not as accessible or robust as other widely used techniques (Moody, 2002). Many NMPs utilize malaria rapid diagnostic tests (RDTs), which are plastic cassettes covering an absorbent paper. Sections of this absorbent paper contain reagents that are activated (i.e., change color) when exposed to their target antigen. Among the 345 million RDTs distributed globally in 2022, 94% of WHO-qualified RDTs target histidine-rich protein 2 (HRP2), specific to *P. falciparum*. Many RDTs also include a pan-*Plasmodium* antigen, either lactate dehydrogenase (pLDH) or aldolase (Thomson et al., 2020, Gendrot et al., 2019). RDTs are easy to use, requiring only a drop of blood, and can detect infections down to 100 parasites/µl. HRP2-based RDTs are much more sensitive than pLDH RDTs at lower parasite density infections, showing, respectively, 85% and 50% sensitivity at 100-1,000 parasites/µl (Varo et al., 2021, Hendriksen et al., 2011).

False positive RDT diagnosis can occur due to persistent HRP2 in the blood up to 15 days post-treatment (Dalrymple et al., 2018). Further, false negative RDT diagnosis can occur due to *P. falciparum* strains with *hrp2* gene deletions that have been observed globally (Gendrot et al., 2019). Indicative of RDT's effectiveness and selective pressure, these deletions have extended to include HRP3, which cross-reacts with HRP2-RDTs in high density infections (Kong et al., 2021). Molecular diagnostic assays have the potential to overcome limitations of microscopy and RDTs, able to detect infections as low as 0.1 parasites/µl. These assays either utilize quantitative polymerase chain reaction (qPCR) targeting multi-copy gene targets, or a RT-qPCR that targets *Plasmodium* 18S nucleic acid material (DNA and RNA) (Kamau et al., 2013, Hofmann et al., 2015, Kamau et al., 2011). The more sensitive PCR protocols are particular advantageous to detect asymptomatic, sub-patent infections as they constitute large parasite transmission reservoirs (Cheaveau et al., 2019), estimated to contribute 20-50% of human to mosquito transmission in low endemicity regions (Okell et al., 2012). To be conducted reliably, molecular assays require specialized equipment, excellent logistics, and well-trained staff, hampering wider accessibility and utility as a field-deployable diagnostic technique (Varo et al., 2021). However, recent advances in technology have resulted in the development of the PlasmoPod device (Diaxxo AG, Zurich, Switzerland), which aims to account for the accessibility limitations of molecular techniques. The cartridge-based design (i.e. pods) comes with dehydrated reagents already included, allowing for transport at room-temperature. Pods only require the addition of extracted materials and immersing oil, and report similar results in terms of sensitivity and specificity to currently used qPCR and RT-qPCR techniques (Bechtold et al., 2023, Bechtold et al., 2021). Although still novel, the device itself is of smaller overall footprint than a standard PCR machine and may be an appropriate tool for molecular malaria surveillance in resource-limited environments.

1.1.6 Vaccine Development

Prevention methods are not limited to targeting the parasite-transmitting mosquitos, but extend to the fight against *Plasmodium* parasites themselves in the human host. Several vaccine candidates currently under development are listed in Figure 1. Based on the life cycle stage in which the target antigen is expressed and active (i.e. vaccine strategy), vaccine candidates are categorized as aiming to neutralize/elicit protection against sporozoites (pre-erythrocytic vaccines), merozoites and infected red blood cells (asexual blood stage vaccines), or gametocytes (transmission-blocking vaccines). Malaria vaccine strategies are further outlined in Figure 2. Significantly, the WHO now recommends two vaccines against *P. falciparum*, RTS,S/AS01 (commercial name: Mosquirix^{TM}), and R21/Matrix-M (WHO, 2023), both of which are pre-erythrocytic vaccines. They are based on the *P. falciparum* circumsporozoite protein (PfCSP), an essential protein involved in sporozoite development in the mosquito (Ménard et al., 1997), hepatocyte binding and invasion (Rathore et al., 2002), and parasite development

in infected hepatocytes (Singh et al., 2007). RTS,S and R21 have been shown to be, respectively, 56% and 78% effective against clinical malaria among children aged 5-17 months (RTS, 2015, Datoo et al., 2024). Analysis of the infective strains among vaccinated *vs*. placebo-receiving individuals illustrates that efficacy might be allele-specific, and hence that the vaccine is not as protective against *P. falciparum* strains with genetically distinct CSP alleles (Neafsey et al., 2015). These results suggest that some strains may evade vaccine-mediated protection (a phenomenon known as vaccine escape). Once these two vaccines become more widely available and used, selection of these strains may ensue, potentially compromising vaccine impact in the longer term (Kiyuka et al., 2020). Monitoring of circulating malaria populations for vaccine-evading strains, as well as development of next-generation malaria vaccines is essential to ensuring continued progress towards malaria elimination.

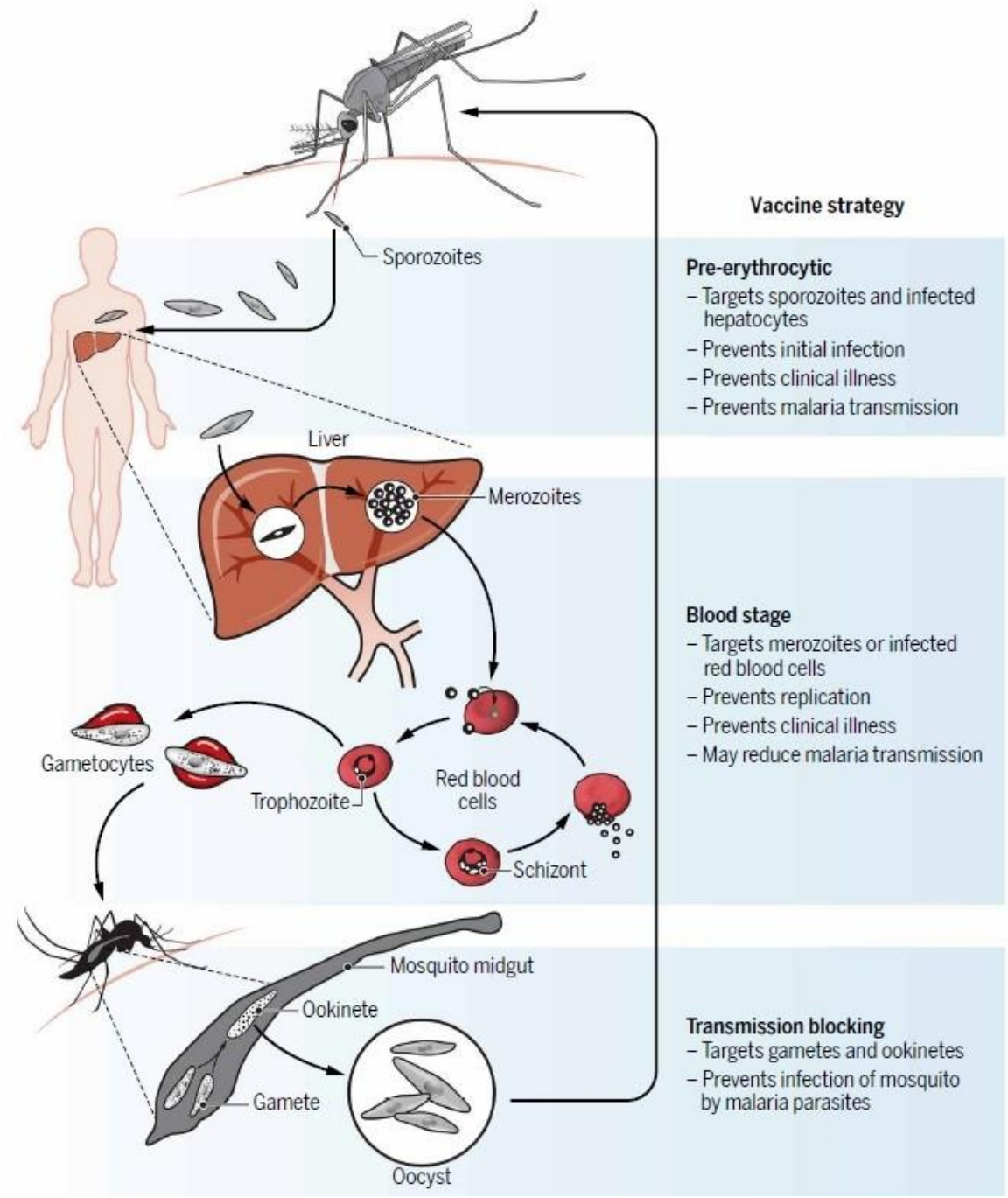


Figure 2. Malaria Vaccine Strategies. The vaccine target antigen implicates the type of elicited protection and interruption in the *Plasmodium* lifecycle stage. **Pre-erythrocytic** refers to vaccine candidates that target sporozoites and infected hepatocytes in the liver. The aim is to prevent initial infection, and subsequent blood and transmission stages. **Blood stage** (or erythrocytic) vaccines target merozoites escaping the liver and infected erythrocytes. This strategy prevents asexual replication and clinical illness of infected individuals, and may reduce the probability of transmission. **Transmission blocking** vaccines target gametocytes preventing further transmission and continuation of the *Plasmodium* life cycle onto the mosquito-stage. This strategy does not protect against human-host infection by *Plasmodium* of the vaccinee (altruistic vaccine approach. Image extracted from Beeson et al., 2019 (Beeson et al., 2019).

Although the WHO-approved vaccines are pre-erythrocytic, naturally acquired immunity does not appear to strongly target pre-erythrocytic parasites (Langhorne et al., 2008, Tran et al.,

2013). However, radiation-attenuated sporozoites (RAS) were observed to elicit sterile protection first in mouse models (Nussenzweig et al., 1967), and later in humans (Clyde, 1975, Luke and Hoffman, 2003), showcasing the potential of pre-erythrocytic vaccines. Advantages of pre-erythrocytic vaccines include targeting *Plasmodium* when fewer parasites are localized in the human host prior to the development and maturation of thousands of merozoites (Prudêncio et al., 2006). This thereby prevents infection and subsequent blood-stages, which can disrupt protective malaria immune responses (Keitany et al., 2016). As symptoms do not present until the blood-stage, human clinical trials have shown these vaccines to be benign and well tolerated (Jongo et al., 2022, Datoo et al., 2024, Syed, 2022). Recent advances in pre-erythrocytic malaria vaccine development have not directly translated into significant protective effects, especially among malaria endemic populations (Stanisic and Good, 2023). In part, this reflects the genetic diversity of circulating parasites within and between regions, and the challenge of identifying conserved antigen targets that elicit protection against geographically distinct *Plasmodium* populations.

Inherent in the complexity of malaria is the parasite's high genetic variability, and researchers difficulty in identifying strong correlates of immune protection (Cockburn and Seder, 2018). Antigens encoded by genes essential to liver-stage progression have been used in single-target vaccine candidates, such as CSP and thrombospondin-related adhesion protein (TRAP), which have been shown to elicit protection (Seder et al., 2013). A RAS vaccine candidate, PfSPZ Vaccine, exposes vaccinees to approximately 2,000 *Plasmodium* proteins to offer broader protection (Hoffman et al., 2015). This wider array of antigens aims to account for *Plasmodium's* diversity, where single target vaccines, like RTS,S, may be more susceptible to vaccine-evading strains than RAS vaccines, like PfSPZ Vaccine (Neafsey et al., 2015, Moser et al., 2020). Several pre-erythrocytic vaccines incorporate multi-epitope antigens or multiple antigen targets to compensate for malaria diversity, however significant protection has yet to be observed (Cockburn and Seder, 2018). Although protective correlates remain elusive, pre-erythrocytic vaccines appear to provide the highest levels of protection so far. Future vaccine iterations will likely need to incorporate either allelic variants of pre-erythrocytic vaccine candidates or a combination of pre-erythrocytic and asexual blood stage candidates to provide robust, longer-lasting protection against genetically variable strains circulating in the field.

1.2 Next Generation Sequencing

1.2.1 Technologies

Next generation sequencing (NGS) refers to the technologies that can sequence selected genes, partial or even whole genomes at low cost and high-throughput (Behjati and Tarpey,

2013). Sanger sequencing was the first high throughput nucleotide sequencing method, initially labelling dideoxynucleotides radioactively, and then fluorescently due to safety concerns, to generate a complementary DNA sequence to a template strand (Heather and Chain, 2016). Sanger sequencing is labor intensive and costly, being the primary method used to complete the $2.7 billion publicly-supported human genome project (Wetterstrand, 2020). However, advancements in technology and sequencing methods have made it more widely accessible and implementable in various fields. The WHO has outlined a global genomic surveillance strategy encouraging member states to incorporate NGS technology into their public health programs to molecularly monitor endemic and epidemic pathogens (Carter et al., 2022). In the case of malaria, NGS has been used to conduct genomic analyses to describe genotype diversity and temporal fluctuations (Laufer et al., 2010, Shrestha et al., 2021, Vareta et al., 2024, Guo et al., 2024), characterize population structure (Amambua-Ngwa et al., 2019), interpret gene selection (Mobegi et al., 2014, Neafsey et al., 2008, Travassos et al., 2018), establish genotype-specific vaccine efficacy (Neafsey et al., 2021, Abukari et al., 2019, Amambua-Ngwa et al., 2019, Thera et al., 2011, Ouattara et al., 2012), identify genes associated with drug resistance (Nwakanma et al., 2014, Cowell and Winzeler, 2019, Dharia et al., 2010a, Miotto et al., 2013, Takala-Harrison and Laufer, 2015), and case tracking (Preston et al., 2014, Tessema et al., 2019).

NGS approaches can be grouped as second-generation or third-generation sequencing technologies. Second-generation refers to technology that uses ligase or polymerase to incorporate nucleotides with an attached fluorescent signal (Heather and Chain, 2016, Metzker, 2010). Platforms utilizing this approach, such as Illumina or Ion Torrent, sequence “short-read” fragments (100-400 base pairs, abbreviated bp) in parallel and on a massive scale, where millions of sequencing reactions occur simultaneously (Heather and Chain, 2016). These techniques typically determine a nucleotide sequence by the detection of a released signal (e.g. pyrophosphate, hydrogen ion, fluorescence) tagged to synthesized nucleotide (Satam et al., 2023). Ion Torrent determines a sequence by changes in pH, where semiconductor chips detect for a single hydrogen ion released upon successful incorporation of a nucleotide (Merriman et al., 2012). The more widely used Illumina technology applies a method of sequencing-by-synthesis, where, through cycles, fluorescently-labelled nucleotides are washed over single-stranded templates, and the sequence is determined by the released nucleotide-specific fluorescent signal (Satam et al., 2023). These short-read technologies have very low error rates (0.1 – 1%), but are not ideal for assembly of genomes with highly biased nucleotide composition, often associated with long stretches of low complexity, and with sequence repeats longer than read length (Fox et al., 2014).

Third-generation sequencing, such as Pacific Biosciences (PacBio) or Oxford Nanopore Technology (ONT) apply a single molecular sequencing technique, able to generate “long

reads". Long read sequencing overcomes the assembly limitations of short reads, but raw reads have error rates 1-2 orders higher in magnitude (Fox et al., 2014). For PacBio, DNA is sequenced with a single-stranded circular template where, similarly to Illumina, released fluorescent light is measured when differentially-labelled nucleotides are successfully incorporated in the growing, complementary read (Rhoads and Au, 2015). Historically, the average length of raw PacBio reads ranged from 2,500 bp to 10,000 bp (Mason and Elemento, 2012). However, due to advancements in PacBio equipment and chemistry, read lengths can reach up to 12 – 20 kbp, consistently (Otto et al., 2018, van Dijk et al., 2023). ONT sequencing is conducted in a much smaller (hand-held) and more accessible device than PacBio, which uses a membrane embedded with nanopores. When a voltage is applied, single- or double-strand DNA/RNA is driven through a pore charged with an ionic current (Lu et al., 2016). Nucleic acid strand sequences are read in real time as each nucleotide passes through the pore causing unique changes in the ion current signal (Lu et al., 2016). Comparatively, MinION error rate is high (5-20%) compared to that of Illumina short reads (<1%) (Jain et al., 2018, Rang et al., 2018). Much like PacBio, several developments have been implemented to significantly increase accuracy, including increasing sequencing coverage (Kono and Arakawa, 2019, Noakes et al., 2019).

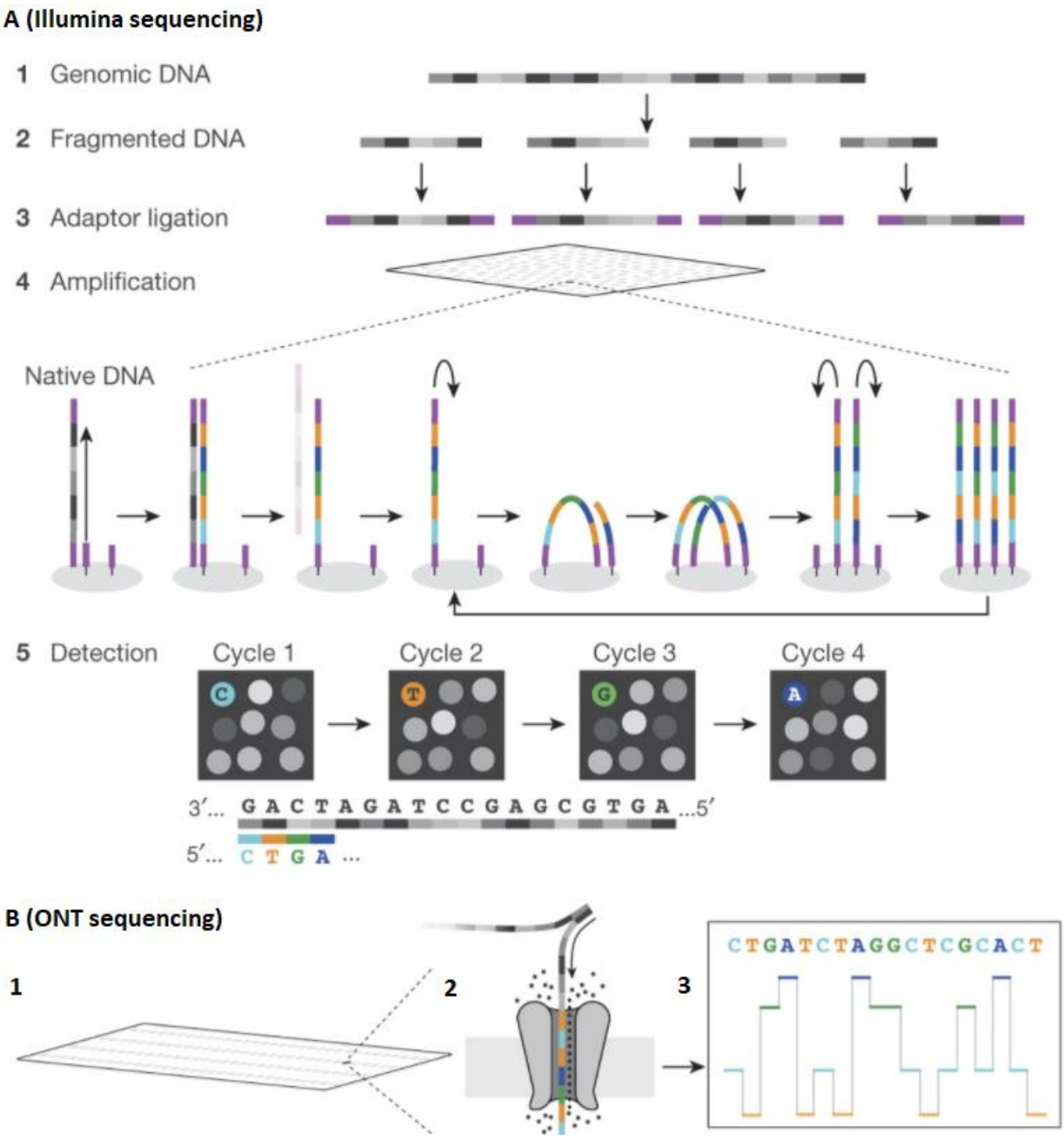


Figure 3. Principles of Illumina and Nanopore sequencing technologies. A. Illumina sequencing begins with (**A1**) sampled genomic material that is (**A2**) fragmented using enzymatic digestion or sonication. (**A3**) Adaptors are ligated to the fragments ends. (**A4**) Fragment-adaptor material is mounted on the flow cell and a cyclical bridge amplification process is conducted, creating fragment-adaptor copies. (**A5**) Fluorescent-labelled nucleotides are based over the flow cell and the device measures the nucleotide-specific light signals incorporated at each cluster site. B. Oxford nanopore technology (ONT) are (**B1**) made up of protein nanopores distributed across a flow cell. (**B2**) An electronic current is applied to the environment causing ions and sample DNA to pass through nanopores. Motor proteins unzip nucleic material and guide single strands through the pore, where each nucleotide causes specific changes in the ionic environment. (**B3**) As the ionic environment changes in response to individual nucleotides, the ONT device infers read sequences based on these electric signatures. Image extracted from Hackl et al., 2022 (Hackl et al., 2022), an adapted image from Shendure et al., 2017 (Shendure et al., 2017).

1.2.2 Whole genome sequencing

Whole genome sequencing (WGS) is the process of sequencing the entire genome of a target organism, including organelle DNA like the mitochondria and apicoplast for *Plasmodium spp*. (Ng and Kirkness, 2010). WGS offers a comprehensive view of sequenced genomes, with thousands of single nucleotide polymorphisms (SNP) identified by mapping generated sequences against a reference genome. These SNPs can fall into multiple genomic partitions, including most protein-coding regions, providing enough detail to conduct genome-wide association studies (GWAS), or population genetics studies (Wu et al., 2016, Neafsey et al., 2021). Library construction for WGS, also referred to as "shotgun sequencing", most commonly starts with the random shearing of entire chromosomes into smaller DNA fragments which, after sequencing, are reassembled into longer contiguous sequences (or contigs), ideally back to full length chromosomes (or supercontigs). A potential limitation of WGS is the uneven coverage across the genome. In the case of *Plasmodium*, hypervariable sub-telomeric regions and multi-gene families can result in high read coverage due to amplification bias during library construction (Oyola et al., 2016, Sabina and Leamon, 2015). Ultimately, this can also result in low coverage of these regions as many reads are discarded due to multi-location mapping and ambiguous alignment resulting in low confidence values. Whole genome assembly can be performed based on reference or *de novo* assembly. Reference-based assembly utilizes a reference genome to map each sequenced read to its most probable genomic location (Olson et al., 2015). Reference-based assembly is unable to place novel or highly divergent sequences in the new assembly, since homologous sequences are absent or unrecognizable in the reference genome. This may underestimate diversity by selecting for reads that can be more easily mapped to the reference genome (Olson et al., 2015). *De novo* refers to the assembly technique of comparing newly sequenced reads to each other, building the novel contiguous sequence based on overlaps between reads. *De novo* assembly has been shown to be accurate for small genomes, like those of bacteria (Kisand and Lettieri, 2013, Rodrigue et al., 2009), but remains a significant challenge for larger genomes due to missing sequencing data, the presence of repetitive or low complexity genome regions (such as high AT-rich regions in some *Plasmodium* genomes), uncorrected sequencing errors, and lack of adequate computational infrastructure (Liao et al., 2019).

1.2.3 Amplicon Sequencing

Targeted amplicon sequencing is a method based on first amplifying by PCR a specific chromosomal region or gene of interest, which is then sequenced. This focused approach allows for increased data quality at the target locus or loci and, due to its reduced time and cost, enables studies with much larger sample sizes (Ranjan et al., 2016). Depending on the target region, amplicon sequencing has been used to conduct population genetic studies and

can detect genetic variants at a low prevalence level in the parasite population (Meek and Larson, 2019). Amplicon-based sequencing can have limitations identifying duplicates, as virtually every read starts and ends with the same sequence of nucleotides, due to the specificity of the PCR primers (Poretsky et al., 2014, Ranjan et al., 2016). Further, achieving uniform coverage across the target region can be difficult, as certain segments may amplify better than others, and much like WGS, can over-represent regions with greater amplification propensity (Schirmer et al., 2015).

1.2.4 Sample collection

An important aspect of sequencing is the type of sample used as the source of nucleic material. Samples collected for DNA extraction and nucleotide sequencing can differ, like nasal swabs for respiratory viruses like SARS-CoV-2 (US FDA, 2020), or stool samples for norovirus (Strubbia et al., 2019). In the context of *Plasmodium*, parasite DNA is commonly extracted from peripheral venous blood samples, and then sequenced. Leukocyte-depleted blood samples, refers to whole blood samples that, at the time of collection, have undergone a filtering process to remove leukocytes, and thereby reduce host DNA content and increase the proportion of the parasite DNA prior to sequencing (Venkatesan et al., 2012). However, depleting leukocytes can be labor intensive for resource-limited settings, where depletion needs to be conducted shortly after sample collection. Alternatively to the removal of leukocytes from whole blood samples, venous whole blood stored as dried blood spots (DBS) can be obtained. These samples are more stable in field settings and require less upfront processing prior to storage and subsequent DNA isolation and downstream processing (Färnert et al., 1999, Bereczky et al., 2005). The challenge in using any of these samples for DNA sequencing is the comparatively low starting DNA material. Therefore, researchers commonly need to perform *Plasmodium* DNA enrichment relative to host DNA, through such processes as hybrid selection (a method based on preferential capture of *Plasmodium* DNA prior to sequencing) (Melnikov et al., 2011, Stucke et al., 2021), preferential enzymatic digestion of human DNA (Oyola et al., 2013), or selective whole genome amplification (sWGA) of Plasmodium DNA prior to sequencing (Oyola et al., 2016, Shah et al., 2020). Sequencing of *Plasmodium* amplicons has even been extended to the use of discarded RDTs as a source of parasite DNA, not only allowing for confirmation of RDT diagnosis, but genotyping of important malaria genes (Hosch et al., 2022). Similar to DBS, these samples may encounter issues related to low initial nucleic acid source material.

1.3 Thesis Aims

The overall aim of this doctoral work is to use molecular monitoring of *P. falciparum* as it pertains to diagnostic test evasion, understanding of parasite transmission dynamics and next generation malaria vaccine development.

For part one, the aim is to assess available malaria diagnostic tools and improve visualization of target genes involved in diagnosis using NGS coverage data by completing the following objectives:

1) Compare diagnostic capacity of ultra-sensitive RDT to conventional RDT, microscopy and a molecular assay
2) Generate a semi-automated computational approach (Python/R) to extract WGS coverage data of user-specified regions of a reference genome and output informative data files and plots
3) Apply computational approaches to detect and describe *hrp2* and *hrp3* gene deletions and validate outcome with an HRP2/HRP3-specific qPCR assay

In part two, the aim is to describe transmission dynamics of the *P. falciparum* population on Bioko Island, Equatorial Guinea, by completing the following objectives:

1) Generate WGS data from blood samples collected during the 2019 malaria indicator survey (MIS) and stored as DBS
2) Apply population genetics and genomic epidemiology principles and metrics to describe genetic diversity, and characterize population structure and dynamics in and around Bioko Island
3) Illustrate local transmission dynamics by combining NGS data with larger epidemiological data collected during the MIS

The aim of part three is to apply bioinformatics to NGS data to inform the next generation of malaria vaccines by completing the following objectives:

1) Assess the global genetic diversity of gene segments of *msp1* using publicly available long-read genome sequencing data
2) Conduct molecular sieve analysis of malaria samples to identify genes that encode targets of partial protective immunity in adult volunteers vaccinated with PfSPZ Vaccine in Burkina Faso and PfSPZ-CVac in Mali

2 MALARIA DIAGNOSTIC TEST PERFORMANCE AND EVASION DETECTION

This chapter contains the following publication:

Mpina M, **Stabler TC**, Schindler T, Raso J, Deal A, Acuche Pupu L, Nyakarungu E, del Carmen Ovono Davis M, Urbano V, Mtoro A, Hamad A. **Diagnostic performance and comparison of ultrasensitive and conventional rapid diagnostic test, thick blood smear and quantitative PCR for detection of low-density Plasmodium falciparum infections during a controlled human malaria infection study in Equatorial Guinea**. *Malaria Journal*. 2022 Mar 24;21(1):99.

Stabler TC, Dwivedi A, Shrestha B, Joshi S, Schindler T, Ouattara A, García GA, Daubenberger C, Silva JC. **Gene Coverage Count and Classification (GC3): a locus sequence coverage assessment tool using short-read whole genome sequencing data, and its application to identify and classify histidine-rich protein 2 and 3 deletions in Plasmodium falciparum**. *Malaria Journal*. 2022 Nov 29;21(1):357.

2.1 Diagnostic performance and comparison of ultrasensitive and conventional rapid diagnostic test, thick blood smear and quantitative PCR for detection of low-density Plasmodium falciparum infections during a controlled human malaria infection study in Equatorial Guinea.

Published in Malaria Journal, 2022

Malaria Journal

RESEARCH Open Access

Diagnostic performance and comparison of ultrasensitive and conventional rapid diagnostic test, thick blood smear and quantitative PCR for detection of low-density *Plasmodium falciparum* infections during a controlled human malaria infection study in Equatorial Guinea

Maxmillian Mpina[1,2,3*], Thomas C. Stabler[1,2], Tobias Schindler[1,2], Jose Raso[4,5], Anna Deal[1,2], Ludmila Acuche Pupu[4], Elizabeth Nyakarungu[3,4], Maria del Carmen Ovono Davis[4], Vicente Urbano[4,5], Ali Mtoro[3,4], Ali Hamad[3,4], Maria Silvia A. Lopez[4,5], Beltran Pasialo[4,5], Marta Alene Owono Eyang[4,5], Matilde Riloha Rivas[5], Carlos Cortes Falla[4], Guillermo A. García[6], Juan Carlos Momo[4,5], Raul Chuquiyauri[4,7], Elizabeth Saverino[7], L. W. Preston Church[7], B. Kim lee Sim[7], Bonifacio Manguire[8], Marcel Tanner[1,2], Carl Maas[8], Salim Abdulla[3], Peter F. Billingsley[7], Stephen L. Hoffman[7], Said Jongo[3,4], Thomas L. Richie[7†] and Claudia A. Daubenberger[1,2*†]

Abstract

Background: Progress towards malaria elimination has stagnated, partly because infections persisting at low parasite densities comprise a large reservoir contributing to ongoing malaria transmission and are difficult to detect. This study compared the performance of an ultrasensitive rapid diagnostic test (uRDT) designed to detect low density infections to a conventional RDT (cRDT), expert microscopy using Giemsa-stained thick blood smears (TBS), and quantitative polymerase chain reaction (qPCR) during a controlled human malaria infection (CHMI) study conducted in malaria exposed adults (NCT03590340).

Methods: Blood samples were collected from healthy Equatoguineans aged 18–35 years beginning on day 8 after CHMI with 3.2×10^3 cryopreserved, infectious *Plasmodium falciparum* sporozoites (PfSPZ Challenge, strain NF54) administered by direct venous inoculation. qPCR (18s ribosomal DNA), uRDT (Alere™ Malaria Ag P.f.), cRDT [Carestart

*Correspondence: mmpina@ihi.or.tz; claudia.daubenberger@swisstph.ch
†Thomas L. Richie and Claudia A. Daubenberger contributed equally to this work
[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
Full list of author information is available at the end of the article

Malaria Pf/PAN (PfHRP2/pLDH)], and TBS were performed daily until the volunteer became TBS positive and treatment was administered. qPCR was the reference for the presence of *Plasmodium falciparum* parasites.

Results: 279 samples were collected from 24 participants; 123 were positive by qPCR. TBS detected 24/123 (19.5% sensitivity [95% CI 13.1–27.8%]), uRDT 21/123 (17.1% sensitivity [95% CI 11.1–25.1%]), cRDT 10/123 (8.1% sensitivity [95% CI 4.2–14.8%]); all were 100% specific and did not detect any positive samples not detected by qPCR. TBS and uRDT were more sensitive than cRDT (TBS vs. cRDT p = 0.015; uRDT vs. cRDT p = 0.053), detecting parasitaemias as low as 3.7 parasites/μL (p/μL) (TBS and uRDT) compared to 5.6 p/μL (cRDT) based on TBS density measurements. TBS, uRDT and cRDT did not detect any of the 70/123 samples positive by qPCR below 5.86 p/μL, the qPCR density corresponding to 3.7 p/μL by TBS. The median prepatent periods in days (ranges) were 14.5 (10–20), 18.0 (15–28), 18.0 (15–20) and 18.0 (16–24) for qPCR, TBS, uRDT and cRDT, respectively; qPCR detected parasitaemia significantly earlier (3.5 days) than the other tests.

Conclusions: TBS and uRDT had similar sensitivities, both were more sensitive than cRDT, and neither matched qPCR for detecting low density parasitaemia. uRDT could be considered an alternative to TBS in selected applications, such as CHMI or field diagnosis, where qualitative, dichotomous results for malaria infection might be sufficient.

Keywords: Malaria, Rapid diagnostic test, Controlled human malaria infection, Thick blood smear, Low parasite density infections, Malaria pre-exposure

Background

Over the past decade, malaria treatment and vector control interventions have significantly decreased malaria burden worldwide. The global incidence rate of malaria decreased between 2010 and 2018 from 71 to 57 cases per 1000 people at risk [1]. However, during 2020, the first full year of the COVID-19 pandemic, the World Health Organization (WHO) World Malaria Report documented an increment of 14 million annual malaria cases and 69,000 additional deaths compared to 2019, much of this attributable to COVID-19-related interruption of malaria control and medical services [2]. A significant challenge faced by malaria control and elimination projects is addressing transmission potential from low parasite density carriers with mild or no symptoms. Low parasite density carriers are estimated to account for 20–50% of human-to-mosquito transmission [3]. It would be beneficial to have a rapid test able to identify these infections.

Current widely deployed diagnostic tools such as conventional rapid diagnostic tests (cRDTs) are affordable and have user-friendly formats and function. In 2018, 259 million cRDTs were distributed mainly in sub-Saharan Africa and utilized to examine suspected malaria cases [1]. cRDTs test for the presence of histidine-rich protein 2 (PfHRP2), an antigen specific to *Plasmodium falciparum*, and many iterations of the test also include a pan-malaria antigen (PAN) common to all 4 major malaria species such as lactate dehydrogenase (LDH). While cRDTs are affordable, provide quick and readable results, and require little training to operate, they cannot detect low-density infections, which can occur frequently in low transmission areas [3–5]. As a result, efforts are underway to develop advanced malaria rapid diagnostic tests that are more sensitive and effective at identifying low-density *P. falciparum* infections [6, 7].

Laboratory-based techniques such as thick blood smear (TBS) and polymerase chain reaction (PCR) are considered to have greater sensitivity than cRDTs [8]. Historically, TBS has been the gold standard for malaria diagnosis. When performed by expert microscopists reading 0.5 μL of blood, detection may range typically between 10 and 50 parasites per microlitre (p/μL), and under carefully controlled CHMI and TBS preparation conditions, expert microscopists can quantify parasite densities at the theoretical lower limit of detection for this blood volume, 2 p/μL. In contrast, cRDTs have reported detection limits of 100–200 p/μL of blood in field studies [9–11]. TBS has the additional advantages of diagnosing infections even in the presence of parasites carrying PfHRP2 deletions, a challenge that RDT manufacturers are currently facing [12, 13], and allowing detection of all species of malaria parasite. In recent years, the further development of the quantitative polymerase chain reaction (qPCR) method has enabled the detection of low parasite densities that frequently go undetected by cRDTs and TBS [14, 15]. Utilizing qPCR, reservoirs of low parasite density cases can be identified and treated, which is an essential component for elimination in low transmission areas [8, 16–18]. The drawback of both TBS and qPCR is the need for specialized laboratory equipment, materials and well-trained staff, which are often in short supply in low income countries, making them unfeasible for use under field conditions on a large scale [19, 20]. TBS can also lead to false negative results when performed by a non-competent microscopist, the parasitaemia is low or there are multiple-species co-infections [21, 22]. An ideal diagnostic tool would

combine the sensitivity of qPCR with the affordability and simplicity of the cRDT.

An ultrasensitive RDT (uRDT; Alere™ Malaria Ag P.f.) has recently been developed and offered as a reliable diagnostic tool that can be used in clinical studies and in field operations [23–25]. Similar to conventional RTDs [cRDTs, such as Carestart Malaria Pf/PAN (PfHRP2/pLDH)] in form and function, the uRDT detects PfHRP2 in *P. falciparum*, but at greater sensitivity, identifying densities as low as 0.1–1.0 p/μL in culture-derived samples [7, 23]. This more sensitive RDT, which has been described as being capable of detecting low parasite density infections, could target these low-density infections for treatment [26]. If the sensitivity reported is confirmed, it could also be used in clinical trials of antimalarial drugs or vaccines to document protection following CHMI or natural *P. falciparum* exposure [27–29]. To date, limited literature exists systematically examining the performance of the uRDT in direct comparison to cRDTs, TBS and qPCR in samples collected from individuals with low parasite density [30, 31]. Existing investigations have been conducted predominantly in the field where it is difficult to monitor factors such as timings of infectious mosquito bites and the waxing and waning of parasite densities [24, 31–35].

A new approach to conducting controlled human malaria infection (CHMI) using *P. falciparum* sporozoites (PfSPZ) has become available over the past five years, based on administering aseptic, purified, cryopreserved, infectious PfSPZ (Sanaria® PfSPZ Challenge) [36]. In this model, PfSPZ are administered by syringe, replacing mosquito bite administration, and the study subjects are then monitored in the standard way [37]. The advantages are that CHMI can be performed by institutions without an insectary or without the need to import infected mosquitoes, the dose of PfSPZ can be standardized, and CHMI can be administered at any time without coordinated mosquito infections. The use of cryopreserved infectious PfSPZ is thus similar to the use of cryostabilates for induced blood stage malaria (IBSM), a controlled human infection model that bypasses the SPZ and liver stages and is similarly free of constraints [38]. CHMI using PfSPZ Challenge is now being utilized extensively to evaluate efficacy of anti-malarial drugs and vaccines in malaria-naive and malaria pre-exposed populations [39–45].

CHMI using injectable PfSPZ provides an opportunity to assess malaria diagnostics under carefully controlled conditions in malaria exposed populations. The exact exposure time is known, the induced infections gradually increase in density, and the time of first detection and associated prepatent period can be precisely determined. The aim of this study was to systematically evaluate and compare uRDT (Alere™ Malaria Ag P.f.) performance against three other commonly used malaria diagnostic tools using whole blood samples collected daily from malaria pre-exposed individuals undergoing CHMI.

Methods

Study site

The Bioko Island Malaria Elimination Program (BIMEP) focuses on developing malaria vaccines and other interventions to decrease malaria-attributable morbidity and mortality on Bioko Island, Equatorial Guinea [46]. BIMEP performs various activities on Bioko such as in-depth epidemiological studies and clinical trials of the malaria vaccine candidates Sanaria® PfSPZ Vaccine and PfSPZ-CVac [44, 47, 48] to collect data on safety and efficacy to support vaccine licensure. In 2018, the BIMEP conducted a regimen optimization trial of PfSPZ Vaccine (ClinicalTrials.gov ID: NCT03590340) at the Baney Research Facility that involved 104 healthy Equatoguinean adults male and female, aged 18–35 years [49]. Study volunteers were recruited from Baney district and the city of Malabo, and were enrolled after providing informed consent. Homologous (the strain of *P. falciparum* in PfSPZ Challenge, PfNF54, was the same as in the vaccine) CHMI with 3.2×10^3 PfSPZ was administered to 95 eligible individuals, 6 to 7 weeks after last vaccination and the observation period was conducted in a hotel at the La Paz Hospital beginning eight days after PfSPZ Challenge injection.

Malaria rapid diagnostic tests

Malaria rapid diagnostic tests were performed with frozen venous whole blood samples anticoagulated with EDTA and stored at − 80 °C. Blood samples were allowed to thaw slowly and equilibrate to ambient temperature for at least 30 min before performing the RDT. The commercially available cRDT [Carestart malaria Pf/PAN (PfHRP2/pLDH) Ag Combo, ACCESSBIO, USA (Lot# MR18F63, expiration: 30th Nov, 2020)] and uRDT [the Alere™ Malaria Ag Pf, Standard Diagnostic Inc., Republic of Korea (Lot# 05LDE001A, expiration: 20th Feb, 2020)] were used throughout the study. Both tests have similar manufacturer instruction workflows, requiring 5 μL blood to be applied to the RDT, followed by addition of the assay buffer and incubation for 20 min. After incubation, diagnosis was determined by the appearance of lines in the test result window. A single control line was considered malaria negative, a line for both control and test were considered malaria positive, and no lines, neither control nor test, was considered an invalid test. Two to three readers were involved in determining each RDT result, the third added when there were discrepancies between the first two readers. The samples were analysed

in batches of 10 by a first reader and the developed RDTs were then given to a second reader to confirm the reading of the first reader. Any discrepancy between first and second readers was resolved using a third reader within a period of 20 ± 1 min. Results were recorded in the database as binary figures with zero representing negatives and one representing positives. Five fresh known malaria positive and five fresh known negative samples were used for quality control for the performance of the assay and for training the technicians prior to conducting the study.

Thick blood smear for quantification of *Plasmodium falciparum*

Two millilitres of fresh whole venous blood from study participants were used for the preparation of TBS. The TBS was prepared by evenly spreading 10 µL of fresh whole venous blood into a 1 cm × 2 cm rectangle. The smears were air dried, stained for 45 min using 4% Giemsa stain and rinsed with buffered water, pH 7.2. The slides were dried and read using a light microscope with a high-power field (immersion oil, 100× objective) of 0.18 mm diameter. 6 × 1 cm passes equivalent to 0.54 µL of blood or 24 × 1 cm passes equivalent to 2.14 µL of blood for symptomatic volunteers were read before a TBS was declared positive or negative. The slides were read by two independent expert microscopists and any discrepancies were resolved by a third microscopist. For Giemsa staining quality control, known positive and negative thin blood smears were included at the beginning of the day and analysed for both parasites and cell staining colour and quality according to a standard operating procedure. Only Giemsa stain that passed the quality control procedures was allowed to be used for the slide staining on that day. Microscopes were maintained on a daily basis.

Quantification of *P. falciparum* parasite density by polymerase chain reaction (qPCR)

DNA was extracted directly from 180 µL of freshly collected venous whole blood using Quick-DNA Miniprep kits (Zymo Research, Irvine, USA) and eluted with 50 µL of elution buffer as recommended by manufacturer. DNA samples were kept at − 20 °C until analysis using the Bio-Rad CFX96 Real-Time PCR System (Bio-Rad Laboratories, California, USA). The PlasQ assay previously described by Schindler et al. [50] was used for quantification of *Plasmodium spp.* and *P. falciparum* parasites in the venous blood sample. This multiplex assay targets two independent *Plasmodium* genes namely the Pan-*Plasmodium* 18 S rDNA sequence (Pspp18S) and the *P. falciparum*-specific acidic terminal sequence of the *var* genes (PfvarATS). The human Ribonuclease P gene (HsRNaseP) was used as a DNA extraction and qPCR amplification control. All qPCR assays were run in duplicate and both non-template control (molecular grade nuclease-free water) and *P. falciparum* 3D7 DNA were included in each PCR run as negative and positive controls, respectively. For the parasite density estimation, a serial dilution was made according to the 1st WHO International Standard for *P. falciparum* DNA Amplification Technique (NIBSC code: 04/176) to establish a calibration curve with the parasite densities ranging between 0.01 and 10,000 p/µL. The actual parasite density of the tested sample by qPCR was then estimated from the calibration curve's y-intercept and slope. The lower limit of detection for this qPCR assay was 50 copies/mL. The sample was considered *P. falciparum* positive if each of the two replicates for both PfvarATS and 18 S RNA gene targets had quantitation cycles (Cq) < 40 and Cq < 28 for qPCR amplification control (HsRNaseP). In case of a discrepancy between duplicates, the assay was repeated, with at least two positive replicates out of four considered a positive result. The final results were used for the qPCR-based estimate of parasite density.

Controlled human malaria infection (CHMI)

From October 2018 to March 2019, 95 healthy Equatoguinean adults underwent CHMI [49]. Prior to CHMI, a full 3-day course of artemether/lumefantrine treatment was given to all volunteers. Eligibility criteria for CHMI were met if volunteers had received a complete regimen of PfSPZ Vaccine and were negative for malaria infection at the time of CHMI. During the ward observation period, volunteers were monitored daily for *P. falciparum* parasitaemia starting on day 8 to detect the parasite early and prevent the development of symptoms. Two millilitres of venous whole blood were collected in EDTA tubes daily on days 8–20 post-infection and transported to the laboratory in cooling boxes (4–8 °C) within 30 min of collection. One mL blood was used for examining malaria parasites positivity and density by TBS and qPCR within 4 h of collection and 1 mL was stored at − 80 °C for 8 months before retrospectively analysing samples using uRDT and cRDT. The standard artemether/lumefantrine treatment was given to subjects once malaria parasites were detected by TBS or on day 28 post CHMI for volunteers who remained negative throughout the post-CHMI follow-up period. Positive TBS results confirmed by qPCR were used as the end-point for initiating participant malaria treatment and termination of ward visits and further diagnostic sample collection. Volunteers diagnosed as malaria positive by TBS during 28 days of CHMI follow-up were considered eligible for participation in this malaria diagnostics study.

Sampling and statistical analyses

The aim of this study was to demonstrate differences in performance sensitivity between TBS, cRDT, and uRDT methods in detecting low density malaria infection, using qPCR as the gold standard for sensitivity. To obtain sufficient samples, Epi Info 7 software was used to calculate the sample size assuming the following parameters: minimum sensitivity of 90% for a reference method (qPCR) and 80% for cRDT, a 95% confidence and 90% power to detect a maximum sensitivity difference of 10%. cRDT was used to determine sample size due to its wider application in the field [51]. This gave a minimum required sample size of 267 samples. Sensitivity of the various tests was compared using Fisher's exact test.

Participating individuals were to be observed from day 8 after challenge until the day of first positivity by TBS. Tests were not performed on samples collected after TBS diagnosis since volunteers were treated and positive results could still occur for RDTs and qPCR due to residual parasite material, which would confound results. Out of the eligible volunteers, a subset of 24 individuals were to be randomly selected to meet sample size requirements while maintaining the distribution of parasite densities observed using qPCR. Considering the low sensitivity of most cRDTs at 100 p/µL, a stratified random sampling method was selected, whereby samples with parasite density > 100 p/µL and < 100 p/µL as detected by qPCR were put into two different strata. Using Microsoft Excel (2016), simple random sampling was performed within each stratum to obtain a total of 24 individuals.

To assess the distributions of the complete set, sampled and unsampled subsets were examined to ensure that they presented similar structure. Sampled individuals provided a total of 279 observed individual sample time points.

For this analysis, a multiplex qPCR targeting Pspp18S and PfvarATS was designated as the reference for detection of infection against which TBS, uRDT and cRDT positivity could be compared. A two-tailed Fishers exact test was used to determine the significant differences between the sensitivities of the various diagnostic tests. In this study, all samples were included in the sensitivity analysis of diagnostics. Only positive samples by either TBS, uRDT and cRDT were included in the analyses of the overall geometric mean (geomean) of parasite density of positive results and the geomean of parasite density at time of first detection (prepatent period). If a TBS, uRDT or cRDT test was negative, the respective sample point was deemed not applicable for the geomean parasite density and time to first detection analysis. TBS and qPCR both provided density measurements but TBS was considered to be more reliable as it did not involve conversion from gene copy number using a reference standard.

Results were recorded by trained and qualified laboratory staff on case report forms (CRFs) during the CHMI ward observation period and later entered onto an Excel spreadsheet (Microsoft, Office 2019 Ver 16). All samples were assigned a sample specific number that was linked to each volunteer ID. No personal information was recorded on laboratory CRFs and for the laboratory staff the connection of each sample with the corresponding donor volunteer was not possible. Retrospective RDT results were recorded on the same Excel spreadsheet. The geomean and geomean confidence intervals of parasite densities were calculated using R 4.0.1. Sensitivity and 95% confidence intervals for all diagnostic methods were calculated in R 4.0.1 using the epiR package [52].

Results

Overview

A total of 48 volunteers were diagnosed positive for malaria by reference method (qPCR), qualifying them for inclusion in the malaria diagnostic comparison. Individuals had an average of 12 time points of observation-days (range 8 to 17), with each day-test-record representing an independent observation since each was obtained from a newly collected whole blood sample. Out of the eligible volunteers, 24 (50%) individuals were randomly selected while maintaining the distribution of parasite densities observed using qPCR. The subset of selected samples was evaluated against unselected samples by parasite density distribution and variance and the selected subset was determined to be an appropriate representation (Additional file 1: Figs. S1–S3).

A total of 279 samples were collected from the 24 selected study participants; 123 and 156 samples were positive and negative for *P. falciparum* infection by qPCR, respectively. All 156 samples negative for *P. falciparum* by qPCR were also negative by TBS, uRDT and cRDT demonstrating 100% specificity for these tests. In total, 24 of 123 positive samples were detected by TBS, 21 by uRDT and 10 by cRDT, providing sensitivities of 19.5% (95% CI 13.1–27.8%), 17.1% (95% CI 11.1–25.1%), and 8.1% (95% CI 4.2–14.8%), respectively. qPCR detected more positives than any of the other tests ($p < 0.001$) and TBS and uRDT were both more sensitive than cRDT (TBS vs. cRDT, $p = 0.015$ by Fishers Exact two-tailed; uRDT vs. cRDT, $p = 0.053$). TBS detected 61.9% (13/21) of uRDT positive infections, while uRDT detected 54.2% (13/24) of TBS positive infections. The uRDT detected 100% (10/10) of cRDT positive infections while TBS detected 90% (9/10). The cRDT detected 47.6% (10/21) of uRDT positive infections and 37.5% (9/24) of TBS positive infections. The summary of the findings are depicted in a Venn-diagram (Fig. 1).

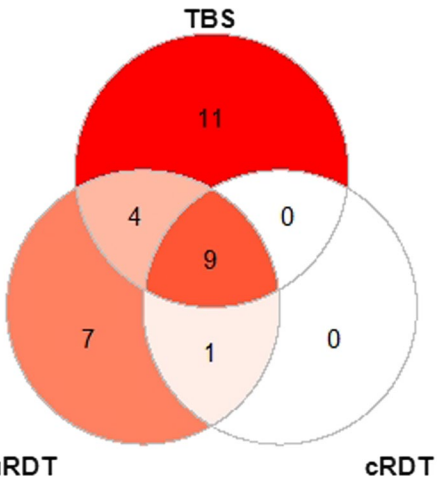


Fig. 1 Similarities and discrepancies in detection of *P. falciparum* cases by different diagnostic methods. Venn-diagram showing distribution of positive results according to the diagnostic test used with qPCR as the reference method. All samples were sorted by thick blood smear (TBS), ultrasensitive rapid diagnostic test (uRDT), and conventional rapid diagnostic test (cRDT). All cases were low parasite density *P. falciparum* infections that occurred during CHMI

Parasite density by qPCR was calculated using a standard curve generated using a WHO reference sample that related copy number to density, and ranged from 0.14 to 603.8 p/μL, with a geomean of 2.57 p/μL. Using this scale TBS detected parasites in the range of 5.9–603.8 p/μL (geomean 97.6 p/μL) and uRDT detected parasites in the range of 0.8–603.8 p/μL (geomean 103.4 p/μL), compared to cRDT, which detected parasites in the range of 2.6–603.8 p/μL (geomean = 149.4 p/μL) (Table 1).

Parasite density by TBS ranged from 3.7 to 201.8 p/μL (geomean 12.81 p/μL). Using this scale, uRDT detected parasites in the range of 3.7–201.8 p/μL (geomean 20.6 p/μL), compared to cRDT, which detected parasites in the range of 5.6–201.8 p/μL (geomean = 30.2 p/μL) (Table 2).

Examining just infections positive by both qPCR and TBS, the geomean ratio established from parasite density of qPCR and TBS (qPCR/TBS) was 7.62 p/μL (Table 3). PCR detects gene copy number in a specimen and each *P. falciparum* genome has at least 5–8 copies [53], and in addition qPCR can detect free DNA in a specimens. These variables make it difficult to calculate parasite density accurately using qPCR, even when using a standard curve to convert copy numbers to density based on a WHO reference sample. TBS detects parasites, regardless of how many genes and nuclei are present, but is hindered by the possible loss of significant numbers of parasites during processing [54]. All these factors were suspected to have contributed to the higher densities found using qPCR compared to TBS.

Table 1 Overall geomean and sensitivity of TBS, uRDT, and cRDT compared to the qPCR method

Diagnostic test	TBS (+)	TBS (−)	uRDT (+)	uRDT (−)	cRDT (+)	cRDT (−)	Total
PCR Pf (+)	24	99	21	102	10	113	123
PCR Pf (−)	0	156	0	156	0	156	156
Total	24	255	21	258	10	269	279
Range of positive samples (p/μL by qPCR)	5.9–603.8		0.8–603.8		2.6–603.8		
Geomean of positive samples (p/μL by qPCR)	97.6		103.4		149.4		
Sensitivity (%)	19.5% (13.1–27.8)		17.1% (11.1–25.1)		8.1% (4.2–14.8)		

Number of positive and negative samples, overall geomean and sensitivity of TBS, uRDT, and cRDT compared to the reference qPCR method. All cases were low parasite density infections that occurred during CHMI. Parasite densities of positive samples by qPCR ranged from 0.14–603.84 p/μL

Table 2 Overall geomean and sensitivity of uRDT and cRDT compared to the TBS method

Diagnostic test	uRDT (+)	uRDT (−)	cRDT (+)	cRDT (−)	Total
TBS (+)	13	11	9	15	24
TBS (−)	8	247	1	254	255
Total	21	258	10	269	279
Geomean of positive samples (p/μL by TBS)	20.6		30.2		
Sensitivity (%)	54.2% (33.2–73.8)		37.5% (19.6–59.2)		

Overall geomean, sensitivity and number of positive and negative samples of uRDT, and cRDT compared to TBS. All cases occurred during CHMI. Parasite densities of positive samples by TBS ranged from 3.7 to 201.8 p/μL

Table 3 Overall geomean, ranges and the ratio of parasite density established from qPCR and TBS

	qPCR (Pf/μL)	TBS (Pf/μL)	Ratio qPCR/TBS
GeoMean	97.57	12.81	7.62
Range	[5.86–603.84]	[3.70–201.80]	[0.07–53.84]

Comparison of geomean, ranges and the ratio of parasite density established from qPCR and TBS using paired samples in which both tests were positive. All cases were low parasite density infections that occurred during CHMI. The Ratio was determined by dividing the geomean of qPCR by geomean of TBS (qPCR/TBS). N = 24

Next, the diagnostic test sensitivities of TBS, uRDT and cRDT were stratified by ranges of parasite density of qPCR (Table 4). None of the three diagnostic tests detected *P. falciparum* infections below 1 p/μL as determined by qPCR. For TBS, the lowest parasite density detected as quantified by qPCR was 5.86 p/μL and as quantified by TBS was 3.7 p/μL. At parasite densities of 1–50 p/μL, uRDT and TBS appeared roughly equally sensitive and both appeared more sensitive than cRDT. At parasite densities of 51–100 p/μL by qPCR, uRDT appeared more sensitive (71% [95% CI 29–96%]) compared to TBS and cRDT with, 29% (4–71%) and 29% (4–71%) sensitivity, respectively. However, numbers were small and differences in sensitivity amongst the tests across the density categories were not statistically significant.

Finally, the uRDT and cRDT diagnostic test sensitivities were stratified by ranges of parasite density measured by TBS (Table 5). At parasite densities range between 1 and 50 p/μL, uRDT had higher sensitivity compared to cRDT; 33% (95% CI 12–62) for 1–10 p/μL and 100% (95% CI 48–100) at 11–50 p/μL compared to 13% (95% CI 2–40) for 1–10 p/μL and 80% (95% CI 28–99) at 11–50 p/μL respectively. Above 50 p/μL, both uRDT and cRDT had the same sensitivity but again numbers were too small to allow a meaningful comparison.

Finally, the range and distribution of parasite densities of samples determined to be positive by qPCR (n = 123), by TBS (n = 24), by uRDT (n = 21) and by cRDT (n = 10) were examined over the follow up period for the 24 volunteers who were TBS positive. TBS and uRDT recorded a trend for lower geomean parasite densities detected compared to cRDT, which did not reach statistical significance [p = 0.19 and p = 0.26, respectively] (Fig. 2).

Time to first detection

To investigate the efficiency of cRDT and uRDT to detect asexual blood stage parasites during CHMI follow-up, the median times to first detection of parasites (in days) by qPCR, TBS, uRDT and cRDT were compared. There

Table 4 Comparison of TBS, uRDT and cRDT sensitivity stratified by parasite density (p/μL) as determined by qPCR

Group density (p/μL)	# samples qPCR (+) (reference)	TBS (+)	TBS sensitivity (95% CI)	uRDT	uRDT sensitivity (95% CI)	cRDT	cRDT sensitivity (95% CI)
< 1	37	0	–	0	–	0	–
1–10	38	1	3% (0–14)	1	3% (0–14)	1	3% (0–14)
11–50	25	7	28% (12–49)	4	16% (5–36)	0	–
51–100	7	2	29% (4–71)	5	71% (29–96)	2	29% (4–71)
> 100	16	14	88% (62–98)	11	69% (41–89)	7	44% (20–70)

Number of samples and sensitivity of TBS, uRDT and cRDT stratified by parasite density (p/μL). All cases were low parasite density samples that occurred during CHMI follow-up and were 100% specific compared to qPCR

Table 5 Comparison of uRDT and cRDT sensitivity stratified by parasite density (p/μL) as determined by TBS

Group density (p/μL)	# samples TBS (+) (reference)	uRDT (+)	uRDT sensitivity (95% CI)	cRDT (+)	cRDT sensitivity (95% CI)
< 1	0	0	–	0	–
1–10	15	5	33% (12–62)	2	13% (2–40)
11–50	5	5	100% (48–100)	4	80% (28–99)
51–100	2	2	100% (16–100)	2	100% (16–100)
> 100	2	1	50% (1–99)	1	50% (1–99)

Number of samples and sensitivity of uRDT and cRDT stratified by parasite density (p/μL). Diagnostic methods are compared to TBS as reference and all cases were low parasite density samples that occurred during CHMI follow-up

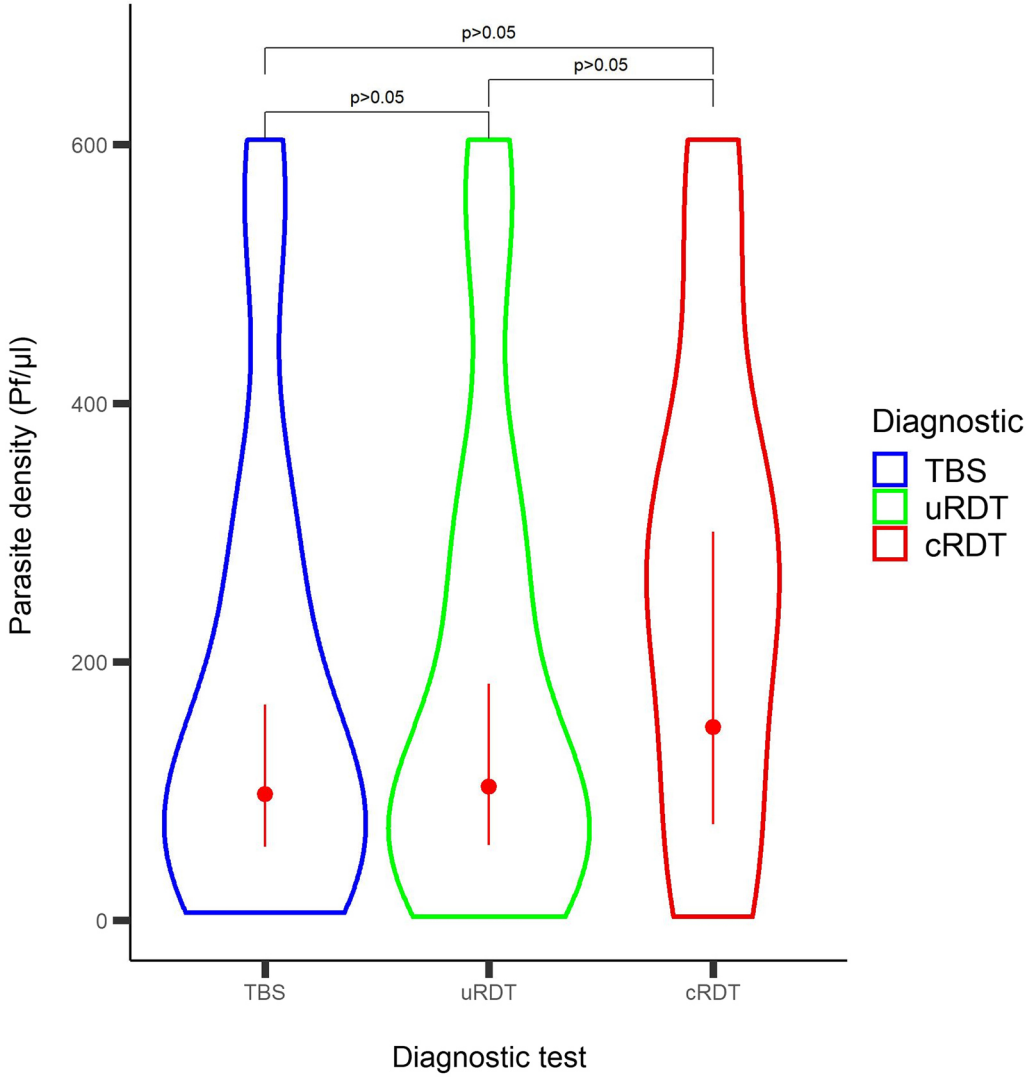


Fig. 2 Distribution of parasite density by qPCR of all malaria positive samples by TBS, uRDT, and cRDT. Red dots represent the geomean and error bars represent the 95% confidence intervals of each respective diagnostic test. Significance values were calculated using two-tailed Wilcoxon-test

was no evidence to support differences in prepatent period when using TBS, uRDT and cRDT since these methods all reported a median of 18.0 days to first parasite detection. The median days to detection of asexual blood stage parasitaemia by qPCR was 14.5, 3.5 days earlier than TBS, uRDT and cRDT ($p < 0.001$ log-rank test) (Fig. 3).

Discussion

As progress against malaria is made, asymptomatic infections at lower parasite densities become a significant challenge for malaria control and elimination efforts due to their contribution to ongoing transmission [3, 4, 55, 56]. Mass testing of a population with treatment of those found positive is one approach to address this problem [57]. However, it is difficult to diagnose low density parasitaemias and the most sensitive and, therefore, the best method, qPCR, is expensive, requires special laboratories and skilled personnel. Thus, the development of an inexpensive rapid test with equivalent sensitivity would be of great benefit, especially as conventional rapid diagnostic tests (cRDTs) are significantly less sensitive. Other applications could also benefit from a simple, rapid test that is more sensitive than cRDTs, such as detection of parasitaemia following sporozoite or blood stage CHMI, both important procedures for evaluating vaccine and drug efficacy [44, 45, 58–61] or for exploring innate and acquired immunity [62–66]. PfSPZ CHMI in particular is

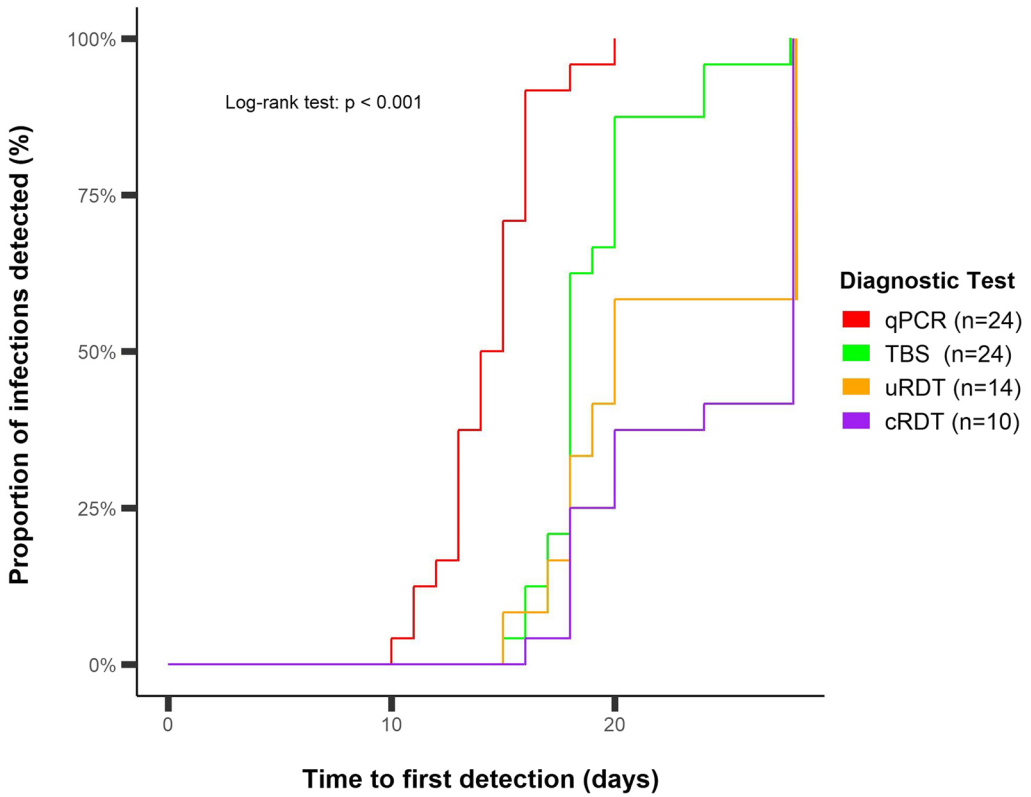


Fig. 3 Comparison of time to detection of parasites. Kaplan–Meier plot of the number of infections detected by time since CHMI qPCR (N = 24), TBS (N = 24), uRDT (N = 14) and cRDT (N = 10). All cases were low parasite density Pf infections that occurred during CHMI. P-value < 0.001 calculated using the log-rank test

now gaining attention by investigators, since it can now be used by any clinical centre without the need for infectious mosquitoes [67–69]. In CHMI, detection of low parasite densities is useful because it allows the identification and treatment of positive study subjects earlier in the course of their parasitaemia thereby preventing or ameliorating clinical manifestations. Therefore, on many fronts, there is a need to develop simpler, highly sensitive methods to diagnose low parasite densities that could augment the success of mass testing and treatment, promote epidemiological studies and simplify and lessen the costs associated with CHMI.

cRDTs have been a tremendous boon to diagnosing clinical malaria, where parasite densities are relatively high and the tests adequately sensitive. An uRDT has recently been developed, and might extend the usefulness of RDTs particularly in low to moderate transmission areas, in pre-elimination settings, and in experimental uses such as CHMI follow-up. For example, it has been reported that the uRDT is significantly more sensitive than cRDTs and TBS, detecting PfHRP2 at parasite densities as low as 0.1–1.0 p/μL in culture-derived samples [7, 23]. For this reason, the current study tested the uRDT and a cRDT during follow-up in a CHMI trial, and compared their sensitivities to those of TBS, using qPCR as the reference standard.

This study indeed found that both TBS and the uRDT were more sensitive than the cRDT. However, the data demonstrated that in samples from malaria-experienced subjects undergoing CHMI with parasites that contained PfHRP2, the sensitivity of uRDT (17.1%) was about threefold lower than that reported for pretreatment specimens from an IBSM study (47%) and in samples from a low transmission setting (44%), and fivefold lower than that reported in samples from a high transmission setting (84%) [23]. None of 37 specimens less than 1.0 p/μL by qPCR were identified. In 15 specimens that had 1–10 p/μL by TBS, uRDT identified 5 (33%), and cRDT 2 (13%). In 5 specimens with 11–50 parasites/μL by TBS, uRDT identified all 5 as positive (100%) and cRDT identified 4 (80%). Overall, uRDT and TBS gave similar results, and both tests were more sensitive than cRDTs in a setting of CHMI with PfHRP2-containing parasites.

Currently, CHMI requires highly trained clinical and laboratory staff including expert microscopists. Considering the comparable outcomes of TBS and uRDT

in this study, uRDT could be considered to replace TBS microscopy, especially in settings with inexperienced microscopists. However, parasite density estimation using qPCR has become a standard method utilized in many malaria studies [70, 71] and has been particularly useful in CHMI follow-up where it can detect parasitaemia earlier than TBS and allow treatment before signs and symptoms of clinical malaria develop [72–74]. Thus, uRDT would need to show advantages over TBS in the early identification of positive study subjects, as TBS itself is now being supplanted by qPCR. As expected, the study showed a significant difference in prepatent period amongst qPCR on the one hand and TBS, uRDT and cRDT on the other, confirming that qPCR is the most sensitive diagnostic method. The study further demonstrated that the median time to first malaria parasite detection by qPCR was 3.5 days earlier (14.5 days) compared to the other tests, and also that TBS, uRDT and cRDT were substantially equivalent to each other, each providing a prepatent period of 18 days. Similar to these findings, the CHMI studies conducted in semi-immune participants [44, 58, 73, 75] and in malaria naïve participants [76] have reported comparable prepatent periods using qPCR as the reference method. The fact that in this setting of progressively rising parasitaemias, the prepatent periods calculated by TBS, uRDT and cRDT were similar even though positive samples diagnosed by uRDT had a lower overall geomean of parasite density by qPCR than did TBS, suggests that uRDT may not have any particular advantage over TBS other than reduced costs and easier performance, or even over cRDT, as in this study the day of treatment would not have been affected had it been cRDT- rather than TBS-based. It would be expected that the same relative detection abilities would hold for blood stage CHMI, although this was not evaluated in this study, and Das et al. reported that uRDT detected parasitaemia 1.5 days earlier than cRDT in this setting [23].

In a field setting, the greater sensitivity of the uRDT over cRDTs could allow the detection of more asymptomatic carriers. This question was not directly examined in this CHMI-based study. However, the results showed that despite the uRDT being hailed as a significant improvement in malaria diagnostics, leading to increased sensitivity and specificity, satisfactory RDT performance for parasite density infections < 10 p/μL remains elusive. Field studies of mass testing and treatment are needed to further explore the potential contribution of the uRDT in identifying and treating asymptomatic carriers with low parasite densities contributing to ongoing transmission.

One important consideration for evaluating RDTs based on the detection of PfHRP2 is the increase in prevalence of parasites carrying PfHRP2 deletions, not only in Southeast Asia [77], but within the study area as well [12]. In 2018, approximately 65% of all suspected malaria cases in public health facilities in sub-Saharan African were tested with RDTs (~ 150 million cases) [1]. During CHMI, a standardized infectious PfSPZ dose of PfSPZ Challenge (NF54) was used, a parasite that expresses PfHRP2 to initiate the infection. Conducting a similar study in hospitals and field environment with important confounders, such as a deleted *Pfhrp2* gene, would likely have had different results.

Limitations

Anticoagulated (EDTA) fresh whole blood was used for to prepare samples for qPCR and TBS assessments. Anticoagulated (EDTA) cryopreserved (temperature of − 80 °C) whole blood held for 8 months and thawed was used to prepare samples for uRDT and cRDT. It is possible, but unlikely that HRP2 degraded during storage. Whole blood samples were temperature monitored during storage. When proper procedures are followed for long-term storage of whole blood, the quality of DNA, RNA or HRP2 is not compromised [7, 78]. Considering the strict temperature monitoring in this study and the fact that samples were only thawed once for processing, the difference in quality of samples over time is unlikely to have been different.

Another limitation is the discrepancy in parasite densities estimated by qPCR and TBS. qPCR may have overestimated parasite density due to variable numbers of copies of the amplification target and the persistence of nucleic acid from non-viable parasites [79], and TBS may have underestimated parasite density due to the loss of parasites during processing.

Because the research subjects were semi-immune, they may have had variable levels of anti-HRP2 or anti-LDH antibodies, which could have affected results [80, 81]. Some may also have had ongoing infections at the time of clearance with artemether/lumefantrine prior to CHMI. Although the current study did not measure the level of HRP2 in the participants before and after CHMI, none of the individuals were RDT positive between days 8 and 14 after CHMI. Therefore, it is likely that this factor did not affect the current performance comparison.

This study was designed to evaluate the performance of malaria diagnostic tests in independent samples and was analysed accordingly, even though several samples were collected from each individual post CHMI. This was based on the reasoning that, since each test was measuring a different parameter (DNA for qPCR, whole parasites for TBS, HRP2 for uRDT, HRP2/LDH for cRDT) and these parameters would vary independently from

day to day due to the presence of multiple clones of NF54 parasites released from individual hepatocytes over several days each with its own asynchronous 48 h reproductive/sequestration cycle, it would be difficult to propose a biological metric characterizing an individual that could introduce bias or similarities in observations. Nevertheless, such a bias or similarities could exist and might have affected the data.

Conclusions

TBS has been the classical approach to malaria diagnosis for clinical use, malaria control programs, research studies such as CHMI and field epidemiology. TBS can distinguish the five malaria species that infect humans, which cannot yet be achieved by using RDTs or a single reaction qPCR, and provides a reasonable estimate of parasite density. TBS, however, requires laboratories that support and maintain microscopes, staining solutions and human resources with the requisite microscopy skills. qPCR, with much greater sensitivity, is now supplanting TBS for many applications such as detection of parasitaemia following CHMI, but also requires a high level of laboratory capability and involves higher costs than TBS. cRDTs have, therefore, been a welcome addition to malaria diagnostics and in many places have supplanted TBS for the clinical diagnosis of malaria, where parasite densities are high, but have not been useful for applications requiring greater sensitivity. This study compared qPCR, TBS, a cRDT to a new uRDT advertised as rivaling qPCR in sensitivity, to assess its value for detection of parasitaemia following PfSPZ CHMI, an application where early diagnosis and treatment is important to reduce the severity of adverse events. The major conclusions were that for this specific application, while the uRDT was better than the cRDT, and approached TBS in sensitivity, it did not close the gap with respect to qPCR, and thus could be considered for replacing TBS only in studies unable to use qPCR or TBS due to resource limitations. The added value of the uRDT in field studies, particularly in mass testing and treatment, requires further study.

Abbreviations

cRDT: Conventional rapid diagnostic test; PfHRP2: Histidine-rich protein 2; PAN: Pan-malaria antigen; TBS: Thick blood smear; p/µL: Parasites per microlitre; qPCR: Quantitative polymerase chain reaction; uRDT: Ultra-sensitive rapid diagnostic test; CHMI: Controlled human malaria infection; DVI: Direct venous inoculation; BIMEP: Bioko Island malaria Elimination Project; CRFs: Clinical report forms.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12936-022-04103-y.

Additional file 1: Figure S1. Trend ofparasite density over time between individuals in selected and unselected groups.Trend of parasite density over time between selected and unselectedparticipants. Parasite density was determined by quantitative polymerase chainreaction assays (qPCR). **Figure S2.** Overall distribution of parasite densityamong groups of selected and unselected participants. Scatter plots of parasitedensity measured by quantitative polymerase reaction assays (qPCR) betweenselected and unselected data points. **Figure S3.** Distribution of parasitedensity by individuals participants in selected and unselected groups. Barplotsof parasite density as measured by quantitative polymerase reaction assays(qPCR) between selected vs. unselected individual volunteers.

Acknowledgements

This work was supported by the National Malaria Control Program and the Ministry of Health and Social Welfare of Equatorial Guinea, and Medical Care Development International (MCDI) through the Bioko Island Malaria Elimination Project (BIMEP). In particular, we thank the entire Equatorial Guinea Malaria Vaccine Initiative (EGMVI) team for their efforts while conducting the clinical trial and CHMI. We would especially like to thank Marathon Oil, Noble Energy, AMPCO (Atlantic Methanol Production Company), and the Ministry of Mines and Energy of Equatorial Guinea for their continued funding and support for malaria control on Bioko Island.

Authors' contributions

SLH posed the research question and the applicability of the CHMI follow-up as a means to evaluate a potential uRDT application. MM and CAD conceived and designed the evaluation plan for this study. MM, JR, AD, LA, EN, MdCOD, TS, VU, AM, AH, MSAL, BP, and MAOE were responsible for collecting and processing volunteer samples. MRR, CCF, GAG, RC, ES, PFB, LWPC, BKLS, TLR, BM, MT, SA, CM, SLH, SJ sponsored and oversaw the management of the PfSPZ Vaccine trial and CHMI. SJ was the principal investigator. CAD managed all laboratory activities and supervised this study. JCM managed quality control and quality assurance of all documentation. TCS programmed and conducted the statistical analysis of the data and generated all tables and figures. MM and TCS wrote the manuscript in consultation with PFB, SLH and TLR. TLR and CAD contributed equally on this work. All authors read and approved the final manuscript.

Funding

This work was funded by the Bioko Island Malaria Elimination Program. The funders had no role in study design, data collection, data analysis, data interpretation, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets used and/or analysis during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Ethics approval was obtained to conduct the malaria vaccine trial from the Comité Ética Nacional de Guinea Ecuatorial (CENGE) in Equatorial Guinea, Ifakara Health Institute (IHI) in Tanzania, Ethikkommission Nordwest- und Zentralschweiz (EKNZ) in Switzerland, and Prime IRB in the USA. Included in this approval was the use of participant biological samples for future use if informed consent was provided. No identifying information was accessible to laboratory staff and was not included in any database or analysis for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland. [2]University of Basel, Basel, Switzerland. [3]Ifakara Health Institute, Ifakara, Tanzania. [4]Medical Care Development International, Malabo, Equatorial Guinea. [5]Equatorial

Guinea Ministry of Health and Social Welfare, Malabo, Equatorial Guinea. [6]Medical Care Development International, Silver Spring, MD, USA. [7]Sanaria Inc., 9800 Medical Center Drive, Rockville, MD 20850, USA. [8]Marathon EG production Ltd., Houston, USA.

Received: 8 November 2021 Accepted: 23 February 2022
Published online: 24 March 2022

References

1. WHO. World malaria report 2020. Geneva: World Health Organization; 2020.
2. WHO. World malaria report 2021. Geneva: World Health Organization; 2021.
3. Slater HC, Ross A, Ouédraogo AL, White LJ, Nguon C, Walker PGT, et al. Assessing the impact of next-generation rapid diagnostic tests on *Plasmodium falciparum* malaria elimination strategies. Nature. 2015;528:94–101.
4. Hemingway J, Shretta R, Wells TN, Bell D, Djimdé AA, Achee N, et al. Tools and strategies for malaria control and elimination: what do we need to achieve a grand convergence in malaria? PLoS Biol. 2016;14:e1002380.
5. Lo E, Zhou G, Oo W, Afrane Y, Githeko A, Yan G. Low parasitemia in submicroscopic infections significantly impacts malaria diagnostic sensitivity in the highlands of western Kenya. PLoS ONE. 2015;10:e0121763.
6. Danwang C, Kirakoya-Samadoulougou F, Samadoulougou S. Assessing field performance of ultrasensitive rapid diagnostic tests for malaria: a systematic review and meta-analysis. Malar J. 2021;20:245.
7. Das S, Peck RB, Barney R, Jang IK, Kahn M, Zhu M, et al. Performance of an ultra-sensitive *Plasmodium falciparum* HRP2-based rapid diagnostic test with recombinant HRP2, culture parasites, and archived whole blood samples. Malar J. 2018;17:118.
8. Mathison BA, Pritt BS. Update on malaria diagnostics and test utilization. J Clin Microbiol. 2017;55:2009–17.
9. Adams M, Joshi SN, Mbambo G, Mu AZ, Roemmich SM, Shrestha B, et al. An ultrasensitive reverse transcription polymerase chain reaction assay to detect asymptomatic low-density *Plasmodium falciparum* and *Plasmodium vivax* infections in small volume blood samples. Malar J. 2015;14:520.
10. Bousema T, Okell L, Felger I, Drakeley C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. Nat Rev Microbiol. 2014;12:833–40.
11. Laban NM, Kobayashi T, Hamapumbu H, Sullivan D, Mharakurwa S, Thuma PE, et al. Comparison of a PfHRP2-based rapid diagnostic test and PCR for malaria in a low prevalence setting in rural southern Zambia: implications for elimination. Malar J. 2015;14:25.
12. Berzosa P, González V, Taravillo L, Mayor A, Romay-Barja M, García L, et al. First evidence of the deletion in the pfhrp2 and pfhrp3 genes in *Plasmodium falciparum* from Equatorial Guinea. Malar J. 2020;19:99.
13. Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. Am J Trop Med Hyg. 2012;86:194–8.
14. Berzosa P, de Lucio A, Romay-Barja M, Herrador Z, González V, García L, et al. Comparison of three diagnostic methods (microscopy, RDT, and PCR) for the detection of malaria parasites in representative samples from Equatorial Guinea. Malar J. 2018;17:333.
15. Hofmann NE, Gruenberg M, Nate E, Ura A, Rodriguez-Rodriguez D, Salib M, et al. Assessment of ultra-sensitive malaria diagnosis versus standard molecular diagnostics for malaria elimination: an in-depth molecular community cross-sectional study. Lancet Infect Dis. 2018;18:1108–16.
16. Amaral LC, Robortella DR, Guimarães LFF, Limongi JE, Fontes CJF, Pereira DB, et al. Ribosomal and non-ribosomal PCR targets for the detection of low-density and mixed malaria infections. Malar J. 2019;18:154.
17. Canier L, Khim N, Kim S, Sluydts V, Heng S, Dourng D, et al. An innovative tool for moving malaria PCR detection of parasite reservoir into the field. Malar J. 2013;12:405.
18. Naeem MA, Ahmed S, Khan SA. Detection of asymptomatic carriers of malaria in Kohat district of Pakistan. Malar J. 2018;17:44.
19. Mabey D, Peeling RW, Ustianowski A, Perkins MD. Diagnostics for the developing world. Nat Rev Microbiol. 2004;2:231–40.
20. Urdea M, Penny LA, Olmsted SS, Giovanni MY, Kaspar P, Shepherd A, et al. Requirements for high impact diagnostics in the developing world. Nature. 2006;444:73–9.
21. Ohrt C, Sutamihardja MA, Tang D, Kain KC. Impact of microscopy error on estimates of protective efficacy in malaria-prevention trials. J Infect Dis. 2002;186:540–6.
22. Payne D. Use and limitations of light microscopy for diagnosing malaria at the primary health care level. Bull World Health Organ. 1988;66:621–6.
23. Das S, Jang IK, Barney B, Peck R, Rek JC, Arinaitwe E, et al. Performance of a high-sensitivity rapid diagnostic test for *Plasmodium falciparum* malaria in asymptomatic individuals from Uganda and Myanmar and naive human challenge infections. Am J Trop Med Hyg. 2017;97:1540–50.
24. Landier J, Haohankhunnatham W, Das S, Konghahong K, Christensen P, Raksuansak J, et al. Operational performance of a *Plasmodium falciparum* ultrasensitive rapid diagnostic test for detection of asymptomatic infections in eastern Myanmar. J Clin Microbiol. 2018;56:e00565-18.
25. Yeung S, McGregor D, James N, Kheang ST, Kim S, Khim N, et al. Performance of ultrasensitive rapid diagnostic tests for detecting asymptomatic *Plasmodium falciparum*. Am J Trop Med Hyg. 2020;102:307–9.
26. Acquah FK, Donu D, Obboh EK, Bredu D, Mawuli B, Amponsah JA, et al. Diagnostic performance of an ultrasensitive HRP2-based malaria rapid diagnostic test kit used in surveys of afebrile people living in southern Ghana. Malar J. 2021;20:125.
27. Manjurano A, Omolo JJ, Lyimo E, Miyaye D, Kishamawe C, Matemba LE, et al. Performance evaluation of the highly sensitive histidine-rich protein 2 rapid test for *Plasmodium falciparum* malaria in North-West Tanzania. Malar J. 2021;20:58.
28. McMorrow ML, Aidoo M, Kachur SP. Malaria rapid diagnostic tests in elimination settings—can they find the last parasite? Clin Microbiol Infect. 2011;17:1624–31.
29. Wu L, van den Hoogen LL, Slater H, Walker PGT, Ghani AC, Drakeley CJ, et al. Comparison of diagnostics for the detection of asymptomatic *Plasmodium falciparum* infections to inform control and elimination strategies. Nature. 2015;528:86–93.
30. Lin JT, Saunders DL, Meshnick SR. The role of submicroscopic parasitemia in malaria transmission: what is the evidence? Trends Parasitol. 2014;30:183–90.
31. Plucinski MM, Candrinho B, Dimene M, Colborn J, Lu A, Nace D, et al. Assessing performance of HRP2 antigen detection for malaria diagnosis in Mozambique. J Clin Microbiol. 2019;57:e00875-19.
32. Galatas B, Mayor A, Gupta H, Balanza N, Jang IK, Nhamussua L, et al. Field performance of ultrasensitive and conventional malaria rapid diagnostic tests in southern Mozambique. Malar J. 2020;19:451.
33. Unwin VT, Ahmed R, Noviyanti R, Puspitasari AM, Utami RAS, Trianty L, et al. Use of a highly-sensitive rapid diagnostic test to screen for malaria in pregnancy in Indonesia. Malar J. 2020;19:28.
34. Slater HC, Ross A, Felger I, Hofmann NE, Robinson L, Cook J, et al. The temporal dynamics and infectiousness of subpatent *Plasmodium falciparum* infections in relation to parasite density. Nat Commun. 2019;10:1433.
35. Felger I, Maire M, Bretscher MT, Falk N, Tiaden A, Sama W, et al. The dynamics of natural *Plasmodium falciparum* infections. PLoS ONE. 2012;7:e45542.
36. Mordmüller B, Supan C, Sim KL, Gómez-Pérez GP, Ospina Salazar CL, Held J, et al. Direct venous inoculation of *Plasmodium falciparum* sporozoites for controlled human malaria infection: a dose-finding trial in two centres. Malar J. 2015;14:117.
37. Laurens MB, Duncan CJ, Epstein JE, Hill AV, Komisar JL, Lyke KE, et al. A consultation on the optimization of controlled human malaria infection by mosquito bite for evaluation of candidate malaria vaccines. Vaccine. 2012;30:5302–4.
38. Engwerda CR, Minigo G, Amante FH, McCarthy JS. Experimentally induced blood stage malaria infection as a tool for clinical research. Trends Parasitol. 2012;28:515–21.
39. Epstein JE, Paolino KM, Richie TL, Sedegah M, Singer A, Ruben AJ, et al. Protection against *Plasmodium falciparum* malaria by PfSPZ vaccine. JCI Insight. 2017;2:e89154.
40. Lyke KE, Ishizuka AS, Berry AA, Chakravarty S, DeZure A, Enama ME, et al. Attenuated PfSPZ vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection. Proc Natl Acad Sci USA. 2017;114:2711–6.

41. Seder RA, Chang L-J, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, et al. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. Science. 2013;341:1359–65.
42. Ishizuka AS, Lyke KE, DeZure A, Berry AA, Richie TL, Mendoza FH, et al. Protection against malaria at 1 year and immune correlates following PfSPZ vaccination. Nat Med. 2016;22:614–23.
43. Mordmuller B, Surat G, Lagler H, Chakravarty S, Ishizuka AS, Lalremruata A, et al. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. Nature. 2017;542:445–9.
44. Jongo SA, Urbano V, Church LWP, Olotu A, Manock SR, Schindler T, et al. Immunogenicity and protective efficacy of radiation-attenuated and chemo-attenuated PfSPZ vaccines in Equatoguinean adults. Am J Trop Med Hyg. 2021;104:283–93.
45. Jongo SA, Church LWP, Mtoro AT, Schindler T, Chakravarty S, Ruben AJ, et al. Increase of dose associated with decrease in protection against controlled human malaria infection by PfSPZ vaccine in Tanzanian adults. Clin Infect Dis. 2020;71:2849–57.
46. Billingsley PF, Maas CD, Olotu A, Schwabe C, García GA, Rivas MR, et al. The Equatoguinean malaria vaccine initiative: from the launching of a clinical research platform to malaria elimination planning in Central West Africa. Am J Trop Med Hyg. 2020;103:947–54.
47. Nchama VU, Said AH, Mtoro A, Bidjimi GO, Owono MA, Maye ERM, et al. Incidence of *Plasmodium falciparum* malaria infection in 6-month to 45-year-olds on selected areas of Bioko Island, Equatorial Guinea. Malar J. 2021;20:322.
48. Olotu A, Urbano V, Hamad A, Eka M, Chemba M, Nyakarungu E, et al. Advancing global health through development and clinical trials partnerships: a randomized, placebo-controlled, double-blind assessment of safety, tolerability, and immunogenicity of PfSPZ vaccine for malaria in healthy Equatoguinean men. Am J Trop Med Hyg. 2018;98:308–18.
49. Jongo SA, Church LW, Nchama VU, Hamad A, Chuquiyauri R, Kassim KR, et al. Multi-dose priming regimens of PfSPZ vaccine: safety and efficacy against controlled human malaria infection in Equatoguinean adults. Am J Trop Med Hyg. 2022. https://doi.org/10.4269/ajtmh.21-0942.
50. Schindler T, Robaina T, Sax J, Bieri JR, Mpina M, Gondwe L, et al. Molecular monitoring of the diversity of human pathogenic malaria species in blood donations on Bioko Island, Equatorial Guinea. Malar J. 2019;18:9.
51. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193–204.
52. Stevenson M, Nunes T, Heuer C, Marshall J, Sanchez J, Thornton R, et al. Tools for the analysis of epidemiological data. 2020. https://mran.microsoft.com/snapshot/2016-05-28/web/packages/epiR/index.html.
53. Mercereau-Puijalon O, Barale J-C, Bischoff E. Three multigene families in *Plasmodium* parasites: facts and questions. Int J Parasitol. 2002;32:1323–44.
54. Bejon P, Andrews L, Hunt-Cooke A, Sanderson F, Gilbert SC, Hill AVS. Thick blood film examination for *Plasmodium falciparum* malaria has reduced sensitivity and underestimates parasite density. Malar J. 2006;5:104.
55. Hawkins K, Burton R, LaBarre P. Diagnostics to support malaria elimination: choosing an appropriate biomarker to target the subclinical *Plasmodium falciparum* transmission reservoir. In: IEEE global humanitarian technology conference (GHTC 2014); 10–13 Oct. 2014;2014. p. 561–8.
56. Mosha JF, Sturrock HJW, Greenhouse B, Greenwood B, Sutherland CJ, Gadalla N, et al. Epidemiology of subpatent *Plasmodium falciparum* infection. implications for detection of hotspots with imperfect diagnostics. Malar J. 2013;12:221.
57. Okell LC, Griffin JT, Kleinschmidt I, Hollingsworth TD, Churcher TS, White MJ, et al. The potential contribution of mass treatment to the control of *Plasmodium falciparum* malaria. PLoS ONE. 2011;6:e20179.
58. Jongo SA, Shekalaghe SA, Church LWP, Ruben AJ, Schindler T, Zenklusen I, et al. Safety, immunogenicity, and protective efficacy against controlled human malaria infection of *Plasmodium falciparum* sporozoite vaccine in Tanzanian adults. Am J Trop Med Hyg. 2018;99:338–49.
59. Sulyok Z, Fendel R, Eder B, Lorenz F-R, Kc N, Karnahl M, et al. Heterologous protection against malaria by a simple chemoattenuated PfSPZ vaccine regimen in a randomized trial. Nat Commun. 2021;12:2518.
60. Sulyok M, Rückle T, Roth A, Mürbeth RE, Chalon S, Kerr N, et al. DSM265 for *Plasmodium falciparum* chemoprophylaxis: a randomised, double blinded, phase 1 trial with controlled human malaria infection. Lancet Infect Dis. 2017;17:636–44.
61. Murphy SC, Duke ER, Shipman KJ, Jensen RL, Fong Y, Ferguson S, et al. A randomized trial evaluating the prophylactic activity of DSM265 against preerythrocytic *Plasmodium falciparum* infection during controlled human malarial infection by mosquito bites and direct venous inoculation. J Infect Dis. 2017;217:693–702.
62. Mpina M, Maurice NJ, Yajima M, Slichter CK, Miller HW, Dutta M, et al. Controlled human malaria infection leads to long-lasting changes in innate and innate-like lymphocyte populations. J Immunol. 2017;199:107–18.
63. Kapulu MC, Njuguna P, Hamaluba M, Kimani D, Ngoi JM, Musembi J, et al. Naturally acquired immunity among Kenyan adults suppresses the West African *P. falciparum* NF54 strain in controlled human malaria infection (CHMI). medRxiv. 2020. https://doi.org/10.1101/2020.08.11.20172411.
64. Lell B, Mordmüller B, Dejon Agobe J-C, Honkpehedji J, Zinsou J, Mengue JB, et al. Impact of sickle cell trait and naturally acquired immunity on uncomplicated malaria after controlled human malaria infection in adults in Gabon. Am J Trop Med Hyg. 2018;98:508–15.
65. Achan J, Reuling IJ, Yap XZ, Dabira E, Ahmad A, Cox M, et al. Serologic markers of previous malaria exposure and functional antibodies inhibiting parasite growth are associated with parasite kinetics following a *Plasmodium falciparum* controlled human infection. Clin Infect Dis. 2019;70:2544–52.
66. Kapulu MC, Njuguna P, Hamaluba M, Kimani D, Ngoi JM, Musembi J, et al. Safety and PCR monitoring in 161 semi-immune Kenyan adults following controlled human malaria infection. JCI Insight. 2021;6:e146443.
67. Roestenberg M, Bijker EM, Sim BKL, Billingsley PF, James ER, Bastiaens GJH, et al. Controlled human malaria infections by intradermal injection of cryopreserved *Plasmodium falciparum* sporozoites. Am J Trop Med Hyg. 2013;88:5–13.
68. Laurens MB, Berry AA, Travassos MA, Strauss K, Adams M, Shrestha B, et al. Dose-dependent infectivity of aseptic, purified, cryopreserved *Plasmodium falciparum* 7G8 sporozoites in malaria-naive adults. J Infect Dis. 2019;220:1962–6.
69. Richie TL, Billingsley PF, Sim BK, James ER, Chakravarty S, Epstein JE, et al. Progress with *Plasmodium falciparum* sporozoite (PfSPZ)-based malaria vaccines. Vaccine. 2015;33:7452–61.
70. Kamau E, Alemayehu S, Feghali KC, Saunders D, Ockenhouse CF. Multiplex qPCR for detection and absolute quantification of malaria. PLoS ONE. 2013;8:e71539.
71. Koepfli C, Nguitragool W, Hofmann NE, Robinson LJ, Ome-Kaius M, Sattabongkot J, et al. Sensitive and accurate quantification of human malaria parasites using droplet digital PCR (ddPCR). Sci Rep. 2016;6:39183.
72. Hodgson SH, Douglas AD, Edwards NJ, Kimani D, Elias SC, Chang M, et al. Increased sample volume and use of quantitative reverse-transcription PCR can improve prediction of liver-to-blood inoculum size in controlled human malaria infection studies. Malar J. 2015;14:33.
73. Walk J, Schats R, Langenberg MCC, Reuling IJ, Teelen K, Roestenberg M, et al. Diagnosis and treatment based on quantitative PCR after controlled human malaria infection. Malar J. 2016;15:398.
74. Hodgson SH, Juma E, Salim A, Magiri C, Kimani D, Njenga D, et al. Evaluating controlled human malaria infection in Kenyan adults with varying degrees of prior exposure to *Plasmodium falciparum* using sporozoites administered by intramuscular injection. Front Microbiol. 2014;5:686.
75. Rothen J, Murie C, Carnes J, Anupama A, Abdulla S, Chemba M, et al. Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period. PLoS ONE. 2018;13:e0199392.
76. Friedman-Klabanoff DJ, Laurens MB, Berry AA, Travassos MA, Adams M, Strauss KA, et al. The controlled human malaria infection experience at the University of Maryland. Am J Trop Med Hyg. 2019;100:556–65.
77. Verma AK, Bharti PK, Das A. HRP-2 deletion: a hole in the ship of malaria elimination. Lancet Infect Dis. 2018;18:826–7.
78. Färnert A, Arez AP, Correia AT, Björkman A, Snounou G, do Rosário V. Sampling and storage of blood and the detection of malaria parasites by polymerase chain reaction. Trans R Soc Trop Med Hyg. 1999;93:50–3.
79. Vafa Homann M, Emami SN, Yman V, Stenstrom C, Sonden K, Ramstrom H, et al. Detection of malaria parasites after treatment in travelers: a 12-months longitudinal study and statistical modelling analysis. EBioMedicine. 2017;25:66–72.
80. Poti KE, Sullivan DJ, Dondorp AM, Woodrow CJ. HRP2: transforming malaria diagnosis, but with caveats. Trends Parasitol. 2020;36:112–26.

81. Rogier E, Plucinski M, Lucchi N, Mace K, Chang M, Lemoine JF, et al. Bead-based immunoassay allows sub-picogram detection of histidine-rich protein 2 from *Plasmodium falciparum* and estimates reliability of malaria rapid diagnostic tests. PLoS ONE. 2017;12:e0172139.

Publisher's Note

2.2 Gene Coverage Count and Classification (GC_3): a locus sequence coverage assessment tool using short-read whole genome sequencing data, and its application to identify and classify histidine-rich protein 2 and 3 deletions in Plasmodium falciparum.

Published in Malaria Journal, 2022

Malaria Journal

METHODOLOGY

Open Access

Gene Coverage Count and Classification (GC_3): a locus sequence coverage assessment tool using short-read whole genome sequencing data, and its application to identify and classify histidine-rich protein 2 and 3 deletions in *Plasmodium falciparum*

Thomas C. Stabler[1,2*], Ankit Dwivedi[3], Biraj Shrestha[4], Sudhaunshu Joshi[4], Tobias Schindler[1,2], Amed Ouattara[4], Guillermo A. García[5], Claudia Daubenberger[1,2] and Joana C. Silva[3,6*]

Abstract

Background: The ability of malaria rapid diagnostic tests (RDTs) to effectively detect active infections is being compromised by the presence of malaria strains with genomic deletions at the *hrp2* and *hrp3* loci, encoding the antigens most commonly targeted in diagnostics for *Plasmodium falciparum* detection. The presence of such deletions can be determined in publically available *P. falciparum* whole genome sequencing (WGS) datasets. A computational approach was developed and validated, termed Gene Coverage Count and Classification (GC_3), to analyse genome-wide sequence coverage data and provide informative outputs to assess presence and coverage profile of a target locus in WGS data. GC_3 was applied to detect deletions at *hrp2* and *hrp3* (*hrp2/*3) and flanking genes in different geographic regions and across time points.

Methods: GC_3 uses Python and R scripts to extract locus read coverage metrics from mapped WGS data according to user-defined parameters and generates relevant tables and figures. GC_3 was tested using WGS data for laboratory reference strains with known *hrp2/3* genotypes, and its results compared to those of a *hrp2/3*-specific qPCR assay. Samples with at least 25% of coding region positions with zero coverage were classified as having a deletion. Publicly available sequence data was analysed and compared with published deletion frequency estimates.

Results: GC_3 results matched the expected coverage of known laboratory reference strains. Agreement between GC_3 and a *hrp2/3*-specific qPCR assay reported for 19/19 (100%) *hrp2* deletions and 18/19 (94.7%) *hrp3* deletions. Among Cambodian (n = 127) and Brazilian (n = 20) WGS datasets, which had not been previously analysed for *hrp2/3* deletions, GC_3 identified *hrp2* deletions in three and four samples, and *hrp3* deletions in 10 and 15 samples, respectively.

*Correspondence: thomas.stabler@swisstph.ch; jcsilva@som.umaryland.edu

[1] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland
[3] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
Full list of author information is available at the end of the article

Plots of *hrp2/3* coding regions, grouped by year of sample collection, showed a decrease in median standardized coverage among Malawian samples (n = 150) suggesting the importance of a careful, properly controlled follow up to determine if an increase in frequency of deletions has occurred between 2007–2008 and 2014–2015. Among Malian (n = 90) samples, median standardized coverage was lower in 2002 than 2010, indicating widespread deletions present at the gene locus in 2002.

Conclusions: The GC_3 tool accurately classified *hrp2/3* deletions and provided informative tables and figures to analyse targeted gene coverage. GC_3 is an appropriate tool when performing preliminary and exploratory assessment of locus coverage data.

Keywords: Malaria, Rapid Diagnostic Test, hrp2, hrp3, Deletion, Gene coverage, Genomics, Bioinformatics

Background

From 2010 to 2020, national malaria control programmes (NMCPs) distributed 2.2 billion rapid diagnostic tests (RDTs) for malaria and 3.1 billion RDTs were sold by manufacturers, the majority of these going to malaria-endemic countries in sub-Saharan Africa [1]. RDTs are an integral part of nearly all NMCP's clinical and field interventions since they provide quick and effective malaria diagnosis. These RDTs include a small cassette detecting *Plasmodium*-specific antigens in the blood of an infected individual and are user-friendly and affordable [2]. Predominantly, RDTs detect the *Plasmodium falciparum*-specific antigen histidine-rich protein 2 (HRP2), which is released into the bloodstream in large quantities when infected red blood cells lyse [3]. *Plasmodium falciparum* accounts for vast majority of the 241 million reported human malaria cases in 2020 and is the primary parasite causing malaria-related mortality and morbidity [1]. Due to considerable sequence similarity between the two proteins, (HRP3 is a truncated protein of HRP2 [4] and the two are encoded by similar loci), HRP3 can also bind to the monoclonal antibody on HRP2-based RDTs, but becomes more apparent in high-density infections [5]. As evidence of their effectiveness, 94% of WHO-qualified RDTs are either HRP2-based or based on a combination of HRP2 and a partner antigen, such as parasite lactate dehydrogenase or aldolase [6–8]. HRP2-based RDTs are an essential diagnostic tool for NMCPs to scale surveillance operations and adequately assess infection, leading to proper treatment administration, measure intervention progress and identify malaria reservoirs.

Recently, however, the effectiveness of HRP2-based RDTs is becoming compromised due to the emergence of deletions in the *hrp2* and *hrp3* (*hrp2/3*) loci that prevent the expression of a detectable protein [6, 8–16]. In particular, full deletions, as well as some partial deletions, in one or both of these genes eliminate HRP2 and/or HRP3 signal on RDTs, preventing accurate malaria diagnosis. Previous estimates of *hrp2/3* deletion prevalence report higher frequencies in South and Central America, followed by Africa, then Asia and Oceania [17]. Low-transmission areas with high treatment rates, characteristics often found in elimination settings, are especially at risk for the spread of strains with *hrp2/3* gene deletions, as models show that, under those conditions, strains with *hrp2/3* deletions have a strong fitness advantage over those with intact genes [18]. Therefore, as NMCPs continue to control and move toward elimination, it is critical to monitor the presence and spread of *hrp2/3* deletions. Without fully understanding the dynamics of *hrp2/3* deletions, and spread of those deletions in particular, undiagnosed infections may lead to an increase in malaria prevalence and mortality, and hinder global progress towards control and elimination.

A computational tool that facilitates detection and classification of deletions in *hrp2/3* (e.g. partial *vs*. complete deletions) in published whole genome sequencing (WGS) datasets will enable rapid and detailed analysis of deletions within datasets, and comparisons between datasets. The development of baseline values as well as the comparison of deletion prevalences across current samples sets as well as temporal comparison between these and previously published datasets may be particularly informative. Previous studies have performed analyses using WGS data [19–21]; however, implementation of the methods used in these studies requires a strong understanding of bioinformatics tools and packages. The development of a more user-friendly computational tool would expand the ability to assess the presence of locus deletions based on WGS coverage data to a wider audience investigating copy number variations, including deletions, in *hrp2/3* or other target genes. This work aimed to fill this gap, by developing a computational tool, termed "Gene Coverage Count and Classification", or GC_3, to provide translatable results on the presence of *hrp2/3* deletions and their classification, based on short-read WGS data, among global *P. falciparum* samples for which WGS data is available.

Methods

Samples

The WGS data used were generated either by direct sequencing of total DNA extracted from each isolate or by sequencing post selective whole genome amplification (sWGA) of extracted DNA, and were reported previously [22]. Some of the WGS datasets were generated as part of the MalariaGEN project [23] and downloaded from the Sequence Read Archive (SRA). A selection of field samples representing 19 different countries from Africa (n = 9), South America (n = 5), Asia (n = 4) and Oceania (n = 1), for a total of 1120 datasets (1114 global samples + 6 reference strains), were evaluated for general results (Additional file 1: Figure S1). The following laboratory reference strains with known *hrp2* and *hpr3* genotype were used for developing and testing GC_3: NF54 (West Africa) – *hrp2* and *hrp3* present, 7G8 (Brazil) – *hrp2* and *hrp3* present, NF135.C10 (Cambodia) – *hrp2* and *hrp3* present, NF166 (Guinea) – *hrp2* and *hrp3* present, Dd2 (Laos) – *hrp2* absent/*hrp3* present and HB3 (Honduras) – *hrp2* present/*hrp3* absent.

Read coverage files were generated by aligning raw reads in fastq format to the *Pf*3D7 reference genome assembly (PlasmoDB release v24) using bowtie2 (v2.2.9 and above). Alignment files in BAM (Binary sequence Alignment/Map) format were processed according to GATK's (Genome Analysis Toolkit) Best Practices documentation. Genome-wide coverage per site was recovered using bedtools' genomecov function [22, 24]. The resulting BED (Browser Extensible Data) file (a tab-delimited text file) is used as the initial input to GC_3. However, any delimited file with columns for molecule identifier (e.g. Pf3D7_08_v3), chromosomal position and coverage value is acceptable. When comparing sample datasets, coverage values per base pair (bp) were standardized by dividing 'locus coverage' by 'subtelomeric mean coverage' to account for differences in sequencing depth among samples.

Computational tool framework and algorithm

For GC_3 to function properly, Python v3.0 and R v4.1.1 (or later versions) with the following libraries must be installed: readxl, writexl, dplyr, reshape2, and ggplot2. GC_3 uses a Python-based script to extract read coverage information for genomic coordinates set by the user and processes these output files using an R script (Fig. 1). Following the framework, the user is required to provide input parameters at two junctions.

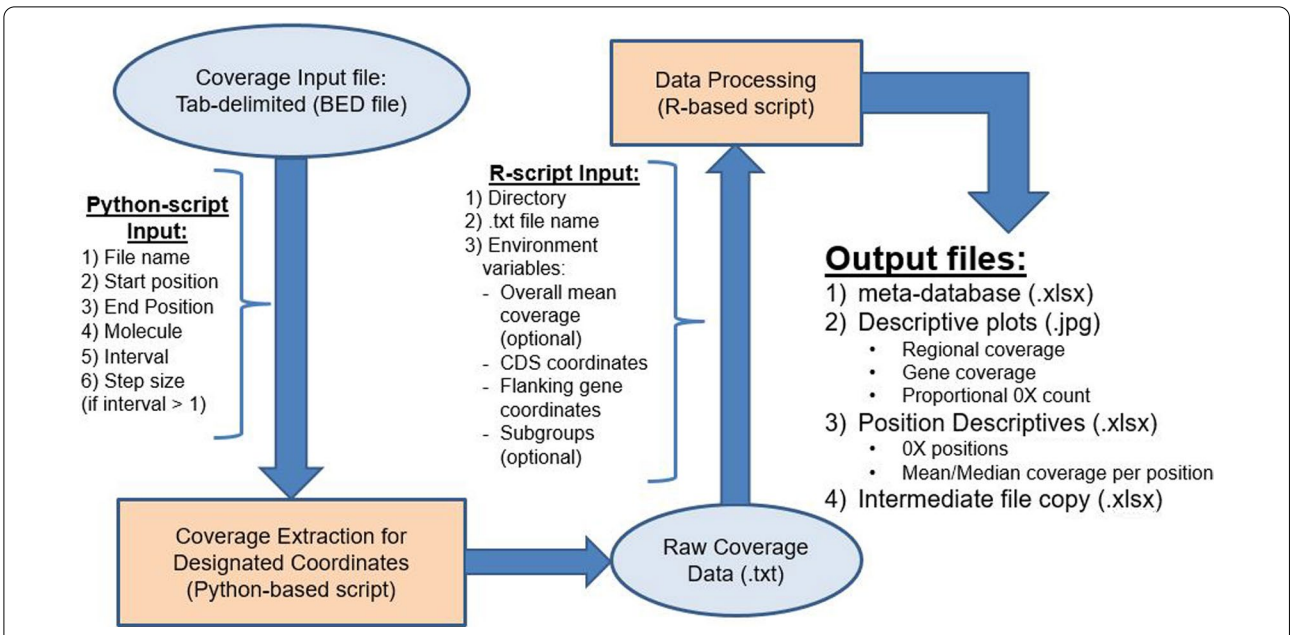


Fig. 1 GC_3 framework. GC_3 extracts read coverage information and processes it into a metric database and descriptive tables/figures. Ovals denote initial/intermediate input(s). Orange rectangles denote scripts for data processing. User input parameters are needed at two junctions in the process and are listed (required and optional). (1) Python script extracts coverage data either using a "sliding window", or coverage at every locus between user-defined start and end coordinates. Overall mean coverage between start and end coordinates can be extracted using a separate function. Output files from the python script (i.e. intermediate output) become the input into the R script, which generates metrics and relevant tables/figures. (2) User input into the R-script is required to define path (directory) to intermediate output as well as the file name, target gene coordinates, intron coordinates (if necessary), coordinates of regions of interest (e.g. flanking genes), and definition of subgroups (optional). Output from R script is Excel versions of intermediate outputs, metrics database, position descriptive database, and relevant figures

Extracting target coverage data – Python-based script

Within the Python script, the user is required to provide parameter inputs depending on the desired output. To use the "sliding window" option, the user must provide: (1) name of input file, (2) start coordinate, (3) end coordinate, (4) molecule identifier containing target locus (e.g. Pf3D7_08_v3), (5) interval size (i.e. window length, in base pairs), and (6) step size (i.e. shift between windows, in base pairs). In the initial window, defined by start coordinate and interval size, the average coverage is obtained by adding read coverage across all positions and dividing by interval size. The start position is then updated by adding step size to the previous start coordinate and the process is repeated until the end coordinate is reached. The output file will report an interval's start and end coordinates separated by a colon and flanked by apostrophes and the interval's average read coverage separated from the coordinates by a colon (e.g. '1,290,240:1,290,740': 294.228).

If individual coverage of all positions in the interval of interest is desired, then the interval size should be set to 1, and GC_3 will extract values for each coordinate between start and end coordinates, inclusively (step size is automatically set to 1). Output file will report the position and respective coverage (e.g. '1,372,236': 387). The intermediate output is a text file with position(s) and corresponding read coverage values.

Additionally, the user can calculate mean coverage between start and end coordinates using a separate GC_3 function. User parameters needed are (1) name of input file, (2) start coordinate, (3) end coordinate, and (4) molecule identifier. This function is needed if the user desires to know, for example, the mean coverage over a wider region or to standardize coverage between different sets of samples (i.e., sample subgroups).

Coverage data processing—R-based script

The user will need to input intermediate output files into the separate GC_3's R script to clean, and generate sample metrics and descriptive plots. At the start of the R script, the user will define (1) path to the intermediate files, (2) name of the intermediate file(s), (3) target locus' coordinates in reference genome, (4) gene's intron coordinates (if necessary), in reference genome, (5) position coordinates of interest (e.g. flanking gene positions), (6) list of subgroup sample identifiers and subgroup name (optional). If read coverage is to be standardized relative to coverage in a reference chromosome or chromosomal segment, then a file of mean read coverage per chromosome or segment (obtained as described above) per sample should also be defined. The GC_3 R script will output several files, namely, (*i*) Excel version of intermediate text files, (*ii*) summary metrics: sample identifier, overall mean coverage—if mean coverage file included, mean target gene coverage, proportion of gene positions with coverage, proportion of smaller regions of interest (including coding regions, exons), deletion classification, and count of positions with 0X coverage, (*iii*) read coverage information for target gene (number of positions with zero coverage, and mean and median coverage per position over all samples), and (*iv*) descriptive plots: sliding window coverage over region of interest (i.e. subtelomeric region), all coordinates coverage over target gene positions (i.e. *hrp2* and *hrp3*), and proportion of positions with zero coverage.

Python and R scripts can be found at the Silva group's GitHub (https://github.com/igs-jcsilva-lab) as well as a README file with detailed instructions and input examples.

Detection of deletions in *hrp2*, *hrp3* and flanking regions

The gene structure of *hrp2* and *hrp3* in the reference 3D7 strain was obtained from PlasmoDB (www.plasmodb.org). Both *hrp2* and *hrp3* consist of two coding exons. Exon1 is 69 bp in length for both genes, and exon2 is 848 bp long in *hrp2* and 758 bp in *hrp3* (Fig. 2). The analysis of WGS data focused on subtelomeric regions of chromosome 8 (*P. falciparum* 3D7 reference strain coordinates 1,290,240–1,443,449, for a total of 153,209 bp), containing the *hrp2* coding DNA sequence (CDS) and intervening intron, and of chromosome 13 (*P. falciparum* 3D7 reference strain coordinates 2,731,041–2,892,340, for 161,299 bp), containing the *hrp3* CDS and intron. Subtelomeric coordinates were chosen to include the closest "essential" gene [25] downstream of *hrp2* or *hrp3* and farthest upstream functional gene (i.e. *Pf*EMP1-encoding *var* gene).

Metrics were generated to classify samples by presence or absence of full or partial deletions in each locus of interest. If ≤ 25% of the CDS was missing (i.e. at most 25% of the reference CDS had zero coverage) the locus was considered present with a "small deletion of uncertain functional impact" (SDUFI). If > 25% (but not 100%) of the reference CDS positions had zero coverage the sample was classified as having a partial deletion (25% < %-positions-with-zero-coverage < 100%); it was classified as having a complete deletion if all CDS positions have zero coverage. This classification is partly informed by Sepúlveda and colleagues [19], who implemented an algorithm to perform deletion calling without having to analyse the coverage profile of the entire genome. They classified deletions based on a 75% threshold of positions with ≤ 2X coverage, but may be too stringent and decreased GC_3 thresholds as explained above to account for "SDUFIs" or partial deletions less than the 75% threshold that might impact protein detection.

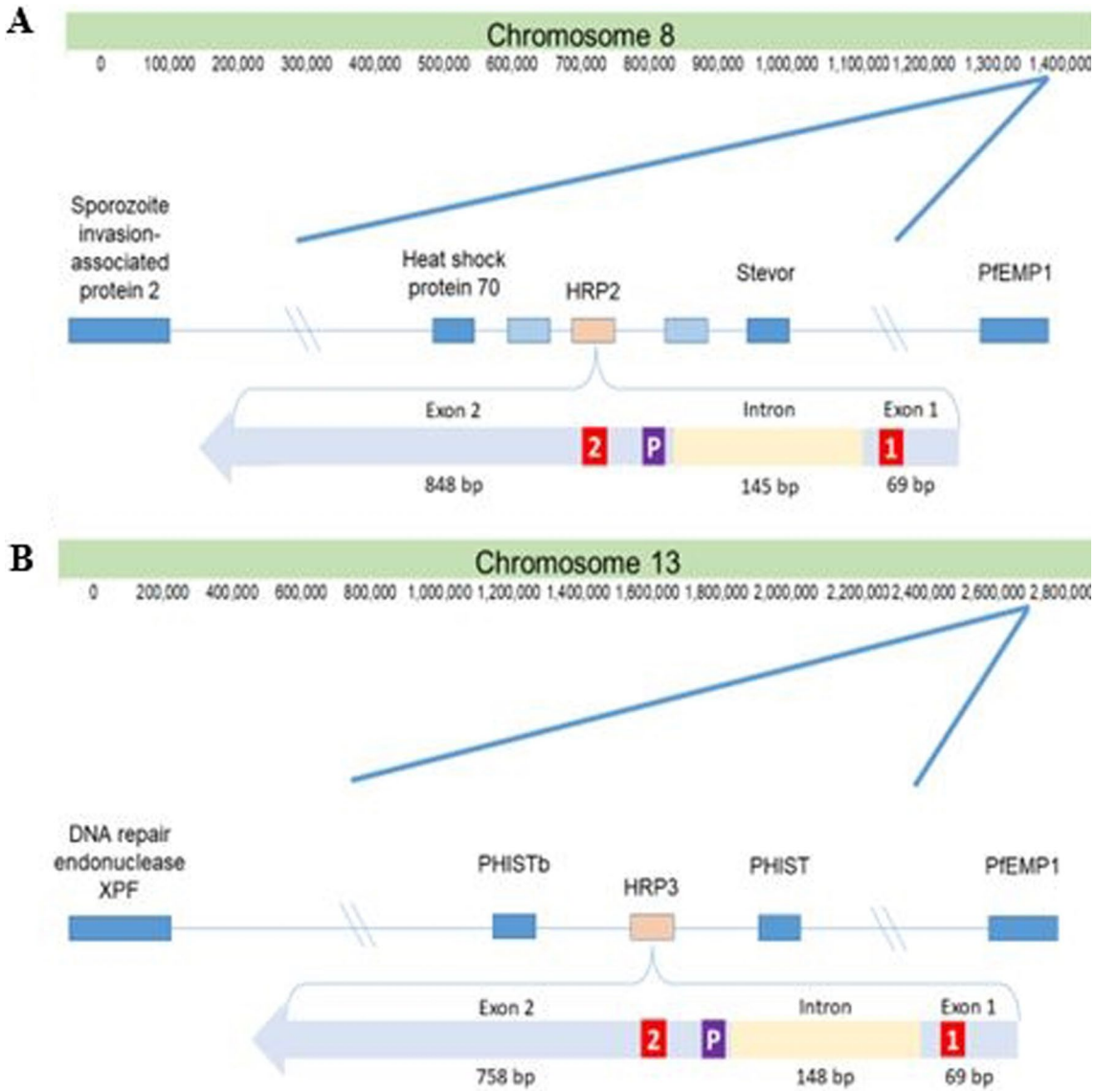


Fig. 2 Schematic of target genomic regions, encoding *hrp2/3*. **A** Schematic of the *hrp2*-containing chromosomal region plus flanking genes (unlabeled light-blue boxes represent pseudogenes), labelled with their respective gene products, and coding regions (exon1 and exon2), their base pair length and direction of transcription. Red highlighted areas denote locations of *hrp2/3*-specific qPCR forward primer (1) and reverse primer (2), and the purple highlighted area denotes the location of the qPCR probe (P). Chromosome 08 PfEMP1 reference – Pf3D7_0833500 (multi-copy gene). **B** Same as **A**, but for hrp3. Chromosome 13 PfEMP1 reference – Pf3D7_1373500 (multi-copy gene)

Deletions of flanking genes were assigned to samples if > 25% of the flanking gene's positions reported zero coverage. Intergenic regions were excluded to reduce the effect of variable read coverages in non-coding regions.

GC_3 agreement with qPCR assay

A previously described *hrp2/3*-specific qPCR assay capable of detecting locus deletions in mono- and poly-clonal infections [26] was utilized to compare with the *hrp2/3* deletion genotype inferred by GC_3. In summary, primer sequences were adapted from conventional PCR [27] to bind to conserved regions of *hrp2*, *hrp3* and an apicomplexan-specific single copy gene used as positive control, *rnr2e2* (ribonucleotide reductase R2_e2, [28]). The computational approach used by GC_3 for the detection of *hrp2/3* deletions was compared to this *hrp2/3*-specific qPCR assay [26], using the following samples:

- NF54—Positive control
- 7G8—Positive control

- Dd2—*hrp2* absent control
- HB3—*hrp3* absent control
- 17 global samples (see Additional file 1: Table S1 for details)

The presence and classification of *hrp2/3* deletions is reported for the four laboratory reference strains mentioned above and for 17 global samples from Brazil (n = 3), Cambodia (n = 6), Mali (n = 3), Malawi (n = 4) and Thailand (n = 1). Global samples with accessible DNA material were randomly selected to represent the following GC_3-inferred genotype subgroups: samples with no deletions, *hrp2* deletion (complete), *hrp3* deletion (complete), double *hrp2/3* deletion, low overall sample mean read coverage (< 20X), possible discordant pairs (partial deletion, with non-zero coverage in qPCR primer binding sites), and PCR primer site deletions (samples with zero coverage in qPCR primer binding site – either in *hrp2* or *hrp3*). Accession ID and subgroup stratification of global samples can be found on Additional file 1: Table S1.

Statistical analyses

When measuring correlation between mean coverage in *hrp2/3* positions and subtelomeric or upstream/downstream gene, Spearman's rank correlation method was used (Additional file 3). Spearman's method accounts for non-parametric distribution and, therefore, mean coverages were not standardized [29]. R v4.1.1 program was used to conduct statistical analysis.

Results

Sample read coverages by sliding windows of 1000 bp intervals and 500 bp step size were generated over the subtelomeric regions of chromosome 8 and chromosome 13 (sum of coverage across all positions in interval/interval length). Additionally, coverage at every position (interval = 1) was generated at every position between coordinates 1,372,236 to 1,377,299 on chromosome 8 and 2,835,756 to 2,847,557 on chromosome 13. These positions corresponded to *hrp2* and *hrp3* coordinates plus 2000 bp on either end of their respective coding regions.

Demonstrating GC_3 features using laboratory strains of known genotype

WGS data from reference laboratory strains were analysed to estimate *hrp2/3* coverage per bp and the proportion of positions with coverage by at least one read (≥ 1X coverage) at *hrp2* and *hrp3* coordinates, and ultimately evaluate the validity of results from GC_3. Expected coverage was estimated using each respective subtelomeric region as reference. Overall, for each lab strain reference, excellent concordance was found between coverage values in each locus and the respective subtelomeric chromosomal regions (Table 1). Mean subtelomeric coverage of the Dd2 strain (with *hrp2* deletion genotype) was high (chromosome 8: 29X; chromosome 13: 48X), and, as expected, mean coverage at the *hrp2* positions was 0X, while *hrp3* mean coverage was 45X (with 100% of CDS coordinates with coverage > 0). The HB3 strain (*hrp3* deletion genotype) was sequenced to ~ 145X coverage (chromosome 08 and chromosome 13 subtelomeric regions with 159X and 131X coverage, respectively). Mean coverage at the *hrp3* positions was 0X (50% proportional coverage) and while *hrp2* was similar to genome-wide coverage (142X, with 100% proportional coverage of gene positions). Residual coverage may have occurred at the *hrp3* gene of HB3 despite its known deletion due to mapping of some reads originating from *hrp2* and mapping to similar but non-orthologous locations. GC_3 correctly identified HB3 as having a *hrp3* gene deletion (Table 1). These results are similar to previously described coverage profiles of Dd2 and HB3 [19].

Plotting of read coverage in subtelomeric region of select reference strains

GC_3 can create plots of the sliding window findings in order to provide a visual perspective of the target region.

Table 1 Subtelomeric coverage, gene coverage, and coding region proportional coverage among known reference strains

Reference Strain	Mean subtelomeric read coverage (sum of coverage/bp)		Mean gene coverage (sum of coverage/bp)		Proportion of coding positions with ≥ 1X coverage	
	Chromosome 08	Chromosome 13	*hrp2*	*hrp3*	*hrp2*	*hrp3*
NF54	156.5	173.7	125.2	134.8	100%	100%
7G8	201.9	236.3	247.0	269.9	100%	100%
NF135.C10	35.3	42.7	36.9	41.7	100%	97%
NF166	280.0	328.4	295.7	344.6	100%	100%
Dd2	29.3	48.3	0.0	45.0	0%	100%
HB3	159.3	130.5	142.2	0.35	100%	50%

To illustrate the coverage data provided in Table 1, the subtelomeric regions containing *hrp2* and *hrp3* of reference strains Dd2, HB3 and NF54 were plotted (Fig. 3). Results were normalized on a log scale to better visualize large fluctuations in coverage generated by whole genome shotgun sequencing. Sliding window plots confirm validation results of Dd2, HB3 and NF54, and clearly illustrate coverage for each respective strain. On chromosome 08, coverage of the Dd2 strain decreases to zero for several thousand base pairs that include the *hrp2* locus, whereas NF54 and HB3 have high coverage in the same region. Noticeably, in the HB3 strain, a section of the subtelomeric region upstream of the *hrp2* CDS has poor coverage (near position 1,400,000 on chromosome 8). This section of poor coverage would not impact *hrp2* presence/absence, and could be due to the presence of one or more deletion(s), or to poor mapping. Poor read mapping can occur in the subtelomeric regions for several reasons, including the presence of multiple members of highly variable multigene families (*var*, *stevor* and *rifin*) that differ between strains or to the presence of low complexity regions. On chromosome 13, it is the HB3 strain that has several thousand base pairs with little or no coverage, including the *hrp3* locus, whereas NF54 and Dd2 coverage remains high. GC_3 visuals showed no deletions in NF54, a complete *hrp2* deletion in Dd2 and a large section of little to no coverage at the *hrp3* locus in HB3, respectively.

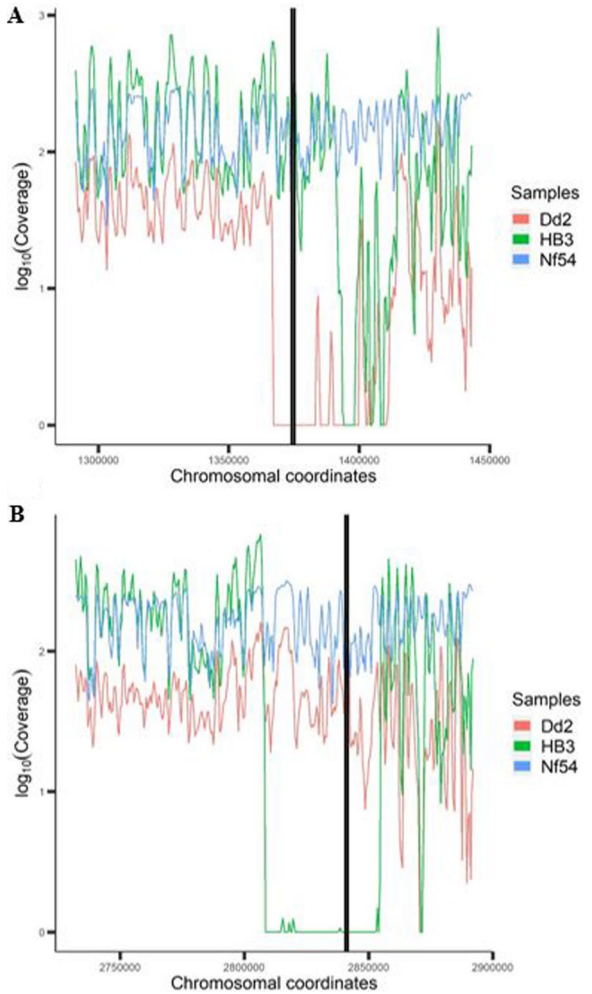


Fig. 3 Subtelomeric read coverage distribution plots. Log_{10} of subtelomeric read coverage for reference strains NF54 (*hrp2* and *hrp3* present), Dd2 (*hrp2* absent) and HB3 (*hrp3* absent). Coverage was measured using a 1000 bp sliding window (i.e. interval) with a 500 bp "step" between windows. A. Chromosome 08 subtelomeric region coverage (black lines denotes *hrp2* coding positions 1,374,236–1,375,299). B. Chromosome 13 subtelomeric region coverage (black lines denotes *hrp3* coding positions 2,840,727–2,841,703)

GC_3 agreement with a hrp2/3-specific qPCR assay on field samples

A subset of global samples (n = 17) and reference strains (n = 4) underwent qPCR specific for *hrp2* and *hrp3* to compare with GC_3 results. Two among the selected global samples were excluded due to low parasitaemia resulting in very low or no detection of the positive control gene by qPCR (Cq threshold cutoff = 37.5). There was very good agreement between GC_3 (computational) and qPCR assay results. Out of four references strains and remaining 15 global samples, GC_3 matched qPCR results 19/19 (100%) for *hrp2* and 18/19 (94.7%) for *hrp3* (Table 2). Only one sample (IGS-CBD-099) had a discordant result between methods. In particular, for this sample, GC_3 classified it as having a partial deletion at the *hrp3* locus, and coverage assessment with base-pair granularity suggested partial lack of read coverage, including the exon 2 primer binding region, between 2,841,390—2,841,412 (Fig. 4). It is noteworthy that the average coverage in this region is very low (~ 1X), however coverage is high for the corresponding chromosomal subtelomeric and core regions (~ 124× and ~ 143×, respectively). On the other hand, the qPCR assay was positive for *hrp3* (Cq = 25.4). Taken together, the results suggest the sample has a partial deletion at the *hrp3* locus, which does not encompass the qPCR primer binding regions, but that is possibly close enough to the binding site of the primer in exon 2 to interfere with read mapping in that region.

Comparison of GC_3 output using Kenyan and Peruvian sample sets previously genotyped for hrp2/3 deletions

Previously, a subset of Kenyan samples (n = 27) was genotyped for *hrp2/3* deletions, with two and one deletions identified in *hrp2* and *hrp3*, respectively [19] (Table 3). In addition, Sepúlveda and colleagues also identified no *hrp2* deletions among twelve Peruvian samples and two

Table 2 Agreement between GC_3 deletion assessment results on global samples and *hrp2/3*-specific qPCR assay[a]

Sample name	Country	Phenotype subgroup	*hrp2* [b]	*hrp3* [b]	*rnr2e2* (control gene)
			GC_3/PCR	GC_3/PCR	PCR
7G8	Reference (Brazil)	Control	Present/Present	Present/Present	Present
NF54	Reference	Control	Present/Present	Present/Present	Present
Dd2	Reference (Laos)	Control—*hrp2* deletion	Absent/Absent	Present/Present	Present
HB3	Reference (Honduras)	Control—*hrp3* deletion	Present/Present	Absent/Absent	Present
IGS-BRA-017sA	Brazil	No deletions	Present/Present	Present/Present	Present
IGS-THL-017	Thailand	No deletions	Present/Present	Present/Present	Present
IGS-BRA-021	Brazil	No deletions	Present/Present	Present/Present	Present
IGS-CBD-026	Cambodia	No deletions	Present/Present	Present/Present	Present
IGS-CBD-031	Cambodia	*hrp2* deletion (complete)	Absent/Absent	Present/Present	Present
IGS-MLI-036	Mali	*hrp3* deletion (complete)	Present/Present	Absent/Absent	Present
IGS-BRA-001sA	Brazil	Double *hrp2/3* deletion	Absent/Absent	Absent/Absent	Present
IGS-CBD-008	Cambodia	Low coverage sample	Present/Present	Present/Present	Present
IGS-MWI-254sA	Malawi	Low coverage sample	Present/Present	Present/Present	Present
IGS-MWI-251sA	Malawi	Low coverage sample	Present/Present	Present/Present	Present
IGS-MLI-039	Mali	*hrp2* discordant pair	Present/Present	Present/Present	Present
IGS-MLI-031	Mali	*hrp3* discordant pair	Present/Present	Present/Present	Present
IGS-CBD-034	Cambodia	*hrp2* PCR primer deletion	Present/Present	Present/Present	Present
IGS-CBD-094	Cambodia	*hrp3* PCR primer deletion	Present/Present	Present/Present	Present
IGS-CBD-099	Cambodia	*hrp3* PCR primer deletion	Present/Present	Absent/Present	Present

[a] Cells in purple and blue denote agreement between GC_3 and qPCR results (i.e. "PCR"); red denotes disagreement between methods

samples with deletions in *hrp3* [19]. In that study, the criterion used to call deletions was > 75% of gene positions with ≤ 2X coverage [19]. To determine how GC_3 performed on a similar set of samples SRA samples were downloaded from the same time point from Kenya (n = 57, including 24 from [19]) and Peru (n = 11, including 7 from [19]). For *hrp3*, the same number of deletions were identified as reported previously [19]. However, the *hrp2* were discordant. One complete deletion was identified among Kenyan samples, and one partial *hrp2* deletion among the Peruvian samples, which differs from previous reporting. Among the Kenyan samples, a previous study reported two *hrp2* deletions [19], one of which was also identified by GC_3. Whereas the discordant Peruvian sample was only identified to have a deletion by GC_3. The difference in assessment is likely due to differences in the criteria used between GC_3 and that used by Sepúlveda and colleagues to call deletions.

Analysis of novel samples for hrp2 and hrp3 deletions

All global samples used in the study (n = 1114) were examined for *hrp2/3* deletions (Additional file 1: Table S2). Cambodian (n = 127) and Brazilian (n = 20) samples were further visualized in more detail at the subtelomeric regions of interest (Additional file 1: Figure S2) and examined for *hrp2/3* exon presence/absence since they have not previously been described (Table 4). Although *hrp2/3* have not been described for these samples, computational results are comparable to previous estimates in each respective region, where deletions have been previously observed [19, 30]. Among Cambodian samples collected in 2009–2011, there were one *hrp2* deletion, eight *hrp3* deletions, and two *hrp2/3* double deletions (both *hrp2* and *hrp3*), with frequencies in this sample set of 0.8%, 6.3% and 1.6%, respectively. All *hrp2* deletions corresponded to absent exons (> 25% zero coverage positions on both exons), but were classified as two partial and one complete *hrp2* deletion as two samples still had coverage in a low proportion on *hrp2* positions. Deletions of *hrp3* among Cambodian samples were classified as seven partial and three complete deletions. Among Brazilian samples collected in 2016, four had *hrp2/3* double deletions, and eleven with *hrp3* deletions, corresponding to frequencies of 20% and 55% respectively. Of the four *hrp2* deletions, two were partial deletions, and two were complete *hrp2* deletions. Of the 15 *hrp3* deletions, one was a partial deletion on exon 2, and 14 were complete *hrp3* deletions.

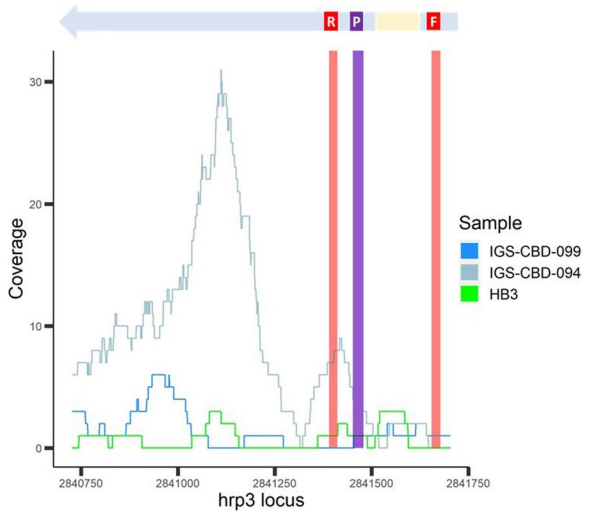


Fig. 4 Coverage of *hrp3* among select validated samples. Coverage plot of *hrp3* locus of validated samples IGS-CBD-099, IGS-CBD-094 and strain HB3 (known *hrp3* deletion genotype) with *hrp3* schematic representation above the plot. Red areas denote the primer binding sites, and purple area denotes the probe binding site of the *hrp2/3*-specific qPCR assay. GC3 assigned IGS-CBD-099 as *hrp3*-absent; however the 5′ end of the gene was amplified by qPCR. As a comparator, IGS-CBD-094 is a similar sample (apparent deletion at qPCR primer binding site in exon 1) that GC_3 assigned as *hrp3*-present. HB3 is a laboratory reference strain known to be missing the *hrp3* locus. It is important to note some position with non-zero coverage in HB3, which suggests that non-orthologous reads from HB3 map the *hrp3* locus of Pf3D7

Quantifying flanking region deletions among Cambodian samples

Deletions at the *hrp2/3* positions may extend to flanking genes, possibly with additional impact on overall parasite fitness. Therefore, an option in the GC_3 R-script was built in to assign deletions of flanking genes. To determine whether deletions extend into these flanking coding regions, a table was generated for Cambodian samples, where samples with > 25% gene positions with zero coverage in upstream and downstream flanking regions were classified as having a locus deletion (Table 5). Of note are the observations that the presence of a *hrp2/3* deletion is not always associated with deletions in flanking genes and, conversely, deletions in flanking genes are not always associated with *hrp2/3* deletions. A subset of Cambodian samples has been plotted to illustrate flanking gene coverage as it relates to *hrp2* or *hrp3* (Additional file 1: Figure S3). Results suggest that deletions can occur independently in *hrp2* (or *hrp3*) and their respective flanking genes.

Temporal comparison of standardized coverage of hrp2/3 positions

Coverage plots of *hrp2/3* coordinates were generated for Cambodian, Malawian, and Malian samples to demonstrate: (1) magnified plots of only *hrp2/3* positions and (2) differences in relative depth of coverage between samples collected at different time points (Fig. 5). Cambodian samples were collected in Battambang, Pailin, Koh Kong, Kampot, Kampong Speu, Oddar Meanchey, Preah Sihanouk, and Preah Vihear districts from volunteers aged 18–65 years in 2010 and 2011, and then sequenced at IGS with 100 bp paired-end Illumina reads [31]. NF135.C10 was cultured in the laboratory and sequenced at IGS with 150 bp paired-end reads. Two datasets of Malawian samples were collected in 2007–08 and 2014–16. Samples from 2007–08 were collected during a malaria drug study in Ndirande, outside Blantyre, from children 6 months to

Table 3 Deletion identification and classification of previously genotyped samples

Country	n	*hrp2*					
		No Deletion	GC_3 Deletion Classification (Partial/Complete)[a]	ex1-/ex2 + [b]	ex1 + /ex2-[b]	ex1-/ex2-[b]	Previous Genotype (Deletion/Total)[c]
Kenya	59	58	0/1	0	0	1	2/27
Peru	11	10	1/0	1	0	0	0/12
Country	**n**	***hrp3***					
		No Deletion	**GC_3Deletion Classification (Partial/Complete)[a]**	**ex1-/ex2 + [b]**	**ex1 + /ex2-[b]**	**ex1-/ex2-[b]**	**Previous Genotype (Deletion/Total)[c]**
Kenya	59	58	1/0	0	0	1	1/27
Peru	11	9	2/0	0	1	1	2/12

[a] *hrp2/3* deletions assigned to isolates with > 25% CDS positions with zero coverage

[b] Exon absence (-) assigned to isolates if > 25% exon positions have zero coverage. (+) signifies exon is present

[c] Previous *hrp2*/3 deletion genotype results from Sepúlveda and colleagues. Deletions called for samples with > 75% of coding region with ≤ 2X coverage [19]

Table 4 Deletion identification and classification of undescribed samples

Country	n	*hrp2*				
		No Deletion	Deletion Classification (Partial/Complete)[a]	ex1-/ex2 + [b]	ex1 + /ex2-[b]	ex1-/ex2-[b]
Cambodia	127	124	2/1	0	0	3
Brazil	20	16	2/2	0	1	3
Country	**n**	***hrp3***				
		No Deletion	**Deletion Classification (Partial/Complete)[a]**	**ex1-/ex2 + [b]**	**ex1 + /ex2-[b]**	**ex1-/ex2-[b]**
Cambodia	127	117	7/3	0	2	8
Brazil	20	5	1/14	0	1	14

[a] *hrp2/3* deletions assigned to isolates with > 25% CDS positions with zero coverage

[b] Exon absence (-) assigned to isolates if > 25% exon positions have zero coverage. (+) signifies exon is present

5 years of age and sequenced at IGS with 150 bp paired-end reads [32]. Samples collected in 2014–16 are from a cohort study of malaria incidence in Chikwawa, south of Blantyre, where samples from volunteers aged 2–8 years were sequenced at IGS with 150 bp paired-end reads [22]. Comparisons were also made between two Malian datasets from 2002 and 2010, both collected in Bandiagara, Mali. Samples from 2002 are from a case/control study of severe malaria among 3 months to 14 year old volunteers and were sequenced at IGS with 100 bp paired-end reads [33]. Malian samples collected in 2010 are from a cohort study of malaria incidence among volunteers aged 1–5 years and sequenced at IGS using 150 bp paired-end reads [34]. It should be noted that comparing these datasets are for illustrative purposes only, since the extent to which relatively small samples sizes and potential confounders (including sampling location and strategy, sample independence, sequencing approach, read length) impact observed deletion frequency is unknown.

Table 5 Frequency of deletions in *hrp2/3* flanking genes, among Cambodian samples

Flanking gene[b] (Upstream/Downstream)[c]	*hrp2*[a]		*hrp3*[a]	
	Present	Absent	Present	Absent
Present/Present	124	2	113	3
Absent/Present	0	1	4	0
Present/Absent	0	0	0	1
Absent/Absent	0	0	0	6
Total	124	3	117	10

[a] Deletion assigned to samples if > 25% of coding region positions had zero coverage

[b] Deletion of flanking genes assigned to samples with > 25% gene positions with zero coverage

[c] For *hrp2*, an upstream gene was a STEVOR family gene, and a downstream gene encoded heat shock protein 70. For *hrp3*, upstream and downstream genes were PHIST-encoding genes of unknown function

To account for differences in sequencing depths between samples, coverage values were standardized. For each sample, site or locus coverage were divided by the expected coverage, obtained from mean coverage in subtelomeric region in which each locus is located (see Methods for coordinates). Standardized coverage of ~ 1 shows locus coverage similar to subtelomeric mean coverage. A strong, positive correlation between subtelomeric and *hrp2/3* gene coverage justifies the use of this standardization approach (Additional file 1: Figures S4 and S5). A decrease in standardized coverage over time would suggest an increase in frequency in *hrp2/3* deletions. To avoid undue impact of outlier standardized values, median standardized coverage was plotted per group.

Standardized read coverage for Cambodian samples was plotted alongside standardized coverage for the geographically representative strain NF135.C10 (Fig. 5A), and showed that the uneven standardized coverage of NF135.C10 is mirrored in the clinical samples. This suggests there are sequence-inherent properties that impact sequencing or mapping success. (see Additional file 2: Table S3 for descriptive coverage of each sample). In contrast, Malawian samples collected in 2014–2016 had lower median standardized coverage than the sample set collected in 2007–2008 (Fig. 5B), everything else being equal, this would suggest an increase in *hrp2* and *hrp3* deletions between the two time points. Interestingly, median standardized coverage is low (< 1) in both time points, showing that read coverage in the target genes is half of that in the respective subtelomeric regions. The majority of Malawian samples underwent sWGA (n = 139) prior to sequencing (Additional file 2: Table S4) which may explain the lower standardized coverage as compared to standardized coverage of directly sequenced samples (Additional file 1: Figure S6). Finally,

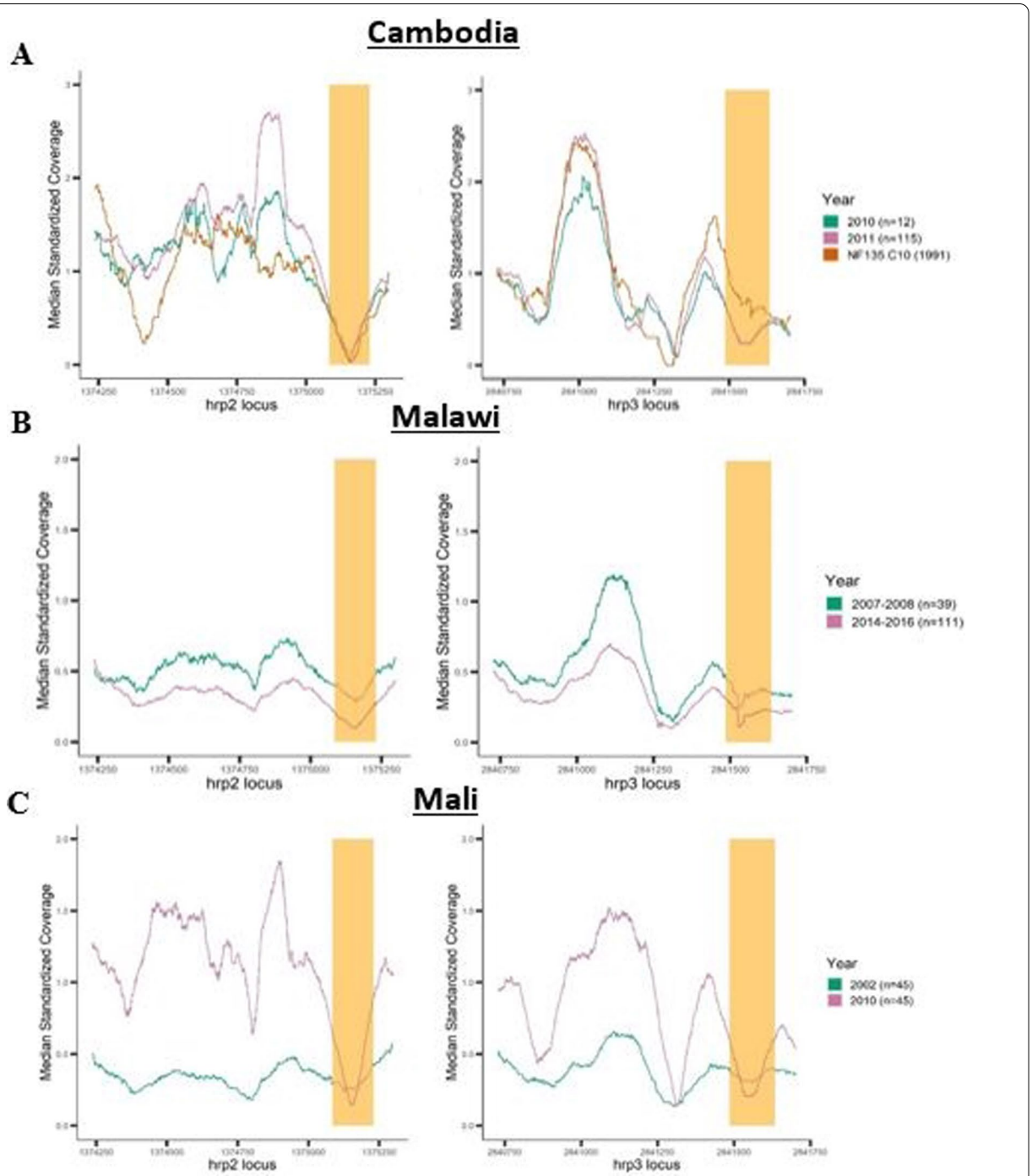


Fig. 5 Median of standardized coverage between longitudinal subgroups. Median of standardized coverage [SUM (Coverage/Subtelomeric Mean Coverage)/ Total Samples] in *hrp2* and *hrp3* grouped by year of sample collection (Year). Countries include A. Cambodian *hrp2* and *hrp3* positions (n = 127), B. Malawi *hrp2* and *hrp3* positions (n = 150), C. Mali *hrp2* and *hrp3* positions (n = 90). Tan shading marks intron positions of target gene, whereas unshaded areas are exon positions (*hrp2*—> exon 1: 1,375,299–1,385,231; intron: 1,375,230–1,375,085; exon 2: 1,375,084–1,374,236; *hrp3*—> exon 1: 2,841,703–2,841,635; intron: 2,841,634–2,841,486; exon 2: 2,841,485–2,840,727). Analysis and figures were generated using R v4.1.1

Mali samples from 2002 showed lower standardized coverage than 2010 samples on both *hrp2* and *hrp3* positions (Fig. 5C). All Malian WGS data was obtained by direct sequencing of total DNA from venous blood, using a similar protocol [22, 33] with high sequence coverage in the core genomes (Additional file 2: Table S5), a strong

suggestion that the quality of WGS data did not contribute to the observed difference. However, it remains unclear if the datasets are directly comparable as differences in, for example, sample strategy/bias and read length could distort observed frequencies [22, 33, 34]. Mali results highlight the potential impact of read length in some of the observed results. Lower coverage may be due to mapping ambiguity in smaller read datasets. Overall, figures offer a visual perspective between different time points as monitoring of *hrp2/3*-deletions become crucial in the possibility of their expansion; however, caution should be applied since the factors that influence WGS coverage need to be considered as part of any interpretation.

Count of positions with zero coverage

To provide a clearer illustration of the proportional frequency of gene coordinates with zero coverage and the location of those positions along the locus, a view of *hrp2* and *hrp3* gene positions by proportional counts of no coverage (0X coverage) *vs.* coverage (≥ 1X coverage) was generated for Cambodian samples (Fig. 6). Additional file 1: Figure S7 provides the same proportional counts of zero coverage per position for Malawian and Mali samples. Among Cambodian samples, there is a clear increase in 0X coverage positions at the intron regions (*hrp2* intron: 1,375,232–1,375,083; *hrp3* intron: 2,841,636–2,841,484) relative to exon coverage. This is to be expected, as the length (145–148 bp) and the nucleotide composition of these *Pf* introns (*hrp2* AT%: 91%; *hrp3* AT%: 91.2%) prevent unambiguous mapping of 101 bp-long reads centered in the middle of the intron. On *hrp3* positions, there are also two spikes in zero coverage positions on either side of coordinate 2,841,250, likely due to differences in Cambodia samples compared to *Pf*3D7 reference, such as indels or rapidly evolving sequence motifs among genetically similar Cambodian strains, which prevent read mapping in a subset of samples. That read mapping pattern is also observed in the troughs in *hrp3* coverage plot in Fig. 5A, for samples collected in 2010 and 2011 (but curiously not in Pf NF135).

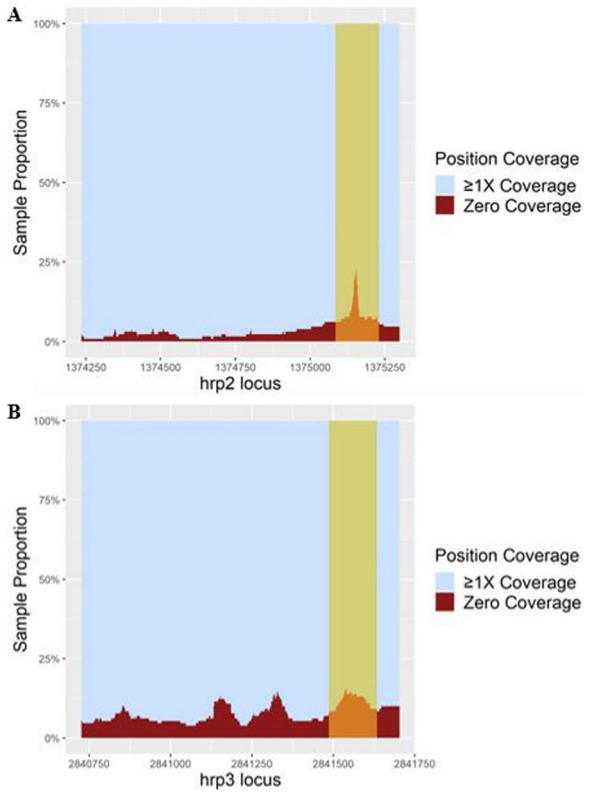


Fig. 6 Proportional count of zero coverage positions at the *hrp2/3* positions Proportional sample count of *hrp2* and *hrp3* gene positions with zero coverage vs. positions with ≥ 1X coverage among Cambodian samples (n = 127). A. shows count of zero coverage positions among *hrp2* positions, and B. shows count of zero coverage positions among *hrp3* positions. The tan section on each plot represents the intron region of each respective gene. Analysis and figures were generated using R v4.1.1

Discussion

Next-generation short-read WGS data has the capability to provide detailed genotype information, but often necessitates a good understanding and use of bioinformatics tools and packages. GC_3 was developed to be a user-friendly computational tool to (1) extract coverage profiles of target genome regions, (2) provide interpretable results regarding location and frequency of deletions, (3) classify samples according to the type of gene deletions, and (4) validate large-scale, qPCR-based, studies conducted to inform NMCPs concerning frequency of copy number variants (including deletions) in genes of translational importance. In this study, it is demonstrated that GC_3 can be used for these purposes by applying it to *Plasmodium falciparum* genome segments, specifically the regions containing *hrp2* and *hrp3* genes. Most NMCPs in malaria-endemic settings rely on HRP2-based RDTs for day-to-day diagnosis in both clinical and field settings. There is evidence of recent expansions of *P. falciparum* strains lacking HRP2, a cause for concern as stated by the WHO [1, 20, 35, 36]. Given the continuing decrease in sequencing costs and the widespread generation of WGS data, computational tools, such as GC_3, that take advantage of such data to efficiently assess the presence of *hrp2*- or *hrp3*-deletion strains, provides a valuable monitoring resource to researchers and public health professionals concerned with malaria RDT effectiveness.

This work demonstrated the validity and utility of the GC_3 tool to assess *hrp2* and *hrp3* deletion frequencies in *P. falciparum* sample sets, as well as to carefully

characterize those deletions in individual samples. The sliding window capability of GC_3 provided a wider view of large deletions in reference lab strains (i.e. Dd2 and HB3) while adjusting for fluctuations in subtelomeric read coverage and allowed for visually-friendly figures. The option of extracting and plotting every chromosome position within an interval allowed for a magnified view of target loci, and better illustrated details in coverage within the target gene. GC_3's genotype results were validated against previously reported genotypes of *P. falciparum* laboratory reference and representative strains, and similar, publicly available sample sets from Kenya and Peru, where *hrp2/3* deletions have been observed [12, 19, 37]. Analysis of samples from Cambodia and Brazil demonstrate GC_3's capability to process novel WGS data from regions other than West Africa, the location of origin of PfNF54, the parental isolate from which the reference 3D7 was cloned [38, 39]. Results are consistent with previous estimates of *hrp2/*3 deletions among each respective country [20, 30, 40]. Of note is the high prevalence of all deletions, and especially *hrp3* deletions among Brazilian samples. Deletions in *hrp2/3*, and reports of high deletion prevalence, were first observed in the American continent [8, 9, 17, 19, 40–42], especially in the *hrp3* locus [8, 43]. Overall, GC_3 can appropriately process and analyse publicly available WGS datasets from a variety of genomic studies.

Additional comparison against a *hrp2/3*-specific qPCR assay demonstrated very good reliability of GC_3's capability. Although there was one discordant result between tools, this may be a reflection of GC_3's sensitivity and the qPCR assay's difficulty to detect partial deletions, as these are only detected by the qPCR-based assay if they overlap the primer-binding or amplicon sites. A potential challenge for GC_3 are the samples with a very low amount of parasite DNA resulting in genomic libraries of substandard quality and overall low depth of coverage and uneven representation of the loci of interest in the genomic library and/or among the WGS data, leading to significant regions of the *hrp2/3* loci with zero coverage (and then a 'deletion' assessment by GC_3), despite the loci being present in the genome. However, this situation was not observed in this study. In the specific case of discordant results in a sample from Cambodia, the sample had very high coverage at the core and subtelomeric regions. In general, partial deletions present a challenge since some cases have shown a qPCR assay can amplify part of *hrp2/3*, but corresponds to false negative RDT diagnoses [44, 45]. Ultimately, there were very few such samples, so their impact on overall results is considered minimal. Although beyond the scope of this study, further examination of *hrp2/3* partial deletions, their specific location within the locus and their respective RDT diagnosis may provide valuable information regarding the most appropriate criteria and thresholds to accurately identify gene deletions with a functional phenotype, i.e., those deletions that abrogate protein expression.

The application of GC_3 to describe and visualize patterns of partial deletions makes it a valuable resource for research purposes. By providing exact genomic coordinates that lack read coverage, GC_3 may inform on genomic sequence backgrounds more prone to such mutations and eventually deletions that prevent RDT detection. Defining exact locations of deletions allows the user to determine whether the observed deletions can be explained by a single deletion event followed by lineage expansion, or whether multiple events need to be invoked to explain the observations. This examination of the evolutionary history of gene deletions can provide insights into mechanisms and rate of origin of indels and allow for improved monitoring of target genes.

Among Cambodian samples, read coverage in flanking genes was further analysed and demonstrated that *hrp2/3* deletions can be restricted to just the locus proper, or extend to flanking genes, but without a discernible pattern. These results are consistent to previous reports [19, 20, 43, 46, 47]. Overall, GC_3 reported similar results in deletion frequency and classification trends within and among global malaria-endemic regions.

Utilizing the function of GC_3 to extract all positions (i.e. interval = 1, step size = 1) and plotting *hrp2/3* coding positions only, coverage is clearer and subsets of samples can be compared, if desired. The greater difference in median relative coverage was visualized between the two sample sets collected in Bandiagara, Mali. Samples collected in 2002 had lower standardized coverage than those collected in 2010, in both *hrp2* and *hrp3*. Were these samples comparable, this result would be unexpected, since the frequency of deletions is expected to have increased over time, due to the selection imposed by parasite detection by RDT. Interestingly, *hrp2* deletions had already been observed in Mali in the late 1990s prior to significant RDT use in the country, indicative of deletion recurrence or persistence in the population at the time despite the absence of pressure from RDT usage [48]. In fact, random polymorphisms occur naturally particularly in the subtelomeric region, including large deletions, without evolutionary pressure [49]. Further, the fitness cost associated with *hrp2* loss is not significant, although a small cost appears to be associated with *hrp3* deletions [3, 19]. However, too many potentially confounding factors exist between the two sample sets to appropriately interpret results as being indicative of a significant change in frequency of these gene deletions. Even slight differences, apparently innocuous, between sample sets, like read length (100 bp in Mali 2002 samples

and 150 bp in Mali 2010 samples), could have an impact when comparing sample sets. Other factors include how malaria-positive cases were detected, study objective (case–control study vs. cohort study), and sample collection strategy (passive vs. active). A carefully controlled study using large sample sizes would be necessary to draw definitive conclusions. Ultimately, this observation showcases GC_3's capability to visualize coverage patterns between sample sets and how other factors can impact deletion rate at these loci.

Closer examination of the proportional coverage at each *hrp2/3* position among Cambodian samples, revealed how mapping artifacts can result in no coverage, and potentially confounding results. On *hrp2*, zero coverage positions increase and then spike around the intron region which, as mentioned before, is likely due to the intron's high AT content [50] that can cause challenges for read mapping. Interestingly, there were two spikes in zero coverage positions on the *hrp3* locus. Examination of a previous whole genome sequence-based hierarchical cluster analysis of the same Cambodian samples [31] revealed that the majority of samples contributing to one or both peaks belong to the same Cambodian subpopulation and hence share a similar genetic background. Ultimately, the figure offers a useful preliminary view of the *hrp2/3* genes and their characteristics.

Some limitations exist when interpreting GC_3's results, particularly when comparing sample sets. In this case, it is critical to ensure that sample sets are directly comparable (similar sampling location, collection design and protocol, sample processing and sequencing approach, etc.) or else that interpretation of results is robust to potential confounding factors. In such studies, GC_3 is most useful when WGS data is all that is available, and biological material has been exhausted. Pertaining to GC_3's results, the quality of computational results is influenced by the depth of the *P. falciparum* sequencing data, as measured by the total number of reads mapped to the reference genome. In the case of *P. falciparum*, it is considered good coverage data if the percentage of genome with coverage asymptotes at ~ 12 million 100-bp reads mapped to the parasite genome, averaging ~ 52X coverage genome-wide. It is also recommended that GC_3 is used to calculate mean coverage over the broader region/chromosome where the target gene is located to estimate expected coverage. Further, high-quality WGS data results obtained with well-described DNA extraction methods and sequencing methods, either by direct sequencing or sWGA [22, 24], and established quality control and filtering protocols should be used when comparing samples from different studies. When comparing standardized coverage between direct and sWGA sequence data, direct sequencing achieves more uniform coverage due to the inefficient amplification on the subtelomeric region by sWGA primers [24], but sWGA still provides good coverage at *hrp2/3* gene positions (Additional file 1: Figure S6). Even with high-quality data, polyclonality adds another layer of complexity, especially in high transmission settings, where these are most common, since the presence of multiple strains can mask the lack of coverage at a target gene absent in some but not all strains [18]. This factor would need to be considered particularly if GC_3 is the only method being used to assess for the presence of deletions. The deletion criteria can be easily adjusted by the user to be more stringent depending on their purposes, much like the deletion criteria used in the comparator study [19]. Despite the limitations, GC_3 appropriately processed *hrp2/3* coverage data and classified deletions. Its utility can be extended to analyse and visualize coverage data of any target gene on any pathogen.

Summary statement

Overall, validation of GC_3 to extract and process WGS data was successful when comparing with expected results using reference strains, well-described samples and a *hrp2*/3-specific qPCR assay. Following the criteria for identifying deletions, the results agreed with previous estimations of *hrp2/3* deletion frequency in each respective country. Apparent in the results is the level of detail that can be extracted from short-read WGS data and viewed using a comprehensive computational tool. Although challenges persist in ensuring high-quality WGS data and achieving similar coverage among low parasitaemia samples using sWGA, GC_3's results are expected to be fairly robust. Further investigation of the partial deletions threshold that results in a false negative RDT diagnosis is needed to validate the deletion criteria. Ultimately, groups investigating a target gene's coverage can use GC_3 to efficiently generate translatable results and figures to understand and interpret broad patterns using hundreds to thousands of previously generated genomic datasets.

Abbreviations

NMCP: National Malaria Control Programme; RDTs: Rapid Diagnostic Tests; *hrp2/3*: Loci encoding histidine-rich proteins 2 and 3; WGS: Whole Genome Sequencing; sWGA: Selective Whole Genome Amplification; SRA: Sequence Read Archive; BAM: Binary sequence Alignment/Map; GATK: Genome Analysis Toolkit; BED: Browser Extensible Data; bp: Base pair; CDS: Coding DNA sequence.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12936-022-04376-3.

Additional file 1: Figure S1. Distribution of global isolates including reference strains NF54 (West Africa), 7G8 (Brazil), NF135.C10 (Cambodia), NF166 (Guinea), DD2 (Laos) and HB3 (Honduras). Figure created using Mapchart.net. **Figure S2.** Subtelomeric read coverage distribution plots. **Figure S3.** Cambodian sample subset of hrp2/3 and flanking genes. **Figure S4.** Scatter plots of subtelomeric mean coverage vs. mean coverage of respective HRP-encoding locus. **Figure S5.** Scatter plots of mean downstream/upstream gene coverage vs. mean coverage of respective HRP-encoding locus. **Figure S6.** Median standardized coverage by direct sequencing vs sWGA. **Figure S7.** Proportion of hrp2/3 positions with 0X vs ≥1X coverage for Malawi and Mali samples. **Table S1.** List of samples that underwent hrp2/3-specific qPCR assay. **Table S2.** GC3 deletion assignments for hrp2/3 per country.

Additional file 2: Table S3. Summaries of Cambodian sample gene and chromosomal coverages grouped by year of collection. **Table S4.** Summaries of Malawian sample gene and chromosomal coverages grouped by year of collection. **Table S5.** Summaries of Malian sample gene and chromosomal coverages grouped by year of collection.

Additional file 3. Samples used for analyses in the manuscript ""Gene Coverage Count and Classification" (GC3), a coverage assessment tool, and its application to identify and classify histidine-rich protein 2 and 3 deletions in Plasmodium falciparum using short-read whole genome sequencing data".

Acknowledgements
We would like to thank the study teams that supported the collection of samples in the Brazil, Malawi, Mali, and Myanmar and Thailand studies. Specifically, we would like to acknowledge the generosity of the UMB PIs from each respective study, namely Dr. Miriam Laufer, Dr. Shannon Takala-Harrison and Dr. Mark Travassos, as well as the University of São Paulo PI, Dr. Marcelo Urbano Ferreira. Parasite isolate DNA from Cambodia and Thailand was derived from clinical samples collected by the Department of Bacterial and Parasitic Diseases at the Armed Forces Research Institute of Medical Sciences in collaboration with Dr. Shannon Takala-Harrison and had undergone whole genome sequencing as part of a previous study [51]. Malawian isolates were generated from samples by the Mfera Health Center in the Chikhwawa district from a previous investigation [52]. Malian samples were collected by the Bandiagara Malaria Project, led by Professor Mahamadou A. Thera. Dr. Drissa Coulibaly supervised the sample collection [53]. Brazilian samples were collected by the University of São Paulo Department of Public Health Practices [22]. Accession numbers for the samples are provided in Additional file 1: Table S1.

Author contributions
TCS conceived the study, designed the bioinformatics toolkit, performed the analyses, interpreted the results and wrote the paper. AD supported the development of GC_3, contributed to the interpretation of results, and edited the manuscript. JCS conceived and designed the study and bioinformatics toolkit, interpreted the results, contributed funding and resources, and wrote the paper. BS and SJ were responsible for all laboratory procedures and edited the paper. TS designed the *hrp2/3*-specific qPCR protocol and edited the paper. AO contributed to the interpretation of results and edited the paper. GG and CD contributed fundamental resources to the study and reviewed the paper. All authors read and approved the final version of the manuscript.

Funding
This work was funded in part by the National Institutes of Health (NIH) awards U19 AI110820 and R01 AI141900 to JCS. AO received funding from the National Heart, Lung and Blood Institute (1K01HL140285-01A1).

Availability of data and materials
The datasets from field studies are available from public repositories, with access information reported previously [22].

Declarations

Ethics approval and consent to participate
All WGS datasets used are publicly available through GenBank. ***P. falciparum*** samples for which DNA material was utilized to perform qPCR were collected during field studies in Brazil, Malawi, Mali, and Thailand. All samples were collected after recruitment/enrollment and with written informed consent from the subject or a parent/guardian, and assent obtained from children under 18 years old. Informed consent included authorization to generate parasite genomic data. The respective field studies were approved by the following Institutional review boards (IRBs): Institute of Biomedical Sciences, University of São Paulo (1368/17/CEPSH), for collection in Acre, Brazil; Malawi National Health Sciences Research Committee, and the IRB at the University of Maryland, Baltimore (FWA#00005976), for collection in Malawi; Comité D'Éthique de la FMPOS (Faculté de Médecine de Pharmacie et d'Odonto-stomatologie), Université des Sciences, des Techniques, et des Technologies de Bamako, Mali, and the IRB at the University of Maryland, Baltimore (HP-00041382), for collection on Mali; Ethics Review Committee of the Department of Medical Research, Human Subject Protection Branch (HSPB) of the Walter Reed Army Institute of Research, Silver Spring, MD (FWA# 00000015, IRB# 00000794), and Institute for the Development of Human Research Protection (IHRP), Ministry of Public Health, Thailand (FWA# 00017503, IRB#00006539), for collections in Thailand. This study conforms to the principles established in the Helsinki Declaration.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland. [2]University of Basel, Basel, Switzerland. [3]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA. [4]Malaria Research Program, Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA. [5]Medical Care Development International, Silver Spring, MD, USA. [6]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA.

Received: 15 July 2022 Accepted: 10 November 2022
Published online: 29 November 2022

References

1. WHO. World malaria report 2021. Geneva: World Health Organization; 2021.
2. Mukkala AN, Kwan J, Lau R, Harris D, Kain D, Boggild AK. An update on malaria rapid diagnostic tests. Curr Infect Dis Rep. 2018;20:49.
3. Poti KE, Sullivan DJ, Dondorp AM, Woodrow CJ. HRP2: transforming malaria diagnosis, but with caveats. Trends Parasitol. 2020;36:112–26.
4. Rock EP, Marsh K, Saul AJ, Wellems TE, Taylor DW, Maloy WL, et al. Comparative analysis of the *Plasmodium falciparum* histidine-rich proteins HRP-I, HRP-II and HRP-III in malaria parasites of diverse origin. Parasitology. 1987;95:209–27.
5. Kong A, Wilson SA, Ah Y, Nace D, Rogier E, Aidoo M. HRP2 and HRP3 cross-reactivity and implications for HRP2-based RDT use in regions with *Plasmodium falciparum* hrp2 gene deletions. Malar J. 2021;20:207.
6. Thomson R, Parr JB, Cheng Q, Chenet S, Perkins M, Cunningham J. Prevalence of *Plasmodium falciparum* lacking histidine-rich proteins 2 and 3: a systematic review. Bull World Health Organ. 2020;98:558-568F.
7. Kozycki CT, Umulisa N, Rulisa S, Mwikarago EI, Musabyimana JP, Habimana JP, et al. False-negative malaria rapid diagnostic tests in Rwanda: impact of *Plasmodium falciparum* isolates lacking hrp2 and declining malaria transmission. Malar J. 2017;16:123.

8. Gendrot M, Fawaz R, Dormoi J, Madamet M, Pradines B. Genetic diversity and deletion of *Plasmodium falciparum* histidine-rich protein 2 and 3: a threat to diagnosis of P falciparum malaria. Clin Microbiol Infect. 2019;25:580–5.
9. Agaba BB, Yeka A, Nsobya S, Arinaitwe E, Nankabirwa J, Opigo J, et al. Systematic review of the status of pfhrp2 and pfhrp3 gene deletion, approaches and methods used for its estimation and reporting in *Plasmodium falciparum* populations in Africa: review of published studies 2010–2019. Malar J. 2019;18:355.
10. Akinyi Okoth S, Abdallah JF, Ceron N, Adhin MR, Chandrabose J, Krishnalall K, et al. Variation in *Plasmodium falciparum* Histidine-Rich Protein 2 (Pfhrp2) and *Plasmodium falciparum* Histidine-Rich Protein 3 (Pfhrp3) Gene Deletions in Guyana and Suriname. PLoS ONE. 2015;10: e0126805.
11. Berzosa P, González V, Taravillo L, Mayor A, Romay-Barja M, García L, et al. First evidence of the deletion in the pfhrp2 and pfhrp3 genes in *Plasmodium falciparum* from Equatorial Guinea. Malar J. 2020;19:99.
12. Beshir KB, Sepúlveda N, Bharmal J, Robinson A, Mwanguzi J, Busula AO, et al. *Plasmodium falciparum* parasites with histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in two endemic regions of Kenya. Sci Rep. 2017;7:14718.
13. Funwei R, Nderu D, Nguetse CN, Thomas BN, Falade CO, Velavan TP, et al. Molecular surveillance of pfhrp2 and pfhrp3 genes deletion in *Plasmodium falciparum* isolates and the implications for rapid diagnostic tests in Nigeria. Acta Trop. 2019;196:121–5.
14. Kumar N, Pande V, Bhatt RM, Shah NK, Mishra N, Srivastava B, et al. Genetic deletion of HRP2 and HRP3 in Indian *Plasmodium falciparum* population and false negative malaria rapid diagnostic test. Acta Trop. 2013;125:119–21.
15. Thomson R, Beshir KB, Cunningham J, Baiden F, Bharmal J, Bruxvoort KJ, et al. pfhrp2 and pfhrp3 gene deletions that affect malaria rapid diagnostic tests for *Plasmodium falciparum*: analysis of archived blood samples from 3 African countries. J Infect Dis. 2019;220:1444–52.
16. Verma AK, Bharti PK, Das A. HRP-2 deletion: a hole in the ship of malaria elimination. Lancet Infect Dis. 2018;18:826–7.
17. Molina-de la Fuente I, Pastor A, Herrador Z, Benito A, Berzosa P. Impact of *Plasmodium falciparum* pfhrp2 and pfhrp3 gene deletions on malaria control worldwide: a systematic review and meta-analysis. Malar J. 2021;20:276.
18. Watson OJ, Slater HC, Verity R, Parr JB, Mwandagalirwa MK, Tshefu A, et al. Modelling the drivers of the spread of *Plasmodium falciparum* hrp2 gene deletions in sub-Saharan Africa. Life. 2017;6:e25008.
19. Sepúlveda N, Phelan J, Diez-Benavente E, Campino S, Clark TG, Hopkins H, et al. Global analysis of *Plasmodium falciparum* histidine-rich protein-2 (pfhrp2) and pfhrp3 gene deletions using whole-genome sequencing data and meta-analysis. Infect Genet Evol. 2018;62:211–9.
20. Gibbons J, Qin J, Malla P, Wang Z, Brashear A, Wang C, et al. Lineage-specific expansion of *Plasmodium falciparum* parasites with pfhrp2 deletion in the Greater Mekong Subregion. J Infect Dis. 2020;222:1561–9.
21. Feleke SM, Reichert EN, Mohammed H, Brhane BG, Mekete K, Mamo H, et al. Emergence and evolution of *Plasmodium falciparum* histidine-rich protein 2 and 3 deletion mutant parasites in Ethiopia. medRxiv. 2021.
22. Moser KA, Drábek EF, Dwivedi A, Stucke EM, Crabtree J, Dara A, et al. Strains used in whole organism *Plasmodium falciparum* vaccine trials differ in genome structure, sequence, and immunogenic potential. Genome Med. 2020;12:6.
23. Malaria GEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. eLife. 2016;5:e08714.
24. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, et al. Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. Malar J. 2016;15:597.
25. Zhang M, Wang C, Otto TD, Oberstaller J, Liao X, Adapa SR, et al. Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. Science. 2018;360:e7847.
26. Schindler T, Deal AC, Fink M, Guirou E, Moser KA, Mwakasungula SM, et al. A multiplex qPCR approach for detection of pfhrp2 and pfhrp3 gene deletions in multiple strain infections of *Plasmodium falciparum*. Sci Rep. 2019;9:13107.
27. Abdallah JF, Okoth SA, Fontecha GA, Torres REM, Banegas EI, Matute ML, et al. Prevalence of pfhrp2 and pfhrp3 gene deletions in Puerto Lempira. Honduras Malar J. 2015;14:19.
28. Munro JB, Jacob CG, Silva JC. A novel clade of unique eukaryotic ribonucleotide reductase R2 subunits is exclusive to apicomplexan parasites. J Mol Evol. 2013;77:92–106.
29. Xiao C, Ye J, Esteves RM, Rong C. Using Spearman's correlation coefficients for exploratory data analysis on big dataset. Concurrency Computation. 2016;28:3866–78.
30. Rachid Viana GM, Akinyi Okoth S, Silva-Flannery L, Lima Barbosa DR, Macedo de Oliveira A, Goldman IF, et al. Histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in *Plasmodium falciparum* isolates from select sites in Brazil and Bolivia. PLoS ONE. 2017;12:e0171150.
31. Agrawal S, Moser KA, Morton L, Cummings MP, Parihar A, Dwivedi A, et al. Association of a novel mutation in the *Plasmodium falciparum* chloroquine resistance transporter with decreased piperaquine sensitivity. J Infect Dis. 2017;216:468–76.
32. Laufer MK, Thesing PC, Dzinjalamala FK, Nyirenda OM, Masonga R, Laurens MB, et al. A longitudinal trial comparing chloroquine as monotherapy or in combination with artesunate, azithromycin or atovaquone-proguanil to treat malaria. PLoS ONE. 2012;7: e42284.
33. Lyke KE, Burges RB, Cissoko Y, Sangare L, Kone A, Dao M, et al. HLA-A2 supertype-restricted cell-mediated immunity by peripheral blood mononuclear cells derived from malian children with severe or uncomplicated *Plasmodium falciparum* malaria and healthy controls. Infect Immun. 2005;73:5799–808.
34. Travassos MA, Coulibaly D, Laurens MB, Dembélé A, Tolo Y, Koné AK, et al. Hemoglobin C trait provides protection from clinical falciparum malaria in Malian children. J Infect Dis. 2015;212:1778–86.
35. Golassa L, Messele A, Amambua-Ngwa A, Swedberg G. High prevalence and extended deletions in *Plasmodium falciparum* hrp2/3 genomic loci in Ethiopia. PLoS ONE. 2020;15: e0241807.
36. Berhane A, Anderson K, Mihreteab S, Gresty K, Rogier E, Mohamed S, et al. Major threat to malaria control programs by *Plasmodium falciparum* lacking histidine-rich protein 2. Eritrea Emerg Incet Dis. 2018;24:462–70.
37. Gamboa D, Ho M-F, Bendezu J, Torres K, Chiodini PL, Barnwell JW, et al. A large proportion of P falciparum isolates in the Amazon Region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. PLoS ONE. 2010;5:e8091.
38. Walliker D, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, London WT, et al. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. Science. 1987;236:1661–6.
39. Preston MD, Campino S, Assefa SA, Echeverry DF, Ocholla H, Amambua-Ngwa A, et al. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. Nat Commun. 2014;5:4052.
40. Góes L, Chamma-Siqueira N, Peres JM, Nascimento JM, Valle S, Arcanjo AR, et al. Evaluation of histidine-rich proteins 2 and 3 gene deletions in *Plasmodium falciparum* in endemic areas of the Brazilian Amazon. Int J Environ Res Public Health. 2020;18:123.
41. Bhasin VK, Trager W. Gametocyte-forming and non-gametocyte-forming clones of *Plasmodium falciparum*. Am J Trop Med Hyg. 1984;33:534–7.
42. Vera-Arias CA, Holzschuh A, Oduma CO, Badu K, Abdul-Hakim M, Yukich J, et al. *Plasmodium falciparum* hrp2 and hrp3 gene deletion status in Africa and South America by highly sensitive and specific digital PCR. medRxiv. 2021.
43. Wurtz N, Fall B, Bui K, Pascual A, Fall M, Camara C, et al. Pfhrp2 and pfhrp3 polymorphisms in *Plasmodium falciparum* isolates from Dakar, Senegal: impact on rapid malaria diagnostic tests. Malar J. 2013;12:34.
44. Houzé S, Hubert V, Pessec GL, Bras JL, Clain J. Combined deletions of pfhrp2 and pfhrp3 genes result in *Plasmodium falciparum* malaria false-negative rapid diagnostic test. J Clin Microbiol. 2011;49:2694–6.
45. Nima MK, Hougard T, Hossain ME, Kibria MG, Mohon AN, Johora FT, et al. A case of *Plasmodium falciparum* hrp2 and hrp3 gene mutation in Bangladesh. Am J Trop Med Hyg. 2017;97:1155–8.
46. Li P, Xing H, Zhao Z, Yang Z, Cao Y, Li W, Yan G, et al. Genetic diversity of *Plasmodium falciparum* histidine-rich protein 2 in the China-Myanmar border area. Acta Trop. 2015;152:26–31.
47. Parr JB, Verity R, Doctor SM, Janko M, Carey-Ewend K, Turman BJ, et al. Pfhrp2-deleted *Plasmodium falciparum* parasites in the Democratic Republic of the Congo: a national cross-sectional survey. J Infect Dis. 2016;216:36–44.
48. Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the

histidine-rich repeat region of the hrp2 gene. Am J Trop Med Hyg. 2012;86:194–8.
49. de Bruin D, Lanzer M, Ravetch JV. The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. Proc Natl Acad Sci USA.1994;91:619–623.
50. Hamilton WL, Claessens A, Otto TD, Kekre M, Fairhurst RM, Rayner JC, et al. Extreme mutation bias and high AT content in *Plasmodium falciparum*. Nucleic Acids Res. 2016;45:1889–901.
51. Takala-Harrison S, Jacob CG, Arze C, Cummings MP, Silva JC, Dondorp AM, et al. Independent emergence of artemisinin resistance mutations among *Plasmodium falciparum* in Southeast Asia. J Infect Dis. 2014;211:670–9.
52. Buchwald AG, Sixpence A, Chimenya M, Damson M, Sorkin JD, Wilson ML, et al. Clinical implications of asymptomatic *Plasmodium falciparum* infections in Malawi. Clin Infect Dis. 2018;68:106–12.
53. Coulibaly D, Travassos MA, Kone AK, Tolo Y, Laurens MB, Traore K, et al. Stable malaria incidence despite scaling up control strategies in a malaria vaccine-testing site in Mali. Malar J. 2014;13:374.

Publisher's Note

3 TRANSMISSION DYNAMICS OF *PLASMODIUM FALCIPARUM* BETWEEN BIOKO ISLAND, AND MAINLAND EQUATORIAL GUINEA

This chapter contains the following publication:

García GA, Janko M, Hergott DE, Donfack OT, Smith JM, Mba Eyono JN, DeBoer KR, Nguema Avue RM, Phiri WP, Aldrich EM, Schwabe C, **Stabler TC**, Rivas MR, Cameron E, Guerra CA, Cook J, Kleinschmidt I, Bradley J. **Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions**. *Malaria journal*. 2023 Mar 1;22(1):72.

Stabler TC, Dwivedi A, Shrestha B, Sudhaunshu J, Rivas MR, Donfack OT, Guerra CA, García GA, Daubenberger C, Silva JC. **Whole genome sequences of *Plasmodium falciparum* on Bioko Island, Equatorial Guinea illustrate connectedness with geographical neighbors and island-wide partial population structure.** Manuscript in preparation. Planned submission to mBio.

Stabler TC, Hosch S, Nyakarungu E, Giger JN, Elonga MK, Bibang RN, Ndong VM, Mometolo IE, Bolopa AL, Bedoung RMO, Bijeri JR, Roka EW, Nguema AG, Ichinda VP, Bakale S, Esokolete CN, Rivas MR, Galick DS, Phiri WP, Guerra CA, García GA, Montemartini L, Schindler T, Silva JC, Daubenberger CA. **Integrating local malaria molecular monitoring into regular malaria indicator surveys on Bioko Island: high association between urban communities and low density infections.** Manuscript in preparation. Planned submission to Malaria Journal.

3.1 Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions.

Published in Malaria Journal, 2023

Malaria Journal

RESEARCH

Open Access

Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions

Guillermo A. García[1*], Mark Janko[2], Dianna E. B. Hergott[3], Olivier T. Donfack[4], Jordan M. Smith[1], Jeremías Nzamío Mba Eyono[4], Kylie R. DeBoer[1], Restituto Mba Nguema Avue[4], Wonder P. Phiri[4], Edward M. Aldrich[5], Christopher Schwabe[5], Thomas C. Stabler[6,7], Matilde Riloha Rivas[8], Ewan Cameron[9], Carlos A. Guerra[1], Jackie Cook[10], Immo Kleinschmidt[10,11] and John Bradley[10]

Abstract

Background Since 2004, malaria transmission on Bioko Island has declined significantly as a result of the scaling-up of control interventions. The aim of eliminating malaria from the Island remains elusive, however, underscoring the need to adapt control to the local context. Understanding the factors driving the risk of malaria infection is critical to inform optimal suits of interventions in this adaptive approach.

Methods This study used individual and household-level data from the 2015 and 2018 annual malaria indicator surveys on Bioko Island, as well as remotely-sensed environmental data in multilevel logistic regression models to quantify the odds of malaria infection. The analyses were stratified by urban and rural settings and by survey year.

Results Malaria prevalence was higher in 10–14-year-old children and similar between female and male individuals. After adjusting for demographic factors and other covariates, many of the variables investigated showed no significant association with malaria infection. The factor most strongly associated was history of travel to mainland Equatorial Guinea (mEG), which increased the odds significantly both in urban and rural settings (people who travelled had 4 times the odds of infection). Sleeping under a long-lasting insecticidal net decreased significantly the odds of malaria across urban and rural settings and survey years (net users had around 30% less odds of infection), highlighting their contribution to malaria control on the Island. Improved housing conditions indicated some protection, though this was not consistent across settings and survey year.

Conclusions Malaria risk on Bioko Island is heterogeneous and determined by a combination of factors interacting with local mosquito ecology. These interactions grant further investigation in order to better adapt control according to need. The single most important risk factor identified was travel to mEG, in line with previous investigations, and represents a great challenge for the success of malaria control on the Island.

Keywords Malaria, Targeted interventions, Importation, Vector control, Risk factors

*Correspondence:
Guillermo A. García
ggarcia@mcd.org
Full list of author information is available at the end of the article

Background

Bioko is the largest Island of Equatorial Guinea (EG), with an area of 2,017 km2. It is located in the Bight of Biafra, off the coast of Cameroon and has an approximate population of around 270,000 people [1]. (Fig. 1). Most of the population lives on the north side of the Island, near and within Malabo, the country capital. Historically, Bioko was hyperendemic for malaria transmission, with some of the highest entomological inoculation rates (EIR) ever recorded: over one thousand infective bites per person per year, 281 for *Anopheles gambiae* sensu lato (*s.l.*) and 787 for *Anopheles funestus* [2–4] This translated into a pre-intervention *Plasmodium falciparum* parasite rate (PfPR) of 45% in 2–14-year-old children [5].

In 2004, the Bioko Island Malaria Elimination Project (BIMEP) was established and has since successfully reduced the malaria burden [6, 7]. PfPR in 2–14-year-old children has declined by about 75%, and all-cause mortality and anaemia in under 5-year-olds by 63% and 86%. Another critical benchmark of the project was the elimination of two of the principal vectors, *An. funestus* and *An. gambiae* sensu stricto, from the Island [2–4], leaving

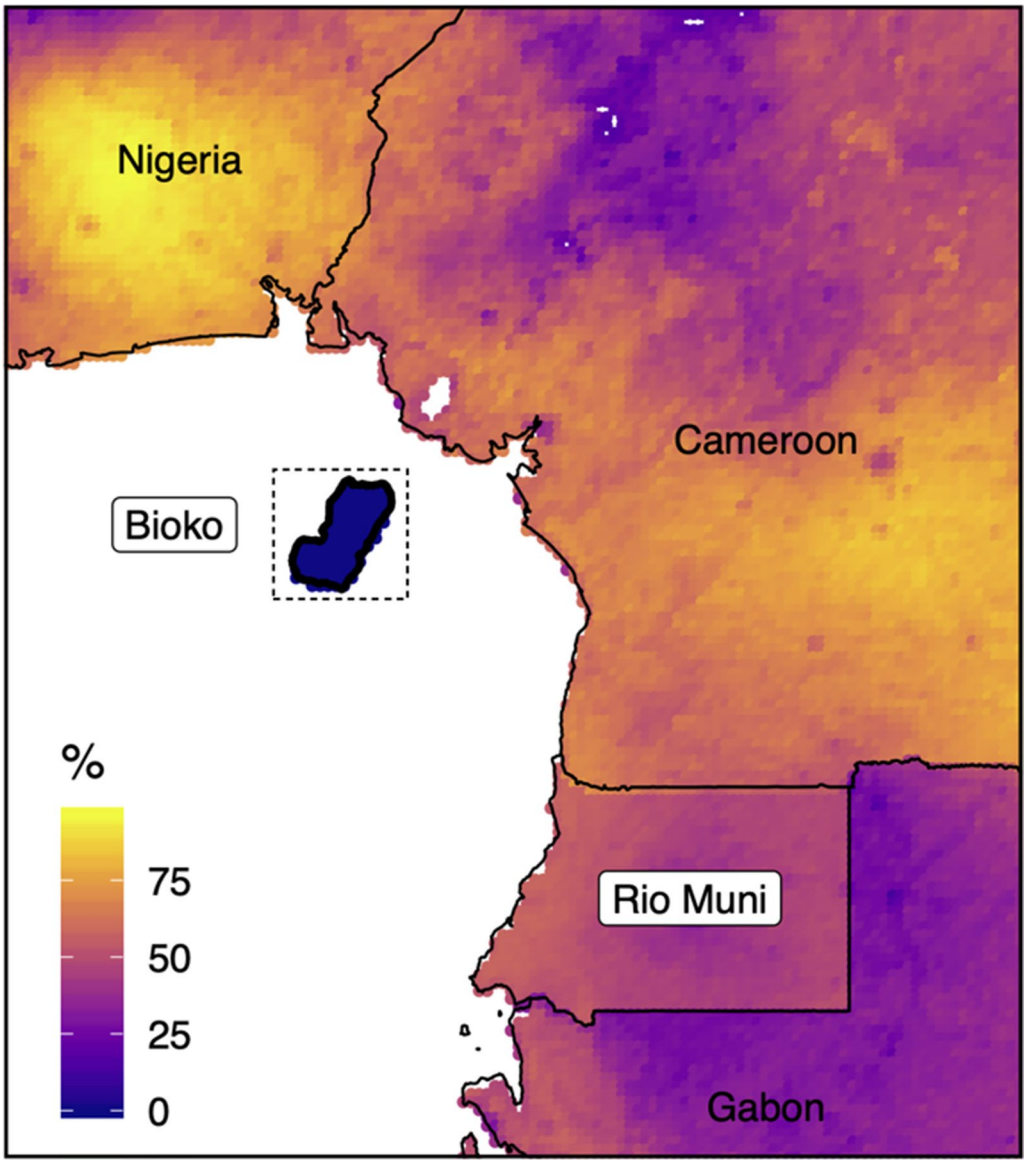


Fig. 1 Bioko Island and its location in the Gulf of Guinea. The continental territory of Equatorial Guinea is known as Río Muni. The color scale represents predicted *Pf*PR in children, reconstructed from data produced by the Malaria Atlas Project, which are available for use under the Creative Commons Attribution 3.0 Unported License

only An. coluzzii and An. Melas [8], which sustain an EIR in order of magnitude lower than in pre-intervention times. PfPR in the lowest year on record (2016) was between 10 and 15% across the Island (Fig. 2), substantially lower than in the surrounding neighbours (Fig. 1). Despite these gains, several areas on the Island have seen a resurgence of malaria similar to other endemic regions across Africa (Fig. 2) [9].

At the inception of the BIMEP, whole-island indoor residual spraying (IRS) was the primary vector control strategy supplemented by limited long-lasting insecticidal net (LLIN) distributions. In the early stages of the Project, it was believed that IRS could rid Bioko of malaria in a few years, but this was soon realized as an overly optimistic goal and this led to attempts to adapt strategies in order to sustain gains. Starting in 2015, two triennial LLIN mass distribution campaigns (MDC) were conducted, one in 2015 and one in 2018, which came to represent the main vector control intervention. LLINs were supported by targeted IRS in high prevalence areas together with intermittent preventive treatment for pregnant women, case management using artemisinin-based combination therapy (ACT), diagnosis through training in microscopy and Rapid Diagnostic Tests (RDTs), and larval source management [6, 10].

A lingering question is whether targeting interventions or universal coverage of the population is the best approach to control malaria while making optimal use of limited resources [11]. Targeting strategies have used statistical models to identify malaria hotspots based on malaria prevalence and malaria incidence data, demographic factors, and transmission etiology [12–19]. Targeting of IRS has proven particularly unsuccessful on Bioko, however, as malaria hotspots represent moving spatial targets around the Island. Therefore, a comprehensive and regularly updated understanding of intervention effectiveness and risk factors for malaria and how they may vary by context, and over time, would prove critical to any targeting strategy. Here, the study evaluated individual, household, and environmental risk factors on Bioko to improve the understanding of some of the main drivers of malaria transmission on the Island.

Methods

MIS Data

The study used individual and household-level data from the 2015 and 2018 annual malaria indicator surveys (MIS). The sampling frame for the MIS was drawn using a comprehensive database that uniquely identifies and geo-references all households on Bioko Island [20]. Clusters were defined by single communities containing at least 20 households and, in the case of sparsely populated communities, an aggregation of communities. A random sample of households was then selected from these clusters comprising 7% of the total island household universe in 2015 and 6% in 2018. Consenting household members present during the survey were tested for *P. falciparum* using RDTs (CareStart G0131 Combo kit, AccessBio Inc., Monmouth, USA). In the case of children, consent was sought from their guardians.

The data on individuals, households, and covariates collected during the MIS and explored in this study are listed in Table 1. Individual-level data used in the analyses included age, sex, history of recent travel to mainland EG (mEG), and LLIN use the previous night. Household-level data included socioeconomic status (SES), household density, and housing characteristics. To derive SES, households were assigned scores based on the type of assets and amenities they own (radio, television, sofa, fan, air-conditioning, car, among others)

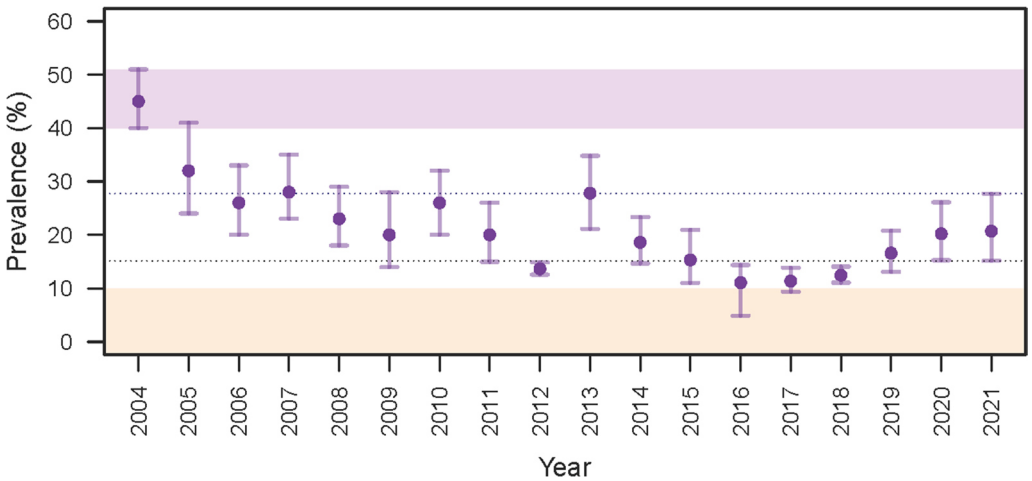


Fig. 2 Malaria prevalence in children 2–14 years old on Bioko Island, 2004–2021

Table 1 Individual, Household, and Environmental covariates investigated in the study

Individual	Source	Included in Model
Age	Malaria Indicator Survey	x
Gender		x
Household size		x
Household density		x
Travel History		x
Socioeconomic Status		x
LLIN use		x
Household		
Wall Type	Malaria Indicator Survey	
Wall Gaps		
Roof Type		
Roof Types		
Floor Types		
Air Conditioner ownership		x
Window Glass		
Window Screen		
Door Screen		
Eaves		x
Water Source		
Light Source		
Toilet Type		
Environmental		
Elevation	Shuttle Radar Topography Mission	x
Slope		x
Size of community	BIMEP Mapping System	
Distance to water streams		x
Distance to military camps		x
Population density		
Tasseled Cap Brightness (TCB)	Nasa Earth Data	x
Tasseled Cap Wetness (TCW)		x
Land Surface Temperature (LST)		x
Enhanced Vegetation Index (EVI)		x

using principal component analysis (PCA). Households were ranked based on their score and divided into quintiles. The first quintile corresponded to the lowest wealth index (WI) and the fifth to the highest WI. Household density was defined as the ratio of the number of household occupants to the number of rooms in the house and was classified into five categories: very low (≤ 0.5), low ($>0.5-<1$), medium ($\geq 1-<1.5$), high ($\geq 1.5-<2$) and very high (≥ 2). Housing characteristics included whether the house had open eaves or air conditioning in place and distance to specific geographical features of interest, namely bodies of water and military camps. Military camps may be important in this context because military personnel represent a source of human infectivity due to frequent travel to mEG [21].

Environmental data

Remote sensing-derived environmental data included the Tasseled Cap Brightness (TCB), the Tasseled Cap Wetness (TCW) [22], Land Surface Temperature (LST) [22], the Enhanced Vegetation Index (EVI) [23], and elevation and slope derived from the Shuttle Radar Topography Mission [24]. Information on how TCB, TCW, LST and EVI are compiled and processed is described elsewhere [7, 25].

Statistical analysis

The primary outcome of interest was individual-level *P. falciparum* malaria infection, as determined by the RDT result reported in the MIS. The covariates listed in Table 1 were investigated for their association with malaria infection using multilevel logistic regression, controlling for potential confounders. To account for the survey design, a random intercept from the MIS survey cluster was included and modeled for its random effect. To account for variability due to differing operational strategies deployed in urban and rural settings, the models were stratified to observe their associations separately. Finally, because the study aimed at identifying the persistence of risk factors across time, the analysis was also stratified by survey year (2015 vs. 2018). All continuous covariates were centered and scaled such that inferences were based on a one standard deviation change in the covariate of interest. All statistical analyses were run in Stata version 16 [26].

Results

Descriptive statistics

In the 2015 and 2018 MIS, 17,016 and 13,906 individuals were tested for malaria. However, some of these individuals had incomplete survey information and were excluded from the analyses. As a result, 16,903 individuals living in 5184 households surveyed in 2015 and 13,734 individuals living in 4762 households surveyed in 2018 were included. Overall, 55.2% of surveyed individuals were female and 84.6% lived in urban areas. The median age of survey participants was 14 years. Roughly a quarter of individuals lived in each of the SES quintiles, with the most in the poorest (26.5%) and the least in the wealthiest (23.4%). Most people lived in medium density households (38.7%), though almost 8% lived in very high density households. The majority of individuals lived in households with open eaves (68.0%) and no air conditioning (81.6%). Around half reported sleeping under a net the night before the survey (55.1%) and 9.5% reported travelling to mEG within 2 months prior to the survey. Descriptive statistics of these variables by each MIS survey and urban/rural stratum are presented in Table 2.

Malaria prevalence

Prevalence of *P. falciparum* infection was 12.7% (CI 11.4–14.2, $p<0.05$) in 2015 and 10.2% (CI 9.3–11.3, $p<0.05$) in 2018. This translated into 25.5% and 20.1% of surveyed households in 2015 and 2018, respectively, having at least one *P. falciparum*-positive household member. Table 3 shows the prevalence of malaria infection by risk factor and survey year. In 2015 and 2018, malaria prevalence was highest in the 10–14 years age group at 21.8% and 16.5%, respectively. Women had a lower prevalence than men in both years (11.9% vs. 13.8% in 2015 and 9.7% vs. 10.9% in 2018, respectively). In 2015 malaria prevalence was significantly higher in rural (17.0%) compared to urban areas (12.0%), though no significant difference was observed in 2018 (10.9% in urban vs 10.1% in rural areas). People living in very low-density households (8.6% in 2015 and 7.7% in 2018) had significantly lower prevalence than those living in very high-density households (15.9% in 2015 and 15.7% in 2018). People with a history of travel to continental EG had more than double the prevalence of those who did not travel (28.7% vs. 10.9% in 2015 and 30.4% vs. 9.2% in 2018).

Odds of malaria infection

Individual level factors

An association between female sex and lower odds of infection was detected in urban settings but this was not significant in rural settings. The odds of infection in 2–14-year-old children were consistently and significantly higher. In both settings, individuals who slept under an LLIN the previous night had lower infection odds than those who did not (OR 0.67, 95% CI 0.48–0.93, $p<0.05$ and OR 0.70, 95% CI 0.48–1.01, $p=0.06$ in 2015 in rural areas, and OR 0.66, 95% CI 0.58–0.75 in 2015 and OR 0.70, 95% CI 0.59–0.83, $p<0.05$ in 2018 in urban areas). In both years travel to mainland Equatorial Guinea was identified as the risk factor most strongly associated with the odds of malaria infection irrespective of setting, with those who travelled showing three to almost five times as likely to be infected than those who did not (OR 3.54, 95% CI 2.92–4.30, $p<0.05$ in 2015 and OR 4.78, 95% CI 3.70–6.17, $p<0.05$ in 2018 in urban areas, and OR 3.69, 95% CI 2.34–5.82, $p<0.05$ in 2015 and OR 3.28, 95% CI 1.82–5.90, $p<0.05$ in 2018 in rural areas; Fig. 3, Additional file 1: Table S1).

Household level factors

At the household level, multivariate regression identified no significant effect for distance to water bodies. Except for rural settings in 2015, wealth had significant protective effects against malaria infection, with individuals living in the wealthiest households having between a third and a half lower the odds of being infected ($p<0.05$), with respect to those in the poorest households (Additional file 1: Table S1). Very high household density was a significant risk factor for those living in the densest households relative to those in the least dense ones, but only in rural areas in 2015 (OR 1.93, 95% CI 1.21–3.10, $p<0.05$) and in urban areas in 2018 (OR 1.50, 95% CI 1.11–2.03, $p<0.05$). A similar inconsistency was found in the effect on individuals living furthest away from military camps, who had significantly lower odds of malaria infection only in rural settings in 2015 (OR 0.88, 95% CI 0.83–0.93,

Table 2 Descriptive statistics for individual, household, and environmental covariates included in regression analyses

Covariate	2015			2018			Total 2015 and 2018		
	Urban	Rural	All	Urban	Rural	All	Urban	Rural	All
	(n = 14,476)	(n = 2427)	16,903	(n = 11,444)	(n = 2290)	13,734	(n = 25,920)	(n = 4,717)	(n = 30,637)
Sex									
Male	6295 (43.5)	1101 (45.4)	7396 (45.4)	5167 (45.2)	1152 (50.3)	6319 (46.0)	11,462 (44.22)	2253 (47.76)	13,715 (44.77)
Female	8181 (56.5)	1326 (54.6)	9507 (54.6)	6277 (54.8)	1138 (49.7)	7415 (54.0)	14,458 (55.78)	2464 (52.24)	16,922 (55.23)
Age group									
< 5 years	3086 (21.3)	470 (19.4)	3556 (21.0)	1646 (14.4)	285 (12.5)	1931 (14.1)	4732 (18.3)	755 (16.0)	5487 (17.9)
5–9 years	1909 (13.2)	329 (13.6)	2238 (13.2)	1450 (12.7)	279 (12.2)	1729 (12.6)	3359 (13.0)	608 (12.9)	3967 (12.9)
10–14 years	1781 (12.3)	276 (11.4)	2057 (12.2)	1421 (12.4)	274 (12.0)	1695 (12.3)	3202 (12.4)	550 (11.7)	3752 (12.2)
15–19 years	1422 (9.8)	160 (6.6)	1582 (9.4)	1127 (9.8)	200 (8.7)	1327 (9.7)	2549 (9.8)	360 (7.6)	2909 (9.5)
20–29 years	3101 (21.4)	374 (15.4)	3475 (20.6)	2378 (20.8)	394 (17.2)	2772 (20.2)	5479 (21.1)	768 (16.3)	6247 (20.4)
30–39 years	1580 (10.9)	253 (10.4)	1833 (10.8)	1534 (13.4)	278 (12.1)	1812 (13.2)	3114 (12.0)	531 (11.3)	3645 (11.9)
40–49 years	731 (5.1)	219 (9.0)	950 (5.6)	642 (5.6)	154 (6.7)	796 (5.8)	1373 (5.3)	373 (7.9)	1746 (5.7)
50–59 years	480 (3.3)	166 (6.8)	646 (3.8)	419 (3.7)	163 (7.1)	582 (4.2)	899 (3.5)	329 (7.0)	1228 (4.0)
> 60 years	386 (2.7)	180 (7.4)	566 (3.4)	827 (7.2)	263 (11.5)	1090 (7.9)	1213 (4.7)	443 (9.4)	1656 (5.4)
Socioeconomic Status (Quantile)									
Poorest	3036 (21.0)	1443 (59.5)	4479 (26.5)	2616 (22.9)	1,022 (44.6)	3638 (26.5)	5652 (21.8)	2465 (52.3)	8117 (26.5)
2nd	3812 (26.3)	532 (21.9)	4344 (25.7)	2911 (25.4)	527 (23.0)	3438 (25.0)	6723 (25.9)	1059 (22.5)	7782 (25.4)
3rd	3840 (26.5)	281 (11.6)	4121 (24.4)	2980 (26.0)	454 (19.8)	3434 (25.0)	6820 (26.3)	735 (15.6)	7555 (24.7)
Wealthiest	3788 (26.2)	171 (7.0)	3959 (23.4)	2937 (25.7)	287 (12.5)	3224 (23.5)	6725 (25.9)	458 (9.7)	7183 (23.4)
Household density									
Very low density	1585 (11.0)	483 (19.9)	2068 (12.2)	1388 (12.1)	448 (19.6)	1836 (13.4)	2973 (11.5)	931 (19.7)	3904 (12.7)
Low density	4217 (29.1)	684 (28.2)	4901 (29.0)	2966 (25.9)	580 (25.3)	3546 (25.8)	7183 (27.7)	1264 (26.8)	8447 (27.6)
Medium density	6104 (42.2)	708 (29.2)	6812 (40.3)	4341 (37.9)	712 (31.1)	5053 (36.8)	10,445 (40.3)	1420 (30.1)	11,865 (38.7)
High density	1802 (12.4)	351 (14.4)	2153 (12.7)	1582 (13.8)	234 (10.2)	1816 (13.2)	3384 (13.1)	585 (12.4)	3969 (12.9)
Very high density	768 (5.3)	201 (8.3)	969 (5.7)	1167 (10.2)	316 (13.8)	1483 (10.8)	1935 (7.5)	517 (11.0)	2452 (8.0)
LLIN Use									
Yes	7681 (53.1)	1312 (54.1)	8993 (53.2)	6633 (58.0)	1251 (54.6)	7884 (57.4)	14,314 (55.2)	2563 (54.3)	16,877 (55.1)
No	6795 (46.9)	1115 (45.9)	7910 (46.8)	4811 (42.0)	1039 (45.4)	5850 (42.6)	11,606 (44.8)	2154 (45.7)	13,760 (44.9)
Travelled to continental Africa									
Yes	1933 (13.4)	134 (5.5)	2067 (12.2)	773 (6.8)	76 (3.3)	849 (6.2)	2706 (10.4)	210 (4.5)	2916 (9.5)
No	12,543 (86.6)	2293 (94.5)	14,836 (87.8)	10,671 (93.2)	2214 (96.7)	12,885 (93.8)	23,214 (89.6)	4507 (95.5)	27,721 (90.55)
House has open eaves									
Yes	10,592 (73.2)	1191 (49.1)	11,783 (69.7)	7712 (67.4)	1345 (58.7)	9057 (66.0)	18,304 (70.6)	2536 (53.8)	20,840 (68.0)
No	3884 (26.8)	1236 (50.9)	5120 (30.3)	3732 (32.6)	945 (41.3)	4677 (34.0)	7616 (29.4)	2181 (46.2)	9797 (32.0)
House has air conditioning									
Yes	2228 (15.4)	167 (6.8)	2395 (14.2)	2819 (24.6)	408 (17.8)	3227 (23.5)	5047 (19.5)	575 (12.2)	5622 (18.4)
No	12,244 (84.6)	2260 (93.1)	14,504 (85.8)	8625 (75.4)	1882 (82.2)	10,507 (76.5)	20,869 (80.5)	4142 (87.8)	25,011 (81.6)
Environmental covariates, mean (sd)									
TCB	0.52 (0.08)	0.5 (0.14)	0.52 (0.09)	0.37 (0.03)	0.38 (0.03)	0.37 (0.03)	N/a	N/a	N/a
TCW	− 0.01 (0.02)	0.03 (0.04)	0.00 (0.03)	− 0.02 (0.02)	0 (0.03)	− 0.02 (0.02)	N/a	N/a	N/a
Elevation (meters)	71.67 (77.62)	284.09 (362.75)	102.98 (172.48)	70.54 (70.21)	299.54 (384.02)	110 (190.05)	N/a	N/a	N/a

Table 2 (continued)

Covariate	2015			2018			Total 2015 and 2018		
	Urban (n = 14,476)	Rural (n = 2427)	All 16,903	Urban (n = 11,444)	Rural (n = 2290)	All 13,734	Urban (n = 25,920)	Rural (n = 4,717)	All (n = 30,637)
Temperature (C)	22.81 (3.61)	23.43 (4.43)	22.88 (3.78)	23.43 (2.12)	23.06 (2.67)	23.36 (2.23)	N/a	N/a	N/a
EVI	0.24 (0.11)	0.46 (0.11)	0.27 (0.14)	0.25 (0.13)	0.49 (0.12)	0.29 (0.16)	N/a	N/a	N/a
Slope	7.61 (4.15)	10.13 (6.33)	7.97 (4.61)	7.5 (4.07)	10.37 (6.30)	7.99 (4.65)	N/a	N/a	N/a

$p < 0.05$), as well as in the effect of air-conditioning in the home, which was also found to significantly reduce the odds only in urban areas in 2015 (OR 0.73, 95% CI 0.58–0.92, $p < 0.05$) (Fig. 3; Additional file 1: Table S1).

Environmental risk factors

Among environmental risk factors, most had non-significant effects across both urban and rural settings and between surveys. The exception was elevation, with each 100-m increase in elevation corresponding to around 20% lower the odds of malaria infection in people living in rural areas in 2015 (OR 0.78, 95% CI 0.68–0.90, $p < 0.05$) and 2018 (OR 0.79, 95% CI 0.58–1.06, $p < 0.05$) and around 50% lower in urban areas in 2015 (OR 0.48, 95% CI 0.35–0.66, $p < 0.05$) and 2018 (OR 0.55, 95% CI 0.40–0.75, $p < 0.05$). The confidence intervals for the other predictors (TCB, TCW, EVI, LST and slope) overlap with the null, precluding any strong inferences about changing risk of infection(Fig. 3; Additional file 1: Table S1).

Discussion

This study investigated the relationship between malaria infection and individual, household, and environmental factors on Bioko Island. This investigation is important to better understand some of the drivers explaining the heterogeneity of malaria prevalence on the Island after many years of successful scaling-up of control interventions. Moreover, despite the intensive and continued efforts, reduction in prevalence on Bioko stalled since 2016, and the situation has worsened more recently in several areas across the Island (Fig. 4). Therefore, understanding some of the principal factors driving infection would prove critical to informing decision-making.

Several factors were investigated but, after adjusting for other covariates and demographic factors such as age and sex, which had a significant effect, only a handful were found to significantly influence the odds of malaria infection. History of travel to mainland EG proved by far the factor most strongly associated with increased odds of infection, with between a threefold and a fivefold increase in travellers. This finding was consistent across survey years and rural and urban areas, and is in agreement with previous investigations that have looked at the impact of malaria importation and its challenge for malaria control on Bioko Island [6, 7, 21, 27]. Human mobility patterns and malaria prevalence determined that travel to mainland EG explained much of the prevalence observed in and around Malabo, where travel prevalence is higher [7]. The fact that travel had a strong effect on the odds of infection in urban and rural communities and also remained consistent over time, despite a declining proportion of travellers in 2018 compared to 2015, points to a pervasive impact of this factor on the local malaria epidemiology. These findings underscore the need for new strategies to reduce the constant flow of parasites from the mainland to the Island, and these will probably be more cost-effective if focused on targeting high-risk frequent traveller groups [7, 25, 27, 28]. Further research to collect data and parameters from the mobile population could benefit in the understanding of this demographic and the development of tailored interventions. The studies can analyse their behaviours regarding access to case management in both locations, as well as their access to the use of bed nets and other vector control methods. In addition, MIS data might be utilized to explore the travel patterns of the mobile population [29], investigate the possible sources of acquired transmission during travel, and identify the risk of parasite importation in the locations where the mobile population goes. However, reducing transmission in mainland EG would be a critical long-term solution, though economically, logistically, and politically onerous.

The analyses also showed that bed net users had around 30% lower the odds of being infected with malaria, after adjusting for other covariates. This effect was similar in urban and rural areas. The BIMEP has deployed significant efforts to provide high coverage with LLINs through two triennial MDCs in 2015 and 2018. During these, households received a number of nets based on the household occupancy, the number of sleeping areas and the number of LLINs they already owned. In 2015 and 2018, 149,097 PermaNet 2.0 (deltamethrin, Vestergaard) and 155,972 Olyset Plus (permethrin plus the synergist, piperonyl butoxide

Table 3 Prevalence of malaria infection by risk factor and malaria indicator survey year

	2015			2018		
	Malaria Prevalence %	**CI**	**p-value**	**Malaria Prevalence %**	**CI**	**p-value**
Sex						
Male	13.8	12.3–15.3	0.0005	10.9	9.8–12.1	0.0109
Female	11.9	10.5–13.5		9.7	8.6–10.8	
Age group						
< 5 years of age	8.8	7.5–10.3	< 0.0001	7.8	6.5–9.5	< 0.0001
5–9 years	17.8	15.3–20.6		13.5	11.7–15.5	
10–14 years	21.8	19.0–24.9		16.5	14.3–18.9	
15–19 years	18.3	16.1–20.8		14.8	12.9–17.3	
20–29 years	11.2	10.0–12.7		9.4	8.2–10.8	
30–39 years	8.3	6.9–9.9		8.2	7.0–9.8	
40–49 years	8.5	6.7–10.8		8.0	6.3–10.2	
50–59 years	7.7	5.6–10.7		5.4	3.8–8.1	
> 60 years of age	5.3	3.5–7.9		3.4	2.4–4.8	
Community location						
Urban	12.0	10.6–13.6	0.0090	10.9	8.3–14.3	0.6014
Rural	17.0	13.5–21.3		10.1	9.1–11.2	
Socioeconomic Status						
Poorest	13.6	11.3–16.1	0.0001	10.4	8.9–12.0	0.1809
2nd	13.6	11.8–15.5		10.8	9.5–12.4	
3rd	14.2	12.4–16.3		10.8	9.3–12.5	
Wealthiest	9.4	8.1–10.8		8.9	7.5–10.6	
Household density						
Very low density	8.6	7.1–10.4	< 0.0001	7.7	6.6–9.1	< 0.0001
Low density	10.5	8.8–12.4		8.5	7.2–9.9	
Medium density	13.8	12.2–15.5		10.3	9.1–11.6	
High density	17.1	14.8–19.7		11.6	9.2–14.5	
Very high density	15.9	12.6–19.9		15.7	13.1–18.7	
LLIN use						
Yes	10.8	9.5–12.2	< 0.0001	8.3	7.4–9.4	< 0.0001
No	14.9	13.2–16.8		12.8	11.4–14.3	
Travel to mainland, EG						
Yes	28.7	26.0–31.6	< 0.0001	30.4	25.7–35.6	< 0.0001
No	10.9	9.6–12.4		9.2	8.4–10.2	
House has open eaves						
Yes	11.9	10.7–13.3	0.0084	10.9	9.4–12.5	0.2758
No	14.6	12.4–17.2		9.9	8.9–11.1	
House has air conditioning						
Yes	9.2	7.5–11.3	0.0008	10.3	8.6–12.3	0.9576
No	13.3	11.9–14.9		10.4	9.2–11.3	
Environmental covariates	Mean (sd)			Mean (sd)		
TCB	0.52 (0.09)			0.37 (0.03)		
TCW	0.00 (0.03)			− 0.02 (0.02)		
Elevation (meters)	103 (172.49)			110 (190)		
Temperature (C)	22.88 (3.78)			23.36 (2.23)		
EVI	0.27 (0.14)			0.29 (0.16)		
Slope	7.97 (4.61)			7.99 (4.65)		

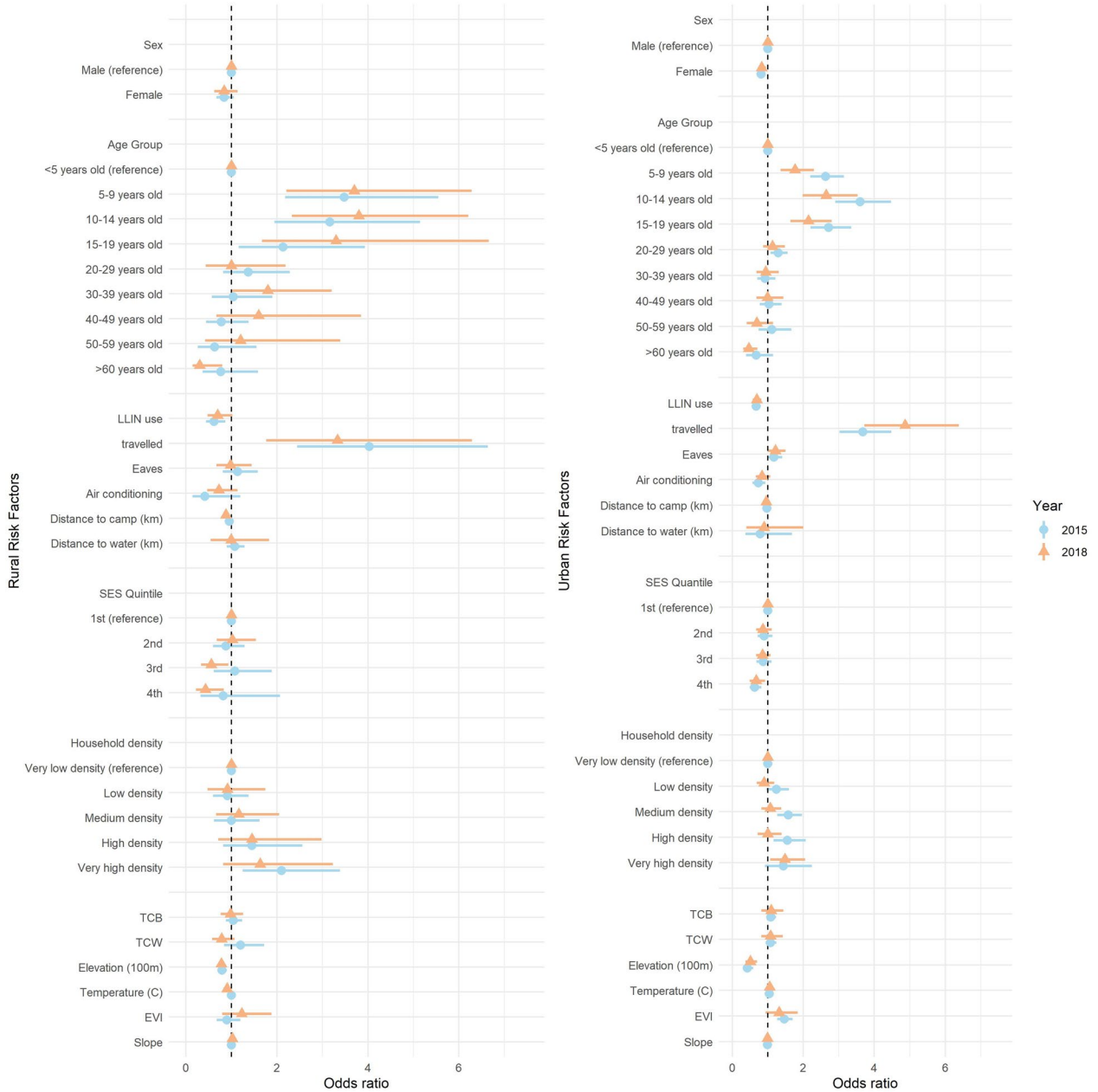


Fig. 3 Odds ratios estimated from regression models of all risk factors for malaria transmission on Bioko island and 95% confidence intervals, stratified by rural (left) and urban (right) settings and malaria annual indicator survey year, 2015 (blue) and 2018 (orange)

(PBO), Sumitomo Chemical) were distributed, respectively. Following these MDCs, universal LLIN coverage was achieved based on the two critical indicators of household ownership and population access. Mean household ownership was 90.9% (IQR: 80.0–100%) and 92.0% (IQR: 82.7–100%) and mean population access 97.3% (IQR: 93.8–100%) and 82.9% (IQR: 73.9–100%) in 2015 and 2018, respectively. Soon after distribution, however, these indicators dropped significantly, as did LLIN use, from 53.6% (IQR: 36.2–67.9%) to 36.6% (IQR: 12.7–58.1%) and 52.2% (IQR: 33.3–68.3%) to 28.7% (IQR: 14.4–45.5%) a year after each MDC (unpublished data), suggesting that, unlike many other malaria endemic settings [30] the net access-use gap on Bioko is substantial, geographically heterogeneous, and driven by socio-demographic and behavioural factors. Despite net-use decreasing significantly soon after distribution, by using data from the MIS that took place soon after

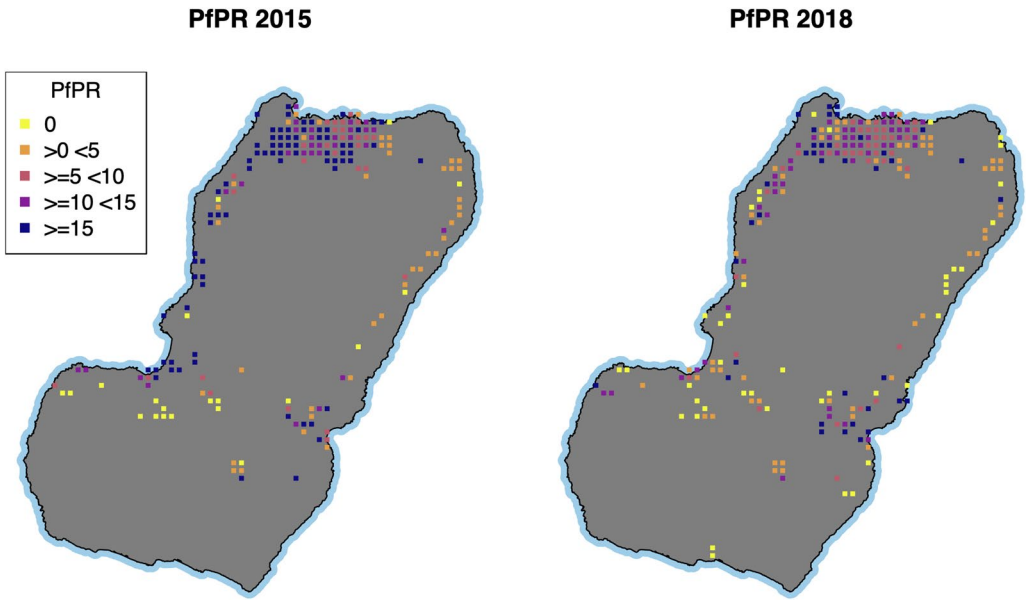


Fig. 4 Malaria prevalence on Bioko Island in 2015 and 2018. Pixels represent 1 × 1 km inhabited areas. Raw *Pf*PR data from each Malaria Indicator Survey (MIS) were interpolated using kernel smoothing

each MDC, we were able to get a reliable measure of the impact of LLIN use on reducing the risk of malaria infection, reinforcing the need to reimagining distribution and communication strategies that can improve adherence to LLINs by the population.

The effect of household-level factors on the odds of malaria infection showed a mixed picture, with confidence intervals often overlapping the null. Larger and overcrowded households were found to generally increase the odds of malaria infection after adjusting for SES, which is a likely confounder. This effect could be driven by low intrahousehold LLIN population access, but the timing of the MIS soon after each MDC again guaranteed that this indicator was still high among surveyed households. In other studies, it has been suggested that smaller density households have a lower risk because fewer people sleeping in the same room at night decreases the probability of transmission of the parasite to other hosts [31]. In nearby Cameroon, household overcrowding was identified as a major determinant of malaria mortality amongst infants [32]. A similar finding was observed for malaria prevalence amongst children in peri-urban areas of The Gambia, who had almost double the odds of malaria infection if they slept in rooms with more than three people [33]. Similarly, a study of the frequency of malaria events in children under five years old in a Sudanese village found that households with more than five people had more than double the odds of malaria infection [34]. People living in houses with a single sleeping room were also at higher risk of malaria in a highland area of Ethiopia compared to households with multiple sleeping rooms [35]. Intriguingly, the effect of household overcrowding on Bioko was not consistent across surveys and across rural and urban settings (Fig. 4). There are many households with low-occupancy on the Island, and targeting malaria control interventions to pre-identified high-risk households could prove a cost-effective malaria control strategy that needs to be further assessed. This would be logistically possible to implement using spatial decision support tools developed by the BIMEP [20, 36]. A potential caveat is that the Island's population is highly mobile [7] and residents are often away during planned interventions in their communities [36], possibly making it difficult to target and getting access to specific households for interventions.

It was also identified that certain housing conditions (*i.e.,* open eaves, absence of air-conditioning) could increase the odds of malaria infection but this effect was identified only in urban areas and only in 2015. Previous analyses had shown associations between housing characteristics and malaria prevalence on Bioko [37], suggesting that improving basic housing standards could significantly reduce the odds of malaria infection, particularly in urban areas, where open eaves were a significant risk factor [38, 39]. Household vector control interventions have been motivated by the assumption that the primary anopheline vectors on Bioko Island are endophagic and bite human hosts within their dwellings. If this were the case, then improving housing characteristics could reduce the risk of malaria infection across

different settings primarily by hampering house entry of vector mosquitoes [35, 38, 40]. In an endemic area of Sri Lanka, poor housing was linked to a 2.5-fold higher risk of malaria infection compared to residents living in houses of good construction [41]. The risk was further increased when housing was located near anopheline breeding sites. In Eswatini, adjusted models showed that people inhabiting medium and low-quality construction homes had one and a half and twice the odds of malaria infection than individuals living in high-quality constructions [39]. Open eaves and unscreened windows were also identified as significant risk factors in a hypoendemic malaria setting [42]. House screening interventions, for example, have been shown to reduce anopheline vector densities and anemia in children in homes in The Gambia [43, 44]. In Ethiopia, entomological inoculation rates, malaria prevalence, and incidence in houses intervened with window, and door screening were lower than in houses without screening [40]. Ceiling modifications reduced indoor vector densities by around 80% in houses in Western Kenya [45]. Considering the cost of LLINs and IRS, investing in improving housing conditions for the general population may prove a cost-effective strategy in the middle to long term [38, 41, 46] and requires further investigation in the context of Bioko Island [37]. This argument, however, is challenged by recent observations of increased outdoor biting activity of mosquito vectors on Bioko, which is subject of current investigation by the BIMEP (unpublished work). This changing mosquito behaviour could partly explain the lack of significant associations between housing characteristics and malaria infection over time and in urban and rural dwellings.

Finally, with the exception of elevation, the investigation of environmental covariates could not reveal any strong effects on the odds of malaria infection on Bioko. This could be explained by the generally homogenous climatic conditions across the Island, where very hot and humid conditions prevail for much of the year. This echoes previous observations of small correlation coefficients of environmental predictors in Bayesian geo-statistical models predicting malaria prevalence [7]. The effect of elevation is not surprising but, unfortunately, this effect is only observable in a few communities on Bioko, as the highest altitude regions correspond to two uninhabited nature reserves in the center of the Island.

In a changing landscape of malaria transmission on Bioko Island, there is an increasing need for adaptively managing malaria control interventions. Between 2015 and 2019, IRS was targeted to areas where prevalence was relatively high, leaning on triennial MDC as the principal vector control strategy to guarantee high LLIN coverage to the whole island population. Malaria prevalence has proven a moving target, however, and the poor uptake of bed nets by the population adds to considerable challenges. Moreover, recent changes in vector ecology point to the need to complement existing interventions with novel strategies that also tackle outdoor and earlier biting [47, 48]. Notwithstanding, knowledge of some of the most important risk factors at the individual and household level could help redefine the optimal suite of malaria control interventions [49]. The effect of housing characteristics on the risk of malaria infection is worth considering for targeting certain household interventions using existing tools [36]. Crucially, the analyses presented here confirmed the very significant contribution of travel to the risk of being infected with malaria parasites, and this represents yet another call for interventions that target malaria importation and travellers. These could take the form of chemoprophylaxis or test-and-treat strategies for travellers at points of entry or, in perhaps a not-too-distant future, travel vaccines. Based on the results presented here and in previous work [27] it is plausible that this approach could substantially impact the local malaria burden.

Conclusion

Malaria risk on Bioko Island is heterogeneous and determined by a combination of factors interacting with local mosquito ecology. These interactions warrant further investigation to better adapt control according to need. The single most important risk factor identified was travel to mEG, in line with previous studies, and represents a significant challenge for the success of malaria control on the Island. Strategies to continue reducing the burden of transmission on the island begin with control activities on the mainland, a better understanding of the migratory population to tackle malaria among this group with targeted interventions, and a focus on importation upon return to the island. For Equatorial Guinea to achieve its goal of eliminating malaria at a national and sub-national level, there must be a focus on malaria prevention among the mobile population.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12936-023-04504-7.

Additional file 1: Table S1. Odds ratios estimated from regression models of all risk factors for malaria transmission on Bioko Island and 95% confidence intervals, stratified by rural and urban settings and malaria annual indicator survey year, 2015 and 2018.

Acknowledgements

This work was supported by the National Malaria Control Program and the Ministry of Health and Social Welfare of Equatorial Guinea, and MCD Global Health, previously Medical Care Development International (MCDI), through the Bioko Island Malaria Elimination Project (BIMEP). In particular, we thank the BIMEP staff for their efforts in collecting the data and implementing malaria control activities on Bioko Island. Finally, we would like to thank Marathon

Oil, Noble Energy, AMPCO (Atlantic Methanol Production Company), and the Ministry of Mines and Energy of Equatorial Guinea for their continued efforts and funding of malaria control on Bioko Island.

Author contributions
GAG and JB conceptualized the study. GAG, JMS, JNME, RMNA, WPP, EMA, CS, TCS, MRR planned, implemented, and supervised field activities. GAG managed and cleaned the data sets. GAG, MJ, and JB performed the analyses. GAG, MJ, CAG drafted the manuscript and prepared the figures and tables. DEBH, OTD, CAG, JC, IK, and JB critically revised the manuscript for intellectual content. All the authors read the initial draft and contributed with their feedback. All authors read and approved the final manuscript.

Funding
This work was funded by a private sector consortium led by Marathon Oil Corporation and the Government of Equatorial Guinea. J.B. received support from the UK MRC and the UK DFID (#MR/R010161/1) under the MRC/DFID Concordat agreement and as part of the EDCTP2 Program supported by the European Union. The funders had no role in study design, data collection, data analysis, data interpretation, the decision to publish, or preparation of the manuscript.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Ethics approval was provided by the ethics committee of the Ministry of Health and Social Welfare of Equatorial Guinea to implement the annual Malaria Indicator Survey.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]MCD Global Health, Silver Spring, MD, USA. [2]Duke Global Health Institute, Duke University, Durham, NC, USA. [3]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA. [4]MCD Global Health, Bioko Island, Malabo, Equatorial Guinea. [5]MCD Global Health, Hallowell, ME, USA. [6]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland. [7]University of Basel, Basel, Switzerland. [8]Equatorial Guinea Ministry of Health and Social Welfare, Bioko Island, Malabo, Equatorial Guinea. [9]Telethon Kids Institute, Perth Children's Hospital, Perth, Australia. [10]MRC International Statistics and Epidemiology Group, London School of Hygiene and Tropical Medicine, London, UK. [11]School of Pathology, Faculty of Health Science, Wits Institute for Malaria Research, University of Witwatersrand, Johannesburg, South Africa.

Received: 8 August 2022 Accepted: 18 February 2023
Published online: 01 March 2023

References

1. Fries B, Guerra CA, Garcia GA, Wu SL, Smith JM, Oyono JNM, et al. Measuring the accuracy of gridded human population density surfaces: a case study in Bioko Island Equatorial Guinea. PLoS ONE. 2021;16: e0248646.
2. Berzosa PJ, Cano J, Roche J, Rubio JM, Garcia L, Moyano E, et al. Malaria vectors in Bioko Island (Equatorial Guinea): PCR determination of the members of *Anopheles gambiae* Giles complex (Diptera: Culicidae) and pyrethroid knockdown resistance (kdr) in *An. gambiae sensu stricto*. J Vector Ecol. 2002;27:102–6.
3. Cano J, Berzosa PJ, Roche J, Rubio JM, Moyano E, Guerra-Neira A, et al. Malaria vectors in the Bioko Island (Equatorial Guinea): estimation of vector dynamics and transmission intensities. J Med Entomol. 2004;41:158–61.
4. Hay SI, Rogers DJ, Toomer JF, Snow RW. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, internet access and review. Trans R Soc Trop Med Hyg. 2000;94:113–27.
5. Kleinschmidt I, Sharp B, Benavente LE, Schwabe C, Torrez M, Kuklinski J, et al. Reduction in infection with *Plasmodium falciparum* one year after the introduction of malaria control interventions on Bioko island, Equatorial Guinea. Am J Trop Med Hyg. 2006;74:972–8.
6. Cook J, Bradley J, Kleinschmidt I, Hergott D, Garcia G, Schwabe C, et al. Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. Malar J. 2018;17:62.
7. Guerra CA, Kang SY, Citron DT, Hergott DEB, Perry M, Smith J, et al. Human mobility patterns and malaria importation on Bioko Island. Nat Commun. 2019;10:2332.
8. Overgaard HJ, Reddy VP, Abaga S, Matias A, Reddy MR, Kulkarni V, et al. Malaria transmission after five years of vector control on Bioko Island. Equatorial Guinea Parasit Vectors. 2012;5:253.
9. WHO. World Malaria Report 2021. Geneva: World Health Organization; 2021.
10. Bradley J, Matias A, Schwabe C, Vargas D, Monti F, Nseng G, et al. Increased risks of malaria due to limited residual life of insecticide and outdoor biting versus protection by combined use of nets and indoor residual spraying on Bioko Island. Equatorial Guinea Malar J. 2012;11:242.
11. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS Med. 2012;9: e1001165.
12. Magalhães RJ, Langa A, Sousa-Figueiredo J, Clements AC, Nery S. Finding malaria hot-spots in northern Angola: the role of individual, household and environmental factors within a meso-endemic area. Malar J. 2012;11:385.
13. Pinchoff J, Chaponda M, Shields T, Lupiya J, Kobayashi T, Mulenga M, et al. Predictive malaria risk and uncertainty mapping in Nchelenge District, Zambia: evidence of widespread, persistent risk and implications for targeted interventions. Am J Trop Med Hyg. 2015;93:1260–7.
14. Hagenlocher M, Castro MC. Mapping malaria risk and vulnerability in the United Republic of Tanzania: a spatial explicit model. Popul Health Metr. 2015;13:2.
15. Sturrock HJW, Cohen JM, Keil P, Tatem AJ, le Menach A, Ntshalintshali NE, et al. Fine-scale malaria risk mapping from routine aggregated case data. Malar J. 2014;13:421.
16. Cohen JM, Dlamini S, Novotny JM, Kandula D, Kunene S, Tatem AJ. Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. Malar J. 2013;12:61.
17. Tatem AJ, Huang Z, Narib C, Kumar U, Kandula D, Pindolia DK, et al. Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. Malar J. 2014;13:52.
18. Noor AM, Clements ACA, Gething PW, Moloney G, Borle M, Shewchuk T, et al. Spatial prediction of *Plasmodium falciparum* prevalence in Somalia. Malar J. 2008;7:159.
19. Raso G, Schur N, Utzinger J, Koudou BG, Tchicaya ES, Rohner F, et al. Mapping malaria risk among children in Côte dIvoire using Bayesian geostatistical models. Malar J. 2012;11:160.
20. García GA, Hergott DEB, Phiri WP, Perry M, Smith J, Osa Nfumu JO, et al. Mapping and enumerating houses and households to support malaria control interventions on Bioko Island. Malar J. 2019;18:283.
21. Bradley J, Monti F, Rehman AM, Schwabe C, Vargas D, Garcia G, et al. Infection importation: a key challenge to malaria elimination on Bioko Island. Equatorial Guinea Malar J. 2015;14:46.
22. NASA Earth Data. Land Processes Distributed Active Archive Center. https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table.
23. NASA Earth Data. MODIS (MOD 13)–Gridded Vegetation Indices (NDVI& EVI). NASA Earth Data. MODIS (MOD 13)–Gridded Vegetation Indices (NDVI& EVI). https://modis.gsfc.nasa.gov/data/dataprod/mod13.php.
24. Farr TG, Rosen PE, Caro E, Crippen R, Duren R, Hensley S, et al. The shuttle radar topography mission. Rev Geophys. 2007;45:RD2004.
25. Marshall JM, Bennett A, Kiware SS, Sturrock HJW. The hitchhiking parasite: why human movement matters to malaria transmission and what we can do about it. Trends Parasitol. 2016;32:752–5.
26. Stata. College Station, TX: Stata Corp.

27. Citron DT, Guerra CA, García GA, Wu SL, Battle KE, Gibson HS, et al. Quantifying malaria acquired during travel and its role in malaria elimination on Bioko Island. Malar J. 2021;20:359.
28. Patel JC, Taylor SM, Juliao PC, Parobek CM, Janko M, Gonzalez LD, et al. Genetic evidence of importation of drug-resistant *Plasmodium falciparum* to Guatemala from the Democratic Republic of the Congo. Emerg Infect Dis. 2014;20:932–40.
29. Guerra CA, Citron DT, García GA, Smith DL. Characterising malaria connectivity using malaria indicator survey data. Malar J. 2019;18:440.
30. Koenker H, Arnold F, Ba F, Cisse M, Diouf L, Eckert E, et al. Assessing whether universal coverage with insecticide-treated nets has been achieved: Is the right indicator being used? Malar J. 2018;17:355.
31. Huldén L, Mckitrick R, Huldén L. Average household size and the eradication of malaria. J R Stat Soc Ser A. 2014;177:725–42.
32. Kuate DB. Epidemiology and control of infant and early childhood malaria: a competing risks analysis. Int J Epidemiol. 1995;24:204–17.
33. Koram KA, Bennett S, Adiamah JH, Greenwood BM. Socio-economic risk factors for malaria in a peri-urban area of The Gambia. Trans R Soc Trop Med Hyg. 1995;89:146–50.
34. el Samani FZ, Willett WC, Ware JH. Nutritional and socio-demographic risk indicators of malaria in children under five: a cross-sectional study in a Sudanese rural community. J Trop Med Hyg. 1987;90:69–78.
35. Tusting LS, Ippolito MM, Willey BA, Kleinschmidt I, Dorsey G, Gosling RD, et al. The evidence for improving housing to reduce malaria: a systematic review and meta-analysis. Malar J. 2015;14:209.
36. García GA, Atkinson B, Donfack OT, Hilton ER, Smith JM, Eyono JNM, et al. Real-time, spatial decision support to optimize malaria vector control: The case of indoor residual spraying on Bioko Island Equatorial Guinea. PLoS Digital Health. 2022;1: e0000025.
37. Bradley J, Rehman AM, Schwabe C, Vargas D, Monti F, Ela C, et al. Reduced prevalence of malaria infection in children living in houses with window screening or closed eaves on Bioko Island Equatorial Guinea. PLoS ONE. 2013;8: e80626.
38. Jatta E, Jawara M, Bradley J, Jeffries D, Kandeh B, Knudsen JB, et al. How house design affects malaria mosquito density, temperature, and relative humidity: an experimental study in rural Gambia. Lancet Planet Health. 2018;2:e498-508.
39. Dlamini N, Hsiang MS, Ntshalintshali N, Pindolia D, Allen R, Nhlabathi N, et al. Low-quality housing is associated with increased risk of malaria infection: a national population-based study from the low transmission setting of Swaziland. Open Forum Infect Dis. 2017;4:ofx071.
40. Getawen SK, Ashine T, Massebo F, Woldeyes D, Lindtjørn B. Exploring the impact of house screening intervention on entomological indices and incidence of malaria in Arba Minch town, southwest Ethiopia: a randomized control trial. Acta Trop. 2018;181:84–94.
41. Gunawardena DM, Wickremasinghe AR, Muthuwatta L, Weerasingha S, Rajakaruna J, Senanayaka T, et al. Malaria risk factors in an endemic region of Sri Lanka, and the impact and cost implications of risk factor-based interventions. Am J Trop Med Hyg. 1998;58:533–42.
42. Ghebreyesus TA, Haile M, Witten KH, Getachew A, Yohannes M, Lindsay SW, et al. Household risk factors for malaria among children in the Ethiopian highlands. Trans R Soc Trop Med Hyg. 2000;94:17–21.
43. Kirby MJ, Ameh D, Bottomley C, Green C, Jawara M, Milligan PJ, et al. Effect of two different house screening interventions on exposure to malaria vectors and on anaemia in children in The Gambia: a randomised controlled trial. Lancet. 2009;374:998–1009.
44. Lindsay SW, Jawara M, Paine K, Pinder M, Walraven GEL, Emerson PM. Changes in house design reduce exposure to malaria mosquitoes. Trop Med Int Health. 2003;8:512–7.
45. Atieli H, Menya D, Githeko A, Scott T. House design modifications reduce indoor resting malaria vector densities in rice irrigation scheme area in western Kenya. Malar J. 2009;8:108.
46. Gimnig JE, Slutsker L. House screening for malaria control. Lancet. 2009;374:954–5.
47. Sougoufara S, Ottih EC, Tripet F. The need for new vector control approaches targeting outdoor biting Anopheline malaria vector communities. Parasit Vectors. 2020;13:295.
48. Gatton ML, Chitnis N, Churcher T, Donnelly MJ, Ghani AC, Godfray HCJ, et al. The importance of mosquito behavioural adaptations to malaria control in Africa. Evolution (N Y). 2013;67:1218–30.
49. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. Bull World Health Organ. 2000;78:1401–11.

Publisher's Note

3.2 Whole genome sequences of *Plasmodium falciparum* on Bioko Island, Equatorial Guinea illustrate connectedness with geographical neighbors and island-wide partial population structure.

Manuscript in preparation. Planned submission to mBio.

Whole genome sequences of *Plasmodium falciparum* on Bioko Island, Equatorial Guinea illustrate connectedness with geographical neighbors and island-wide partial population structure

Thomas C. Stabler*[1, 2], Ankit Dwivedi[3], Bing Guo[4], Biraj Shrestha[4], Sudhaunshu Joshi[4], Matilde Riloha Rivas[5], Olivier Tresor Donfack[6], Carlos A. Guerra[7], Guillermo A. García[7], Claudia Daubenberger[1, 2], Joana C. Silva*[3, 8, 9]

[1]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland

[2]University of Basel, Basel, Switzerland

[3]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA

[4]Malaria Research Program, Center for Vaccine Development and Global Health, University of Maryland Baltimore, Baltimore, MD, USA

[5]Equatorial Guinea Ministry of Health and Social Welfare, Malabo, Equatorial Guinea

[6]Medical Care Development Global Health, Malabo, Equatorial Guinea

[7]Medical Care Development Global Health, Silver Spring, Maryland, USA

[8]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

[9]Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade NOVA de Lisboa (NOVA), Lisbon, Portugal

*Corresponding authors

Email: thomas.stabler@swisstph.ch ; jcsilva@som.umaryland.edu

Keywords: Malaria, Plasmodium falciparum, Population Genetics, Identity-by-Descent, Bioko Island

Abstract:

Introduction:

Whole genome sequences (WGS) of *Plasmodium falciparum* on Bioko Island, Equatorial Guinea, were analyzed to assess the genetic diversity and population structure of malaria parasites. Islands offer unique opportunities for malaria elimination due to their natural geographical barriers limiting parasite mixing, theoretically. However, previous epidemiological studies have shown Bioko Island is vulnerable to imported strains from nearby neighbors in continental Africa through human migration. The impact of these imported strains has yet to be characterized at the genome level.

Methods:

This study aimed to determine the connectedness of *P. falciparum* populations on Bioko Island with neighboring mainland regions and to explore the island-wide population structure. Dried blood spot samples (DBS) from the 2019 MIS were used to generate whole genome sequencing (WGS) data. Genomic analyses were conducted using population genetics metrics to measure local genetic diversity and structure, as well as the connectedness to other sub-Saharan African *P. falciparum* populations using public available samples.

Results:

The analysis revealed significant genetic diversity and partial population structure, indicating some degree of isolation among parasite populations between urban and rural communities on Bioko Island. Results showed a greater connectedness between the Bioko Island and Gabon parasite populations than Bioko Island to Cameroonian populations. Not only supporting previous evidence of significant migration to the island, but suggests importation may originate from Gabon or nearby continental regions, like Rio Muni, the continental regions of Equatorial Guinea. These findings highlight the complex dynamics of malaria transmission on the island.

Background

In the effort to eliminate malaria, islands hold a distinct geographical advantage over other malaria-endemic areas with land borders. The presence of surrounding water should, theoretically, provide a barrier from parasite mixing between local and non-local populations, and, with the introduction of effective control/prevention methods, facilitate rapid progress to malaria elimination. This has been observed for Cabo Verde, where no indigenous cases have been reported since 2017, and was declared malaria free in 2024 (1, 2). Unfortunately, for other island contexts, like Zanzibar and São Tomé and Principe, malaria continues to persist despite significant decreases in malaria (3, 4). Migration has been attributed as a significant obstacle to achieve sustainable progress toward malaria elimination (5, 6). Critically, assessing whether elimination is possible on an island without addressing migration could provide valuable evidence for malaria control programs to distribute their resources.

Bioko Island, Equatorial Guinea, is located 32 km west of the coast of Cameroon and has historically high levels of malaria transmission (7) (**Figure 1A**). Stringent and intensive malaria control interventions have been conducted on the island since 2004 by the Bioko Island Malaria Elimination Project (BIMEP) (formerly the Bioko Island Malaria Control Project – BIMCP), which reduced malaria prevalence from 43.3% to 10.5% by 2016 (7). Despite these interventions, which also resulted in significant reductions in malaria mortality on the island (7, 8), malaria prevalence has fluctuated around 10-12% since 2016 (8, 9). Epidemiological, questionnaire-based studies by BIMEP demonstrate strong associations between on-island infections and travelers recently returning from mainland Equatorial Guinea, suggestive of case importation and represent a source of parasites immigrating to the island (8, 10-12). However, the actual contribution of parasite migration to the Bioko Island remains to be characterized.

The mainland region of the country, Rio Muni, has higher malaria prevalence and has not conducted significant malaria control campaigns since 2011 due to funding constraints (13). Cameroon and Gabon, neighbors flanking Rio Muni, are hyper-endemic regions with year-round transmission and high parasite genetic diversity (1, 14, 15). If migration from the continent to Bioko Island is high, the *P. falciparum* population on the island is expected to have a similarly high genetic diversity and little differentiation from Cameroon and Gabon. However, if the ocean forms a significant geographical barrier, and importation does not significantly contribute to Bioko's malaria transmission and genetic diversity, then the *P. falciparum* population might differentiate from parasite genomes observed on the mainland.

Bioko Island has an approximate population of 270,000, of which 90% reside in Malabo city, the capital of Equatorial Guinea, and is composed of two provinces (Bioko Norte and Bioko Sur) and four districts (Malabo, Baney, Luba and Riaba) (8, 16, 17) (**Figure 1B**). Historically, the greatest reductions in malaria burden have been achieved in rural communities, especially

in the Baney and Luba districts (7). However, rural residents also account for most of the on-island travel, Malabo being the primary destination for employment and education (11). Considering the disparity in human distribution and reported travel, more than one parasite subpopulation may exist. Conversely, importation and inter-island connectedness may be sufficiently high as to prevent differentiation and the formation of subpopulations.

The use of whole genome sequence data to estimate genetic diversity and relatedness among malaria isolates can provide a highly nuanced assessment of the impact of control interventions over time and illuminate parasite transmission routes (18, 19). Consequently, WHO encourages malaria control programs to incorporate parasite genome sequencing technologies as a molecular surveillance technique (20). Previous studies have investigated demography of *P. falciparum* populations using whole genome data in sub-Saharan African (21); however, the vastness of this region has resulted in geographic "pockets" with incomplete molecular characterization of *P. falciparum*. Especially in areas with active malaria control campaigns but limited history of molecular surveillance, generating *P. falciparum* genomic descriptions and estimating genetic relationships can provide insights into the response of a parasite population to ongoing interventions and provide valuable data to inform future approaches to parasite control (18, 22-24).

Here, we aim to characterize the population structure of *P. falciparum* on Bioko Island and determine its relationship with parasite populations in neighboring regions on the mainland. A subset of dried blood spot (DBS) samples obtained from participants in the BIMEP's 2019 malaria indicator survey (MIS) were used to generate whole-genome sequence (WGS) data. Population genetics principles and metrics were applied to measure differentiation between Bioko Island and continental *P. falciparum* parasite populations, and understand local transmission dynamics by describing island-specific markers and relatedness between *P. falciparum* isolates.

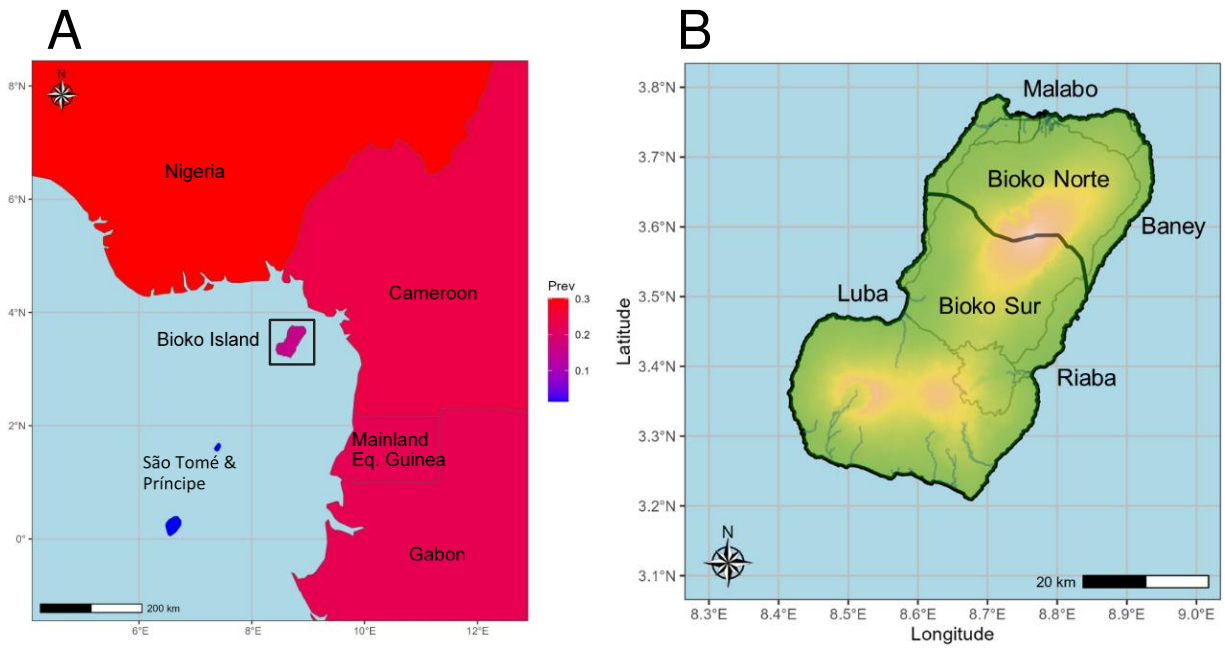


Figure 1. **Map of Bioko Island, Equatorial Guinea**. A. Geographical location of Bioko Island within the Gulf of Guinea. Color gradient denotes the estimated country-wide *P. falciparum* prevalence in 2019 reported in the 2023 World Malaria Report. Bioko Island color was amended to show island-wide prevalence as measured from the 2019 MIS. B. Detail of Bioko Island with province borders (black lines), roads (grey lines), rivers (blue lines) and elevation (green/red gradient – red denotes higher elevation).

METHODS

Sample collection

DBS samples were collected on Whatman filter papers (GE healthcare Ltd, Forest farm, Cardiff, UK; Product code: 11962089) from August to October 2019. On each filter paper are blood spots of 0.5 inches in diameter (n=4), each spot representing a volume of approximately 50 µl of blood. Survey and laboratory data used in this study include: gender, rapid diagnostic test (RDT) diagnosis (*P. falciparum* or mixed infection), qPCR diagnosis (*P. falciparum* or mixed infection), travel history (previous two months), location (urban or rural, or Malabo, Baney, Luba, Riaba districts), age group in years (<5, 5-15, 15-18, 18+), and parasite density by qPCR (Low: Cq $\geq$ 20: High: Cq < 20). Household sampling was based on primary sampling units (PSUs) constructed from 1x1 km map-areas that make up the Bioko mapping grid (25). PSUs were assigned to either a rural or an urban stratum based on household density and local residual transmission, with 25% and 5% of households sampled from each, respectively. All MIS participants provided informed consent, or legal guardian consent for underage participants (<18 years). Filter papers were selected for DNA extraction from 202 individuals with a positive RDT and reported fever.

Additionally, publicly available WGS data were downloaded from the Sequence Read Archive (SRA) provided by the MalariaGEN *Plasmodium falciparum* Community Project (26).

Samples were selected from West, Central and East African countries as representative of their respective continental regions (**Supplementary File S2**).

DNA extraction

A DNA extraction method based on guanidine and silica purification protocols, and developed by the Malaria Research Program, at the University of Maryland Baltimore (UMB), was applied to selected DBS filter paper samples as described (27). Briefly, one circle/DBS was cutout, incubated, and submerged in lysis buffer for 2 hours at 65°C. Samples then underwent two washes before extracting DNA with TE buffer. Extracted DNA material was stored at -80°C until use.

Polymerase chain reaction (PCR)

The Qiagen QuantiTect® Multiplex PCR was used to conduct qPCR (Qiagen Sciences, Germantown, Maryland, USA). The master mix was prepared according to manufacturer's instructions, but adapted to exclude ROX or UDG. RNAseq free water was used in all PCR reactions. The following PCR program was used: [1] 20 minutes at 50°C; [2] 15 minutes at 95°C; [3] 45 seconds at 94°C; [4] 75 seconds at 60°C. Steps 3 and 4 were repeated 45 times. For each sample, all PCR reactions were performed in duplicate. Samples were considered positive if the mean quantification cycles (Cq) of duplicate qPCR reactions was Cq < 40. If one sample result was reported as RDT positive and the corresponding PCR result was negative, the assay was repeated to make a definitive conclusion, if necessary.

Selective Whole Genome Amplification (sWGA)

Selective whole genome amplification (sWGA) was applied on extracted samples with optimal parasite density by Cq, as previously described (28, 29). In short, selected samples underwent a vacuum filtration step and were then amplified in 0.2 mL 96-well PCR plate with the following reaction mixture: 1X BSA, 1mM dNTPs, 2.5 µM of amplification primers, 1X Phi29 reaction buffer and 20 units of Phi29 polymerase. Primers are the same used in Oyola SO, et al. that preferentially bind at adequate distance and amplify *P. falciparum* genomes (29). From each extraction, 17 µL of the DNA sample was added to the reaction mixture for a total of 50 µL final volume. Amplification occurred in a thermocycler with the following stepdown protocol: 35°C for 20 minutes, 34°C for 10 minutes, 33°C for 15 minutes, 32°C for 20 minutes, 31°C for 30 minutes, 30°C for 16 hours. The stepdown protocol was followed with a heating step at 65°C for 20 minutes and cooled to 4°C.

Whole genome sequencing

Preparation of genomic DNA libraries are previously described (30). To summarize, genomic DNA libraries were generated from amplified samples using the KAPA Library Preparation Kit (Kapa Biosystems, Woburn, MA), and then fragmented to approximately 200 base pair (bp) lengths. A modified version of the manufacturer's protocol was used. AMPure XT beads were utilized to inform the library size selection. DNA concentration and fragment size was conducted using the DNA High Sensitivity Assay on the LabChip GX (Perkin Elmer, Waltham, MA) tool. All sample libraries were uniquely barcoded, pooled and sequenced on 150 bp paired-end Illumina NovaSeq 6000 run (Illumina, San Diego, CA).

Read mapping and identification of single nucleotide polymorphisms (SNPs)

Raw fastq files were mapped to the reference *P. falciparum* genome, 3D7, using bowtie2 (31). BAM file processing followed GATK Best Practices (32, 33). Coverage and depth estimates from reads were generated using bedtools (34). Variant calling of each sample was conducted utilizing the Haplotype Caller toolkit to generate genomic variant call format files (GVCF) and perform joint SNP (single nucleotide polymorphisms) calling. When appropriate, diploid calls were allowed since polyclonal infections were expected; otherwise major alleles were called (70% threshold) to genotype the most prevalent strain in the infection (polymorphic sites that did not reach the threshold were set to missing). Variant calls were filtered to omit potential false positive results with the following criteria: DP<5, FS>14.5, MQ<20.0, QUAL<50. Additional filtering was performed to exclude rare allele events (frequency less than 0.05%), sites with >10% missingness, samples with >20% missing genotype values (35).

Genomic Analysis

Genetic diversity in each geographic region was estimated by nucleotide diversity (π) (Vcftools v.0.1.16), the average pairwise difference between samples. Within-host diversity was measured by the F-statistic F_{WS} (36) using the R package moimix (v0.0.2.9001, https://bahlolab.github.io/moimix/), where diversity of *P. falciparum* in each sample was compared against the diversity of the entire sample set. Infections were considered polyclonal if $F_{WS} < 0.95$. P-values were calculated using Chi-squared test to determine statistical significance for differences in proportions of polyclonal infections, as determined by F_{WS}, between African countries.

Principal components analysis (PCA) was applied to a dataset of biallelic SNPs passing filtering criteria, to investigate the extent to which geographic origin contributes to differences between isolates and the genome-wide level. Clustering by PCA was estimated using the R package SNPRelate (v.1.28.0; https://github.com/zhengxwen/SNPRelate) (37, 38). Admixture analysis (ADMIXTURE v1.2) was used to obtain an estimate of contributions

of inferred ancestral subpopulations to each sample. PCA and admixture datasets were pruned for sites in linkage disequilibrium (LD) (window size of 5 kbp, $r^2 \geq 0.2$) prior to analysis. Among samples with a clonal pair (i.e. nearly identical genomes), one sample was selected to represent the clonal group, and the other(s) excluded from the analysis. ANOVA was used to measure differences in composition in inferred ancestral subpopulation, between countries.

The F-statistic F_{ST} was applied to measure overall differentiation (mean F_{ST}) between sampled parasite populations from different geographic regions, and to identify SNPs contributing to differences between populations (39). Significance of F_{ST} per genomic site was estimated empirically, with 5,000 permutations of samples by geographic region, using custom scripts. Overall relatedness between strains as measured by identity-by-descent (IBD) was estimated using hmmIBD (version 2.0.0; https://github.com/glipsnort/hmmIBD) (40). Prior to conducting IBD, samples were deconvoluted using dEploid v0.6-beta (41). Within each sample the strain accounting for the highest proportion of the sample was selected for downstream analysis. Network analysis and community detection was conducted using the Infomap software (42). To estimate effective population size (Ne), IBDNe was applied to the dataset (43). Previously published Python-based scripts were used to process files and run hmmIBD, infomap and IBDNe analyses, excluding small IBD segments under 2cM and highly conserved IBD peaks to minimize bias (44). Samples were considered related (siblings or clonal) if >25% of genomes were IBD (45). The R package sf (46) were used to generate maps of Bioko. R (v4.1.3) was used with all appropriate R-based packages, such as ggplot2, to generate figures. Cluster plots were generated using Cytoscape v3.10.2 (https://cytoscape.org) (47).

RESULTS

WGS data descriptives

Samples with higher parasite density infections were prioritized for sWGA (mean Cq = 32.9). Whole genome shotgun sequencing (WGS) data was generated for 90 samples collected on Bioko Island. Of the 90 samples for which WGS data was generated, fifty-two samples were collected in urban communities compared to 38 from Bioko's rural areas (**Figure 2**). Per sample, an average of 28,276,493 total reads were generated and on average across samples, reads mapped 75.4% of the reference 3D7 genome with 5X read coverage (Supplemental Figure S1). Of the 90 samples, 16 samples were excluded due to inadequate quantity of mapped reads (< 8 million) or missingness (>20% SNPs missing).

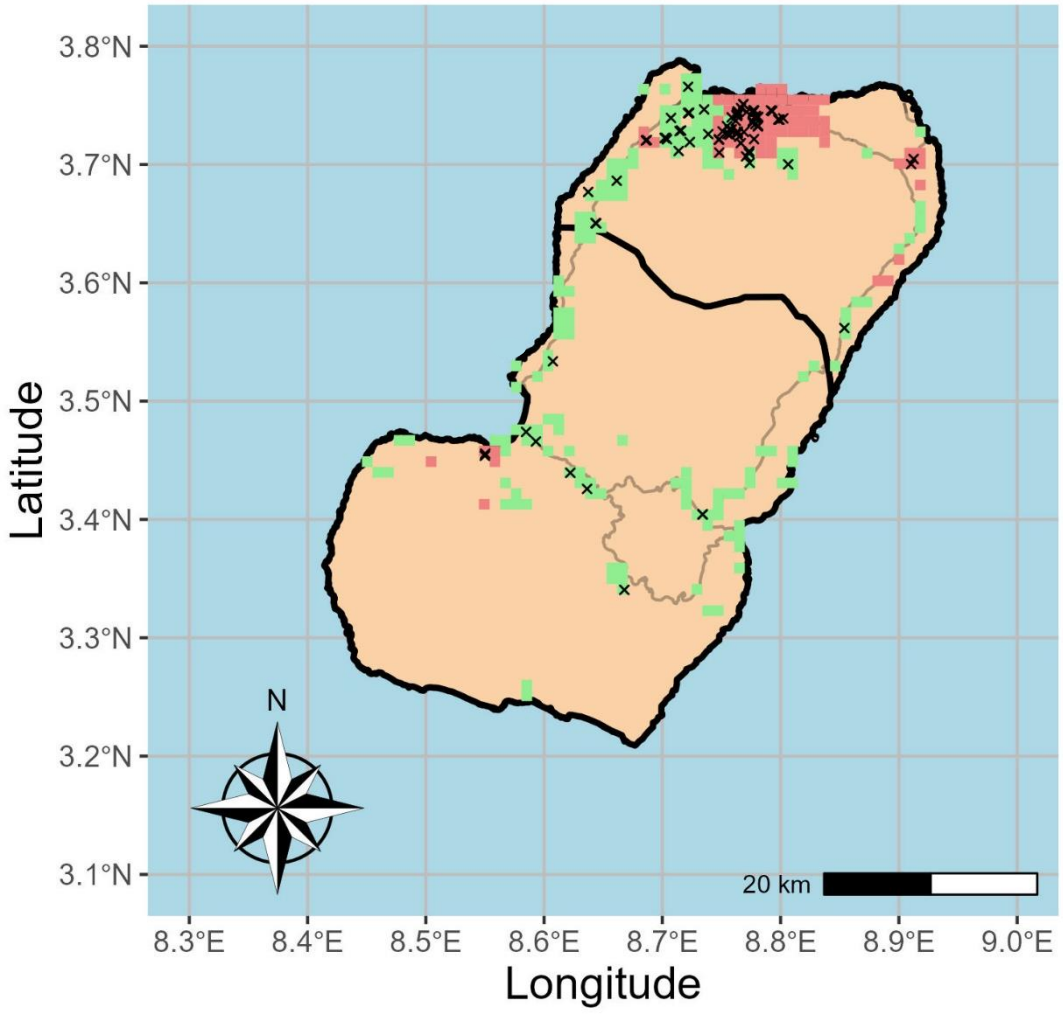


Figure 2. **Urban and Rural Communities on Bioko Island**. Map of Bioko Island with communities color marked as urban (red) or rural (green) assignments. The urban/rural status was determined by a combination of population density and estimated local residual transmission (11). Black crosses represent geographical location of collection of all sequenced samples (n=90).

Bioko Island parasites were combined with 992 sub-Saharan African WGS datasets downloaded from the MalariaGEN *P. falciparum* Community Project (26). Public samples were selected from countries representative of West (Guinea, Mali, Burkina Faso, and Nigeria), Central (Cameroon, Gabon and Democratic Republic of Congo) and East (Kenya, Tanzania, Malawi, and Mozambique) Africa. Variable sites in the highly polymorphic sub-telomeric genomic regions were excluded due to reduced accuracy (48, 49). After calling major alleles, a total of 1,076,963 biallelic SNPs were included, of which 375,921 (34.9%) were intergenic, 205,240 (19.1%) were synonymous and 426,551 (39.6%) were non-synonymous. Overall, average read coverage across all sites was 63.0X (59.9X for Bioko Island samples).

Bioko Island parasite population appears partially structure and polyclonal

To describe on-island diversity and transmission intensity, nucleotide diversity and within-host diversity (F_{WS}) were measured among Bioko samples (n=74). Per-site nucleotide diversity appeared similar between urban/rural communities, but when stratifying by district, higher diversity across all SNPs was observed among Malabo city and Malabo suburb samples over Baney or Luba, 0.058, 0.056, 0.050 and 0.052, respectively. There was some evidence of recent migration between a few strains, measured by IBD (Supp. Figure S2), however overall relatedness suggests significant mixing between parasites ($IBD_{Bioko} = 0.003$). Relatedness is higher in urban communities ($IBD_{rural} = 0.004$; $IBD_{urban} = 0.002$; $p = 0.01$) suggestive of greater parasite mixing. The lack of agreement between urban and rural relatedness may reflect a partial population structure on Bioko, since the entire Bioko *P. falciparum* population is not panmitic. Within-host diversity measured by F_{WS} revealed the majority (46 out of 74) of samples were polyclonal (**Figure 3**), where most polyclonal infections (n=41) originated in Malabo communities. However transmission intensity appears similar as there was no proportional difference of polyclonal infections between Malabo, Baney and Luba communities (X^2 p-value = 0.31). Population subgroups associated with complex infections complemented previous epidemiological observations among subgroups with higher transmission intensity (e.g. travelers), however there was no observable statistical difference (Supp. Table S1) (8, 11, 17). In fact, complexity of infections does not significantly differ from those in Cameroon or Gabon, the closest African countries in the African mainland (X^2 p-value = 1; X^2 p-value = 0.56, respectively). In general, relatedness between Bioko samples supports the presence of a partial population structure, however, transmission intensity, as measured by polyclonal infections, does not appear to differentiate. Island-wide parasites appear connected, however this connection is less pronounced between urban and rural communities.

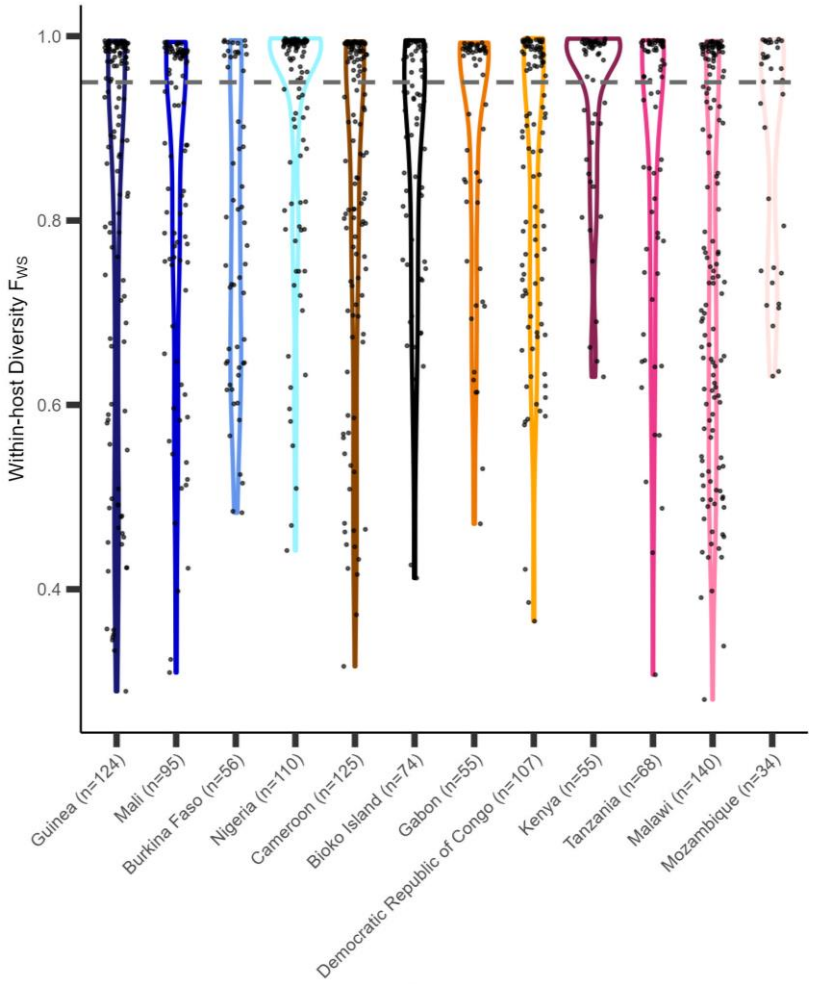


Figure 3. **Within-host diversity as measured by *F*ws**. Y-axis represents F_{WS} values of samples between 0 and 1. Isolates with F_{WS} value below 0.95 are inferred to be polyclonal (grey dotted line). Color groups denote the African region (West, Central, and East) as assigned by the WHO, and individual colors represent a country were samples were collected (x-axis).

Differentiation between Malabo, Baney and Luba parasites occurs on loci associated with cellular adhesion, invasion and sequestration

The fixation index, F_{ST}, was applied to measure genetic differentiation among Bioko subgroups. As above, less differentiation was observed between urban/rural communities ($F_{ST\ urban\ vs\ rural}$ = 0.01), however, when stratifying by district, differentiation increased where within-Malabo parasites had less differentiation ($F_{ST\ Malabo\ city\ vs\ Malabo\ suburbs}$ = 0.007) compared to Malabo and non-Malabo districts ($F_{ST\ Malabo\ city\ vs\ Baney}$ = 0.04; $F_{ST\ Malabo\ city\ vs\ Luba}$ = 0.04), suggestive of some variation between on-island parasite populations. To look at specific genomic signals driving differentiation, F_{ST} per site was calculated between Malabo, Baney and Luba. Among a total of 163,291 SNPs, Baney and Luba consistently differentiated from Malabo isolates at gene loci associated with red blood cell invasion (*clag3.2* - PF3D7_0302200; *msp* - PF3D7_1035600), hepatocyte invasion (DBL-containing protein -

PF3D7_0113800), and parasite sequestration (*var2csa* - PF3D7_1200600) (**Figure 4**). PF3D7_0302200 encodes cytoadherence-linked asexual protein 3.2 (*clag3.2*), associated with RBC invasion and nutrient acquisition post-invasion (50, 51). PF3D7_1035600 encodes for a merozoite protein (*msp* or H101), although whose exact function is unknown, is linked to merozoite invasion (50). PF3D7_0113800 encodes for a DBL domain protein, associated with Rho GTPases and involved in hepatocyte/erythrocyte invasion, and may be a target of acquired immunity (52-56). PF3D7_1200600 encodes for *var2csa*, a protein associated with placental adhesion and placental malaria, where infected red blood cells (RBCs) interact with placental cells and leukocytes (57).

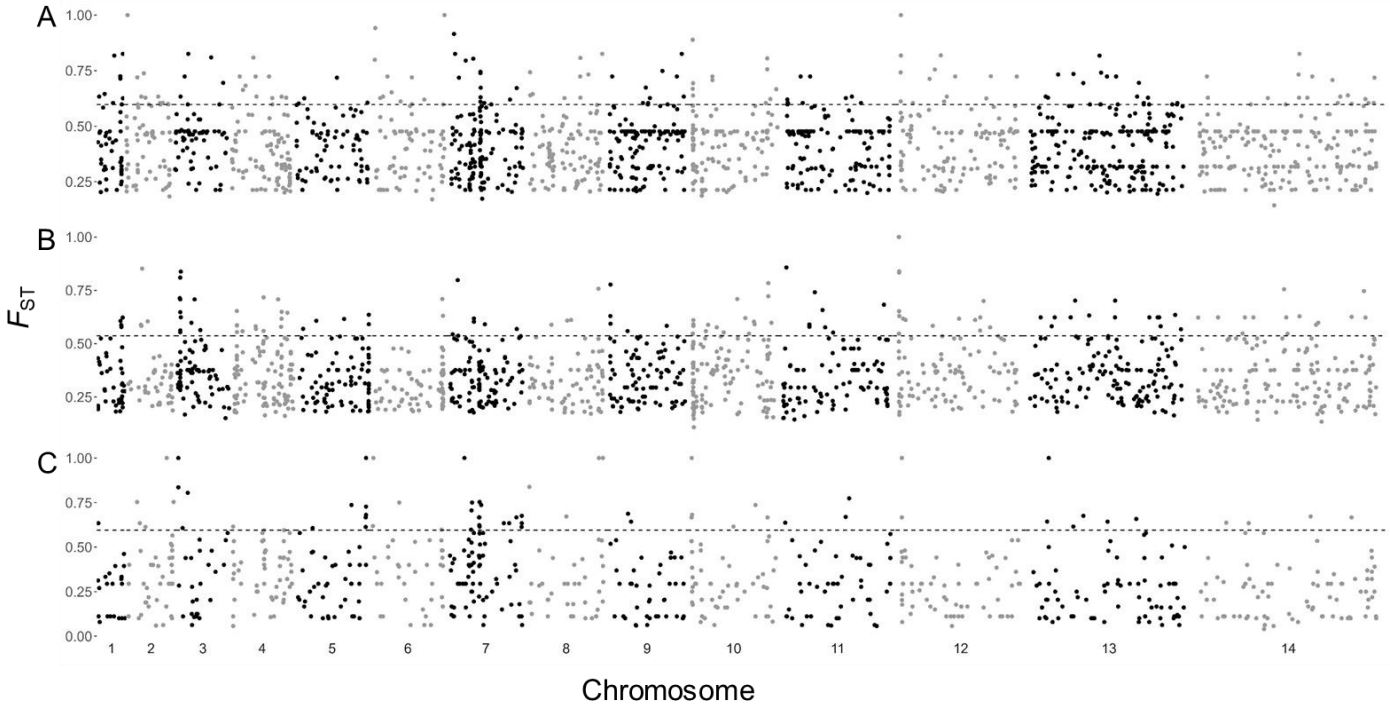


Figure 4. **Genome-wide SNP diversity by F_{ST}.** Allele differentiation was measured between three Bioko subpopulations based on location: A. Malabo vs Baney (Total SNPs = 1,699); B. Malabo vs Luba (Total SNPs = 1,420); C. Baney vs Luba (Total SNPs = 790). A total of 163,291 SNPs were included in each sample set with heterozygous sites included and minor allele excluded. Dotted line in each panel represents the 10% of highest F_{ST} values per sample set. 5,000 permutations were run using vcftools v.0.1.16 and custom UMB pipeline. Rv4.1.3 with package ggplot2 was used to generate results.

Bioko Island P. falciparum population does not appear uniquely dissimilar from continental neighbors

Despite Bioko's clear geographical barrier, results illustrate the lack of differentiation between *P. falciparum* populations with the continent. Bioko had high nucleotide diversity

compared to reported prevalence at time of collection (All SNPs = $8.7x10^{-3}$; Nonsynonymous SNPs = $8.8x10^{-3}$; Synonymous SNPs = $8.5x10^{-3}$), suggestive of several contributing factors, including importation when observed against other African groups (Supp. Fig. S3). Although intense malaria control interventions are applied annually, incoming cases from regions with higher prevalence and more diverse parasites may be contributing to high diversity on the island. This effect can also be observed after conducting PCA, where Bioko parasites do not significantly differentiate from its closest geographical neighbors (Fig 2). For the PCA analysis, of the initial 1,066 samples, 988 samples were included. Samples were excluded based on missingness (n=25), outlier genotype data (n=39), and highly related, or clonal, clusters, where only one sample was included to represent the clonal group (n=14). Variance within the sample set could be explained as geographical distance between country groups (West/Central versus East = 0.48% variance; West versus Central = 0.24% variance). Values appear small considering variance is measured across 38,206 SNPs. The lack of a unique Bioko population complements previous evidence of continued importation and mixing on Bioko with mainland Africa *P. falciparum* populations (10, 11).

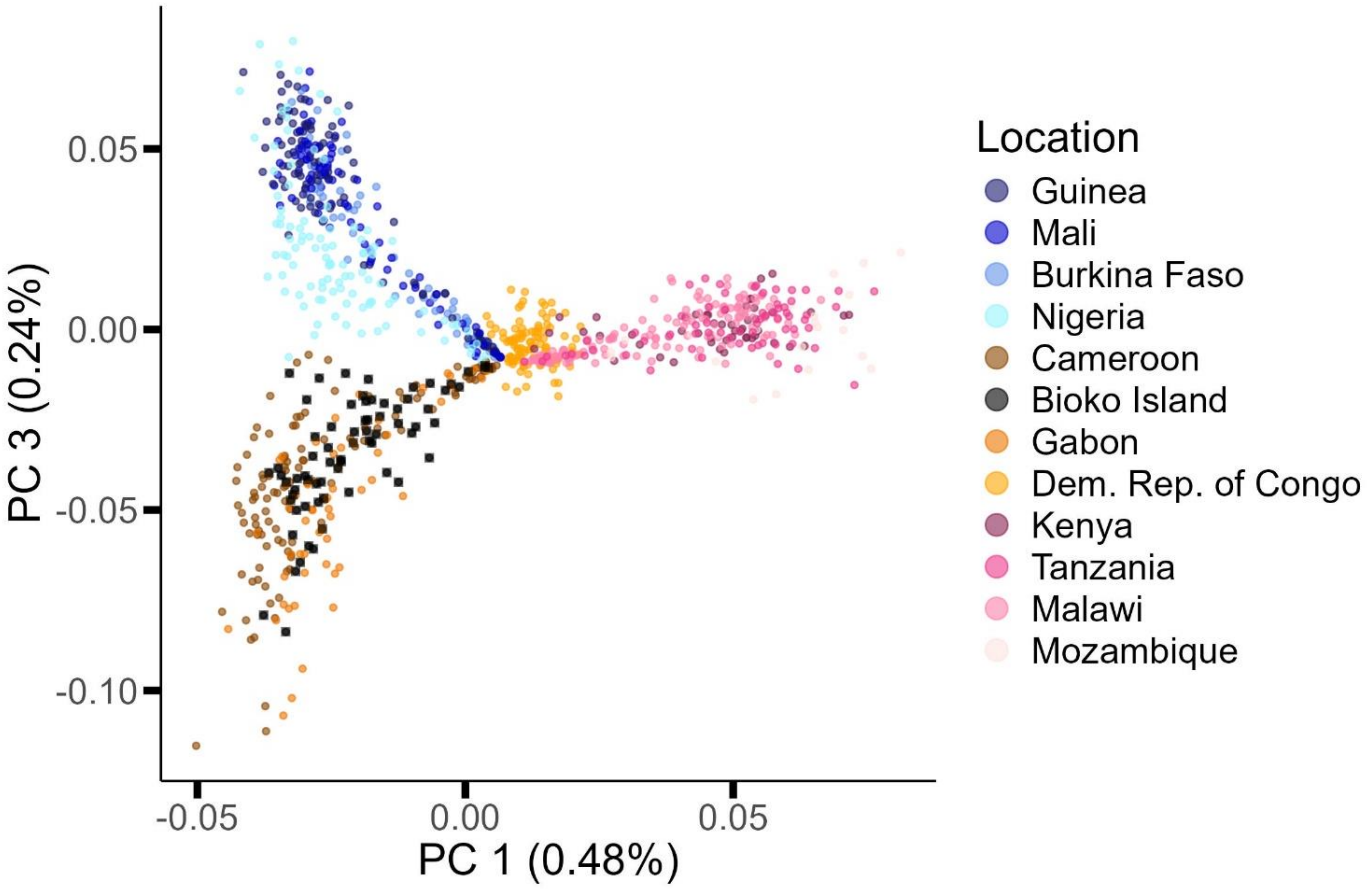


Figure 5. **Principal Components Analysis (PCA) of Bioko and African genomes**. A sub-Saharan African sample set was merged with Bioko Island, Equatorial Guinea samples. WGS data were downloaded from MalariaGEN. Color groups denote the African region (West, Central, and East) as assigned by WHO, and colors represent country were samples were collected. Prior to PCA, major alleles were called for all samples, excluding heterozygous calls that did not meet applied threshold (70%). PC2 is not plotted since it is based on subsets of missingness with no particular pattern. Figure was generated in R v4.1.3 with package SNPRelate v1.28.0.

After LD-pruning and minimizing CV error, five populations are inferred among samples (n SNPs = 9,427; n samples = 988) (**Figure 3**). The largest ancestral contributor to Bioko Island parasites is predominant in Central-West Africa, followed by subpopulations more common in West and Southeast African ancestry, likely due to historical, migratory connections between these regions. Interestingly, a subpopulation (see grey population in **Figure 3**) predominant in East Africa and the DRC contributing to Gabon ancestry appears much less pronounced among the Bioko population ($ANOVA_{Bioko\text{-}GAB}$ $p = 0.008$). Further, the population predominant in Southeast Africa appears to contribute significantly more to Bioko ancestry than its nearby neighbor of Cameroon ($ANOVA_{Bioko\text{-}CAM}$ $p < 0.001$). This may hint at an ancestral connection between Bioko and Southeast Africa that does not exist between Bioko and East African populations. Ultimately, Bioko's population structure appears admixed, where very few samples' ancestry comes from a single ancestral source.

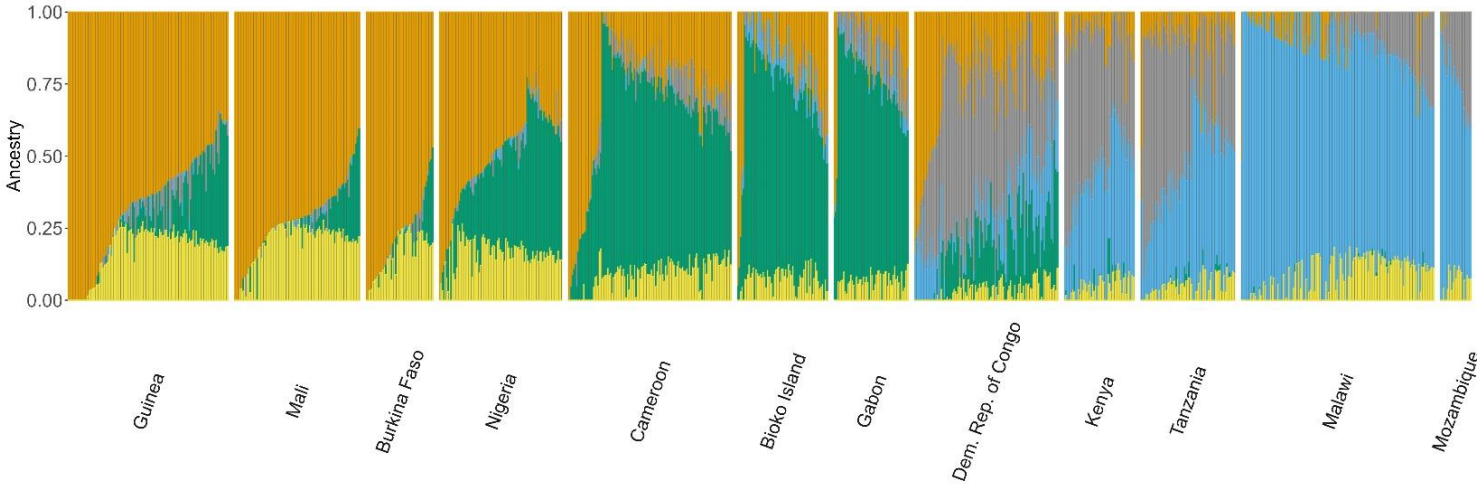


Figure 6. **Admixture of Bioko and sub-Saharan parasites.** Colors denote each ancestral subpopulation identified for K=5, when cross validation error was lowest. Samples are represented as columns, Y-axis refers to the proportion of ancestry attributable to ancestral subpopulation, in each sample. Orange represents the subpopulation predominant in West Africa, green for Central Africa, grey for Democratic Republic of Congo and East Africa and blue for Southeast Africa. Yellow represents a subpopulation not predominant in any major African region. Results were generated using the ADMIXTURE software and plots generated using Rv4.1.3.

Greater connectedness between Bioko and Gabonese parasite populations

To assess the impact of importation on the Bioko *P. falciparum* population, Cameroon and Gabon samples were used as a representative sample set for the suspected source of most imported strains to Bioko. All samples were deconvoluted, and the predominant strain within each sample used. After merging and filtering, a total of 344,703 SNPs were used. After removal of highly conserved IBD peaks, Bioko samples had lower overall relatedness to each other than Cameroon and Gabon (IBD_{Bioko} = 0.003; $IBD_{Cameroon}$ = 0.005; IBD_{Gabon} = 0.006) (**Figure 7A**). Relatedness between Bioko and Gabon was similar to the Bioko population alone ($IBD_{Bioko\text{-}Bioko}$ = 0.003; $IBD_{Bioko\text{-}Gabon}$ = 0.002; ANOVA p = 0.66), whereas this was not the same between Bioko and Cameroon ($IBD_{Bioko\text{-}Bioko}$ = 0.003; $IBD_{Bioko\text{-}Cameroon}$ = 0.001; ANOVA $p < 0.001$). Further, community detection by infomap identified 33 Bioko samples that clustered with Gabon samples (Gabon-linked Bioko samples), as opposed to 9 clustering with Cameroon (Cameroon-linked Bioko samples), and 38 Bioko samples clustering with neither Gabon nor Cameroon (Suspected local strains). Network analysis illustrates suspected local Bioko strains have closer connection to Gabon-linked Bioko strains versus Cameroon-linked strains (**Figure 7B**). When excluding Gabon-linked Bioko samples, estimated effective population size (N_e) on Bioko was greatly reduced (Bioko N_e = 14,700; Bioko + $Bioko_{Cameroon}$ N_e = 24,400; Bioko + $Bioko_{Gabon}$ N_e = 58,100), reflecting the magnitude of genetic variation contributed from Bioko samples linked to continental neighbors, especially Gabon. Ultimately, despite Cameroon's closer geographic location to Bioko, the Gabonese *P. falciparum* population appears to have a greater influence and impact on Bioko diversity. This suggests imported strains to Bioko more commonly come from Gabon, or nearby regions to Gabon, such as Rio Muni.

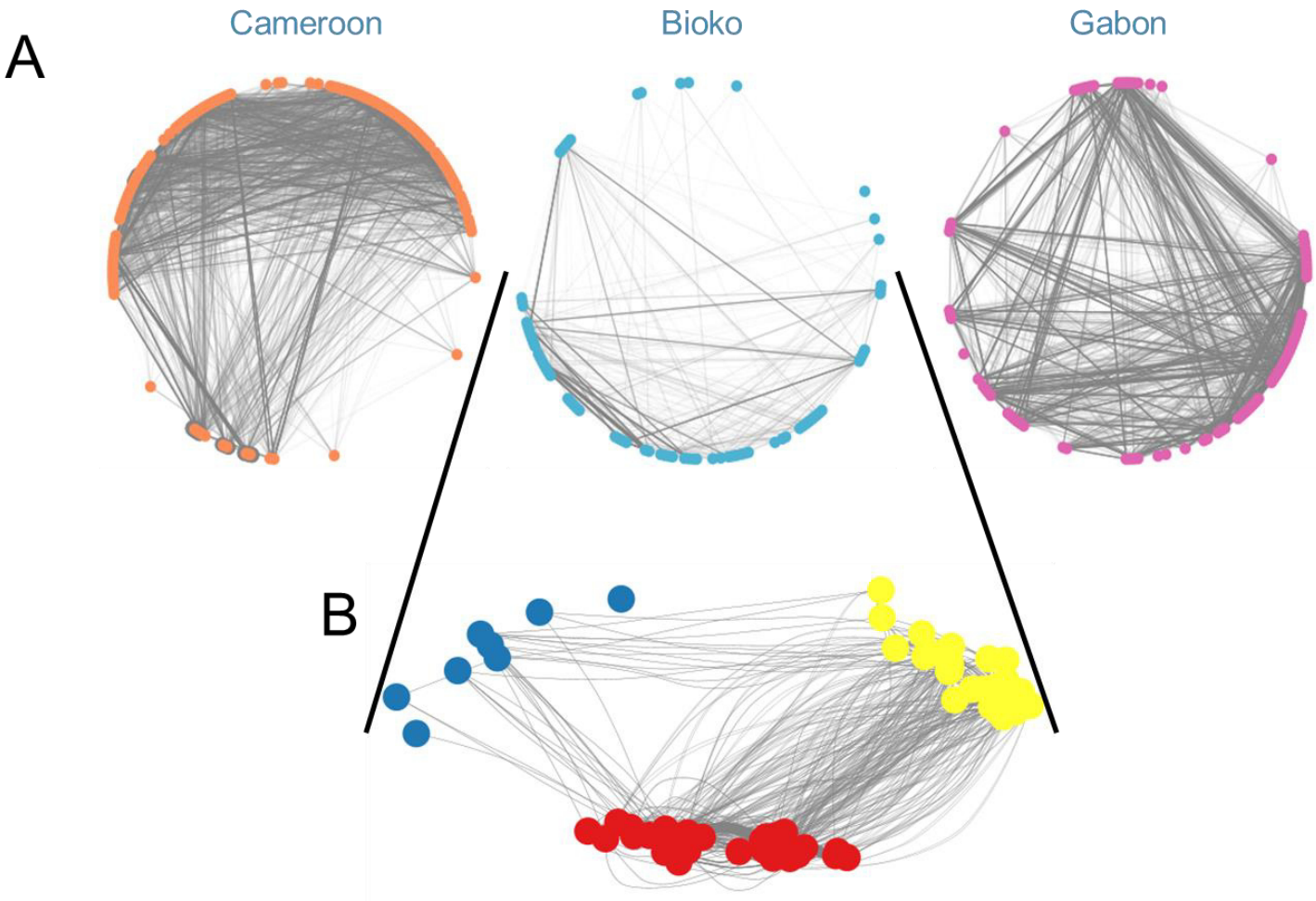


Figure 7. **Network analyses of P. falciparum between Bioko, Cameroon and Gabon**. Relatedness between suspected local, Cameroon-linked and Gabon-linked Bioko sample set was measured by IBD. A. Within group IBD-sharing. Each dot represents a WGS sample and clustered according to communities assigned by Infomap. Line thickness represents the relatedness between sample pairs (thicker line = higher relatedness). B. Bioko samples separated into subgroups according to links with Cameroon or Gabon clusters as assigned by Infomap. Red = Bioko samples linked to neither Cameroon nor Gabon clusters (i.e. suspected local strains), blue = Cameroon-linked Bioko samples, and yellow = Gabon-linked Bioko samples. Figures were generated using Cytoscape v3.10.2.

Discussion

Malaria on Bioko island offers an important view into the effect of control interventions within Central-West Africa considering its midpoint location between Nigeria and the Democratic Republic of Congo, the two biggest contributors to malaria cases in 2022, 27% and 12% respectively (1). As sequencing becomes more assessable, national malaria control programs are incorporating sequencing methodologies as an essential molecular tool to provide improved accuracy to supplement traditional epidemiological approaches and inform control and elimination strategies (20, 58-60). As part of the BIMEP's effort to eliminate malaria from the island, WGS data was generated from the 2019 MIS field samples to describe the parasite population at the genomic level. The Bioko *P. falciparum* population appears partially structured, where some differentiation exists between parasites in the Malabo district *versus* the more rural districts of Baney and Luba. There were some small indications of connectedness between urban and rural communities, however relatedness measurements differed, which may reflect a small isolation effect occurring in the island's rural periphery. Other island contexts observed a similar heterogeneous effect (61, 62), suggesting the impact of imported strains do not influence rural island parasites as quickly as urban ones. This is, in part, evidence of imported infections mixing with the local Bioko parasite population, but is not mixing with Bioko parasites homogeneously. As importation has likely been occurring over many years, Bioko parasites did not significantly differentiate from its geographical neighbors, and transmission intensity appeared higher than expected of a region with moderate transmission rates (9, 63). Although BIMEP has significantly decreased malaria burden on the island, the *P. falciparum* diversity and structure on Bioko reflect a malaria endemic region with moderate transmission that does not form its own unique population from continental neighboring regions.

Differential SNPs between Baney and Luba was much lower than their respective comparison to Malabo, suggesting less differentiation between parasites despite being on opposites sides of the island. As the ethnic Bubi tribe make up the majority of residents in these districts, this may be a reflection of complemental cultural proximity and accessibility in addition to geographical (64). Conserved regions and differential SNPs among rural strains were membrane-bound genes associated with adhesion, sequestration and invasion. Typically, these types of genes are under greater host immune selection (65-67), and some, like PF3D7_0113800, have been identified as potential vaccine targets (unpublished). However further evidence is needed to support this observation or whether these signals are due to differences in selection pressure. Local transmission dynamics suggest that as BIMEP reaches very low transmission rates, rural, non-Malabo communities may be open to elimination without significant concern for reseeding from Malabo.

P. falciparum on Bioko appears admixed with high genetic diversity and connection to the African continent. Most infections were polyclonal, a reflection of sustained transmission intensity and complex infections (68), similar to other low-moderate transmission regions in Africa (61, 69, 70). Beyond human migration, other factors likely contribute to Bioko's transmission rate and high genetic diversity like high mosquito biting rates and breeding sites (71, 72). The presence of an ancestral subpopulation among Bioko samples predominant in East Africa may hint at a transmission route between Bioko, Gabon and DRC that is not as pronounced in Cameroon. This is further supported by the IBD analysis, showing higher relatedness between Bioko and Gabon samples, a relatedness that is not as pronounced between Bioko and Cameroon, despite being geographically closer. This connection suggests imported strains to Bioko more commonly come from Gabon or nearby regions, such as Rio Muni, the continental region of Equatorial Guinea. Although speculative, there may be closer cultural and ethnic ties between Rio Muni and Gabon, as the predominant demographic in both countries is the Fang tribe (73-75). Results show the magnitude of imported infections is significant as more than half of Bioko parasites clustered with non-Bioko subpopulations. For there to be sustained, prolonged progress towards malaria elimination, the BIMEP should consider targeted screening and control interventions for travelers to the island, and if possible, reestablish malaria control to Rio Muni.

Within this study, there were some limitations. No DBS samples have been collected from the continental region of Equatorial Guinea, a likely source of importation to Bioko, which could have better distinguished between local and imported strains. Unfortunately, this is unavoidable as no MIS has occurred in the Rio Muni since 2011 to our knowledge. Also, Bioko samples were only sequenced from infections with positive RDT and fever to optimize for WGS, but may over-represent complex infections. This selection criteria was used to ensure quality data from field infections with lower initial nucleic material. Sample size was also limited due to resource constraints and may affect results, but is considered negligible since whole genome data provide thousands of data points for analysis. Ultimately, this study follows accepted methodologies (61, 65, 69, 76), and accounts for the limitations within its analysis and interpretations.

Using sWGA, high-quality WGS *P. falciparum* data was generated from DBS collected during the 2019 MIS on Bioko Island, Equatorial Guinea to describe local transmission dynamics and measure connectedness to mainland Africa. Locally, partial population structure appears among rural infections outside of the Malabo district. However, there is significant connectedness between the island and nearby mainland neighbors. Bioko-Gabon parasites appear to be most related, where the Fang tribe are the predominant human demographic and thus are more connected through culture and language.

Considerable reductions in island-wide malaria transmission are needed, and importation to the island needs to be addressed to ensure continued progress. Additional sequencing on Bioko and the Equatorial Guinea mainland would further illuminate local transmission dynamics and temporal changes, ideally from both symptomatic and asymptomatic infections. Identifying local *versus* imported strains would help understand Bioko transmission dynamics on a more granular level. The current WGS dataset provides an informative snapshot of *P. falciparum* dynamics, successfully sequencing parasites from field samples, and illustrating the heterogeneous transmission occurring on the island and the effect imported strains may be having on local malaria elimination efforts.

List of abbreviations

BIMEP – Bioko Island Malaria Elimination Project
COI – complexity of infection
DBS – dried blood spot
IBD – identity-by-descent
LD – linkage disequilibrium
MIS – Malaria Indicator Survey
PCA – principal component analysis
PSU – primary sampling units
RBC – red blood cell
SNP – single nucleotide polymorphism
SRA – sequence read archive
sWGA – selective whole genome amplification
WGS – whole genome sequencing

Declarations

Ethics approval and consent to participate

Ethics approval for the 2019 MIS was provided by the Equatorial Guinea Ministry of Health and Social Welfare and the ethics committee of the London School of Hygiene and Tropical Medicine (approval number 5556). Written informed consent was sought from each participating adult and on behalf of participating children under 18 years of age.

Consent for publication

Not applicable

Availability of data and materials

WGS data will be made available on GenBank.

Competing Interests

The authors declare that they have no competing interests.

Funding

Work related to the BIMEP activities was funded by a private sector consortium led by Marathon Oil Corporation and the Government of Equatorial Guinea. The funders had no role in study design, data collection, data analysis, data interpretation, the decision to publish, or preparation of the manuscript. Work related to the generation of WGS data was funded by the National Institutes of Health (NIH) awards R01 AI141900 to JCS.

Author's contributions

TCS conceived the study, performed all DNA extractions, qPCR assays and sWGA, conducted analyses, interpreted results and wrote the paper. AD supported the study design and methodology, contributed to the interpretation of results and edited the manuscript. BS and SJ advised and supported all laboratory procedures and edited the paper. MRR sponsored and oversaw the 2019 MIS. OTD managed the 2019 MIS teams collecting all field samples, organized sample logistics and edited the paper. CAG contributed to the study design, supported interpretation of 2019 MIS epidemiological data, assisted in generating

Bioko Island figures, and edited the paper. GAG contributed fundamental funding and resources to the study, contributed to the study design, and edited the paper. CD conceived of the study, contributed essential resources to the study and wrote the paper. JCS conceived of the study, contributed to the study design, interpretation of results, contributed funding and resources and wrote the paper. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was supported by the National Malaria Control Programme and the Ministry of Health and Social Welfare of Equatorial Guinea, and MCD Global Health through the Bioko Island Malaria Elimination Project (BIMEP). Specifically, we would like to thank the entire BIMEP team for their efforts conducting the 2019 MIS. Ultimately, we would like to thank Marathon Oil, Noble Energy, AMPCO (Atlantic Methanol Production Company), and the Ministry of Mines and Energy of Equatorial Guinea for their continued efforts and support for malaria control on Bioko Island. This work was also supported by the National Institutes of Health through grant R01AI141900. We thank the staff of the IGS's Maryland Genomics for genomic DNA library construction and sequencing.

References

1. WHO. World malaria report 2023: World Health Organization; 2023.
2. WHO. WHO certifies Cabo Verde as malaria-free, marking a historic milestone in the fight against malaria. 2024 [Available from: https://www.who.int/news/item/12-01-2024-who-certifies-cabo-verde-as-malaria-free--marking-a-historic-milestone-in-the-fight-against-malaria.
3. Ali MH, Kitau J, Ali AS, Al-Mafazy AW, Tegegne SG, Ussi O, et al. Malaria elimination in Zanzibar: where next? Pan Afr Med J. 2023;45(Suppl 1):7.
4. Wang Y, Li M, Guo W, Deng C, Zou G, Song J. Burden of Malaria in Sao Tome and Principe, 1990–2019: Findings from the Global Burden of Disease Study 2019. International Journal of Environmental Research and Public Health. 2022;19(22):14817.
5. Markwalter CF, Menya D, Wesolowski A, Esimit D, Lokoel G, Kipkoech J, et al. Plasmodium falciparum importation does not sustain malaria transmission in a semi-arid region of Kenya. PLOS Global Public Health. 2022;2(8):e0000807.
6. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. Quantifying the impact of human mobility on malaria. Science. 2012;338(6104):267-70.
7. Cook J, Hergott D, Phiri W, Rivas MR, Bradley J, Segura L, et al. Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. Malaria Journal. 2018;17(1):62.
8. García GA, Janko M, Hergott DEB, Donfack OT, Smith JM, Mba Eyono JN, et al. Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions. Malaria Journal. 2023;22(1):72.
9. Nchama VUNN, Said AH, Mtoro A, Bidjimi GO, Owono MA, Maye ERM, et al. Incidence of Plasmodium falciparum malaria infection in 6-month to 45-year-olds on selected areas of Bioko Island, Equatorial Guinea. Malaria Journal. 2021;20(1):322.
10. Bradley J, Monti F, Rehman AM, Schwabe C, Vargas D, Garcia G, et al. Infection importation: a key challenge to malaria elimination on Bioko Island, Equatorial Guinea. Malaria Journal. 2015;14(1):46.
11. Guerra CA, Kang SY, Citron DT, Hergott DEB, Perry M, Smith J, et al. Human mobility patterns and malaria importation on Bioko Island. Nature Communications. 2019;10(1):2332.
12. Citron DT, Guerra CA, García GA, Wu SL, Battle KE, Gibson HS, Smith DL. Quantifying malaria acquired during travel and its role in malaria elimination on Bioko Island. Malaria Journal. 2021;20(1):359.
13. Rehman AM, Mann AG, Schwabe C, Reddy MR, Roncon Gomes I, Slotman MA, et al. Five years of malaria control in the continental region, Equatorial Guinea. Malaria Journal. 2013;12(1):154.
14. Atuh NI, Anong DN, Jerome F-C, Oriero E, Mohammed NI, D'Alessandro U, Amambua-Ngwa A. High genetic complexity but low relatedness in Plasmodium falciparum infections from Western Savannah Highlands and coastal equatorial Lowlands of Cameroon. Pathogens and Global Health. 2022;116(7):428-37.
15. Mvé-Ondo B, Nkoghe D, Arnathau C, Rougeron V, Bisvigou U, Mouele LY, et al. Genetic diversity of Plasmodium falciparum isolates from Baka Pygmies and their Bantu neighbours in the north of Gabon. Malaria Journal. 2015;14(1):395.
16. Fries B, Guerra CA, García GA, Wu SL, Smith JM, Oyono JNM, et al. Measuring the accuracy of gridded human population density surfaces: A case study in Bioko Island, Equatorial Guinea. PLOS ONE. 2021;16(9):e0248646.
17. Guerra CA, Citron DT, García GA, Smith DL. Characterising malaria connectivity using malaria indicator survey data. Malaria Journal. 2019;18(1):440.
18. Neafsey DE, Taylor AR, MacInnis BL. Advances and opportunities in malaria population genomics. Nature Reviews Genetics. 2021;22(8):502-17.
19. Holzschuh A, Lerch A, Gerlovina I, Fakih BS, Al-mafazy A-wH, Reaves EJ, et al. Multiplexed ddPCR-amplicon sequencing reveals isolated Plasmodium falciparum populations amenable to local elimination in Zanzibar, Tanzania. Nature Communications. 2023;14(1):3699.

20. WHO. Global technical strategy for malaria 2016–2030, 2021 update. J Geneva: World Health Organization. 2021:1-40.
21. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, et al. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. 2019;365(6455):813-6.
22. Oberstaller J, Zoungrana L, Bannerman CD, Jahangiri S, Dwivedi A, Silva JC, et al. Integration of population and functional genomics to understand mechanisms of artemisinin resistance in Plasmodium falciparum. International Journal for Parasitology: Drugs and Drug Resistance. 2021;16:119-28.
23. Rocamora F, Winzeler EA. Genomic Approaches to Drug Resistance in Malaria. Annual Review of Microbiology. 2020;74(Volume 74, 2020):761-86.
24. Mensah BA, Akyea-Bobi NE, Ghansah A. Genomic approaches for monitoring transmission dynamics of malaria: A case for malaria molecular surveillance in Sub–Saharan Africa. Frontiers in Epidemiology. 2022;2.
25. García GA, Hergott DEB, Phiri WP, Perry M, Smith J, Osa Nfumu JO, et al. Mapping and enumerating houses and households to support malaria control interventions on Bioko Island. Malaria Journal. 2019;18(1):283.
26. MalariaGEN, Abdel Hamid M, Abdelraheem M, Acheampong D, Ahouidi A, Ali M, et al. Pf7: an open dataset of Plasmodium falciparum genome variation in 20,000 worldwide samples [version 1; peer review: 3 approved]. Wellcome Open Research. 2023;8(22).
27. Zainabadi K, Adams M, Han ZY, Lwin HW, Han KT, Ouattara A, et al. A novel method for extracting nucleic acids from dried blood spots for ultrasensitive detection of low-density Plasmodium falciparum and Plasmodium vivax infections. Malaria Journal. 2017;16(1):377.
28. Shah Z, Adams M, Moser KA, Shrestha B, Stucke EM, Laufer MK, et al. Optimization of parasite DNA enrichment approaches to generate whole genome sequencing data for Plasmodium falciparum from low parasitaemia samples. Malaria Journal. 2020;19:1-10.
29. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, et al. Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. Malaria Journal. 2016;15(1):597.
30. Moser KA, Drábek EF, Dwivedi A, Stucke EM, Crabtree J, Dara A, et al. Strains used in whole organism Plasmodium falciparum vaccine trials differ in genome structure, sequence, and immunogenic potential. Genome Medicine. 2020;12(1):6.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. J Nature Methods. 2012;9(4):357.
32. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. J Nature genetics. 2011;43(5):491.
33. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinform. 2013;43(1):11.0.1-.0.33.
34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
35. Karamoko Niaré BG, Jeffrey A Bailey. An Optimized GATK4 Pipeline for Plasmodium falciparum Whole Genome Sequencing Variant Calling and Analysis. Malaria Journal. 2023;PREPRINT.
36. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. 2012;487(7407):375-9.
37. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326-8.
38. Zheng X, Gogarten SM, Lawrence M, Stilp A, Conomos MP, Weir BS, et al. SeqArray—a storage-efficient high-performance data format for WGS variant calls. Bioinformatics. 2017;33(15):2251-7.

39. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38(6):1358-70.
40. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. Malaria Journal. 2018;17(1):196.
41. Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infections in Plasmodium falciparum from high throughput sequencing data. Bioinformatics. 2017;34(1):9-15.
42. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences. 2008;105(4):1118-23.
43. Browning Sharon R, Browning Brian L. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. The American Journal of Human Genetics. 2015;97(3):404-18.
44. Guo B, Borda V, Laboulaye R, Spring MD, Wojnarski M, Vesely BA, et al. Strong positive selection biases identity-by-descent-based inferences of recent demography and population structure in Plasmodium falciparum. Nature Communications. 2024;15(1):2499.
45. Shetty AC, Jacob CG, Huang F, Li Y, Agrawal S, Saunders DL, et al. Genomic structure and diversity of Plasmodium falciparum in Southeast Asia reveal recent parasite migration patterns. Nature Communications. 2019;10(1):2665.
46. Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal. 2018;10(1):439-46.
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003;13(11):2498-504.
48. Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, et al. Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. Genome Res. 2016;26(9):1288-99.
49. Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, et al. Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. Wellcome Open Res. 2018;3:52.
50. Proellocks NI, Coppel RL, Mohandas N, Cooke BM. Malaria Parasite Proteins and Their Role in Alteration of the Structure and Function of Red Blood Cells. In: Rollinson D, Stothard JR, editors. Advances in Parasitology. 91: Academic Press; 2016. p. 1-86.
51. Tripathi J, Zhu L, Nayak S, Stoklasa M, Bozdech Z. Stochastic expression of invasion genes in Plasmodium falciparum schizonts. Nature Communications. 2022;13(1):3004.
52. Aghazadeh B, Zhu K, Kubiseski TJ, Liu GA, Pawson T, Zheng Y, Rosen MK. Structure and mutagenesis of the Dbl homology domain. Nat Struct Biol. 1998;5(12):1098-107.
53. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. Nature. 2002;420(6916):629-35.
54. Schepis A, Kumar S, Kappe SH. Malaria parasites harness Rho GTPase signaling and host cell membrane ruffling for productive invasion of hepatocytes. Cell Reports. 2023;42(1).
55. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-Wide Analysis of Selection on the Malaria Parasite Plasmodium falciparum in West African Populations of Differing Infection Endemicity. Molecular Biology and Evolution. 2014;31(6):1490-9.
56. Mbye H, Mane K, Diop MF, Demba MA, Bojang F, Mohammed NI, et al. Plasmodium falciparum merozoite invasion ligands, linked antimalarial resistance loci and ex vivo responses to antimalarials in The Gambia. Journal of Antimicrobial Chemotherapy. 2022;77(11):2946-55.
57. Tomlinson A, Semblat J-P, Gamain B, Chêne A. VAR2CSA-Mediated Host Defense Evasion of Plasmodium falciparum Infected Erythrocytes in Placental Malaria. Frontiers in Immunology. 2021;11.
58. Dalmat R, Naughton B, Kwan-Gett TS, Slyker J, Stuckey EM. Use cases for genetic epidemiology in malaria elimination. Malaria Journal. 2019;18(1):163.

59. Golumbeanu M, Edi CAV, Hetzel MW, Koepfli C, Nsanzabana C. Bridging the Gap from Molecular Surveillance to Programmatic Decisions for Malaria Control and Elimination. The American Journal of Tropical Medicine and Hygiene. 2023:tpmd220749.
60. Tessema SK, Raman J, Duffy CW, Ishengoma DS, Amambua-Ngwa A, Greenhouse B. Applying next-generation sequencing to track falciparum malaria in sub-Saharan Africa. Malaria Journal. 2019;18(1):268.
61. Moss S, Mańko E, Vasileva H, Da Silva ET, Goncalves A, Osborne A, et al. Population dynamics and drug resistance mutations in Plasmodium falciparum on the Bijagós Archipelago, Guinea-Bissau. Scientific Reports. 2023;13(1):6311.
62. Morgan AP, Brazeau NF, Ngasala B, Mhamilawa LE, Denton M, Msellem M, et al. Falciparum malaria from coastal Tanzania and Zanzibar remains highly connected despite effective control efforts on the archipelago. Malaria Journal. 2020;19(1):47.
63. Simpson SV, Nundu SS, Arima H, Kaneko O, Mita T, Culleton R, Yamamoto T. The diversity of Plasmodium falciparum isolates from asymptomatic and symptomatic school-age children in Kinshasa Province, Democratic Republic of Congo. Malaria Journal. 2023;22(1):102.
64. Taylor AR, Echeverry DF, Anderson TJC, Neafsey DE, Buckee CO. Identity-by-descent with uncertainty characterises connectivity of Plasmodium falciparum populations on the Colombian-Pacific coast. PLOS Genetics. 2020;16(11):e1009101.
65. Kassegne K, Komi Koukoura K, Shen H-M, Chen S-B, Fu H-T, Chen Y-Q, et al. Genome-Wide Analysis of the Malaria Parasite Plasmodium falciparum Isolates From Togo Reveals Selective Signals in Immune Selection-Related Antigen Genes. Frontiers in Immunology. 2020;11.
66. Bockhorst J, Lu F, Janes JH, Keebler J, Gamain B, Awadalla P, et al. Structural polymorphism and diversifying selection on the pregnancy malaria vaccine candidate VAR2CSA. Molecular and Biochemical Parasitology. 2007;155(2):103-12.
67. Gupta A, Thiruvengadam G, Desai SA. The conserved clag multigene family of malaria parasites: Essential roles in host–pathogen interaction. Drug Resistance Updates. 2015;18:47-54.
68. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. Molecular Ecology. 2013;22(2):273-85.
69. Roh ME, Tessema SK, Murphy M, Nhlabathi N, Mkhonta N, Vilakati S, et al. High Genetic Diversity of Plasmodium falciparum in the Low-Transmission Setting of the Kingdom of Eswatini. J Infect Dis. 2019;220(8):1346-54.
70. Amambua-Ngwa A, Jeffries D, Mwesigwa J, Seedy-Jawara A, Okebe J, Achan J, et al. Long-distance transmission patterns modelled from SNP barcodes of Plasmodium falciparum infections in The Gambia. Scientific Reports. 2019;9(1):13515.
71. Guerra CA, Fuseini G, Donfack OT, Smith JM, Ondo Mifumu TA, Akadiri G, et al. Malaria outbreak in Riaba district, Bioko Island: lessons learned. Malaria Journal. 2020;19(1):277.
72. Reddy MR, Overgaard HJ, Abaga S, Reddy VP, Caccone A, Kiszewski AE, Slotman MA. Outdoor host seeking behaviour of Anopheles gambiae mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malaria Journal. 2011;10(1):184.
73. Stokes J. Encyclopedia of the Peoples of Africa and the Middle East: Infobase Publishing; 2009.
74. CIA. Equatorial Guinea: People and Society 2024 [Available from: https://www.cia.gov/the-world-factbook/countries/equatorial-guinea/.
75. CIA. Cameroon: People and Society 2024 [Available from: https://www.cia.gov/the-world-factbook/countries/cameroon/.
76. Vanheer LN, Mahamar A, Manko E, Niambele SM, Sanogo K, Youssouf A, et al. Genome-wide genetic variation and molecular surveillance of drug resistance in Plasmodium falciparum isolates from asymptomatic individuals in Ouélessébougou, Mali. Scientific Reports. 2023;13(1):9522.

3.3 Integrating local malaria molecular monitoring into regular malaria indicator surveys on Bioko Island: high association between urban communities and low density infections.

Manuscript in preparation. Planned submission to Malaria Journal.

Integrating local malaria molecular monitoring into regular malaria indicator surveys on Bioko Island: high association between urban communities and low density infections.

Thomas C. Stabler[1,2], Salome Hosch[1,2], Elizabeth Nyakarungu[3], Johanna Nouria Giger[1,2], Monserrat Kobe Elonga[4], Rosalia Nzang Bibang[4], Victoria Mangue Ndong[4], Ilda Elobe Mometolo[4], Alejandro Lopelo Bolopa[4], Rosa Maria Oko Bedoung[4], Jose Raso Bijeri[4], Esperanza Willy Roka[4], Ana Gertrudis Nguema[4], Victoria Palacios Ichinda[4], Salvador Bakale[4], Cristina Nka Esokolete[4], Matilde Riloha Rivas[5], David S. Galick[3], Wonder Philip Phiri[3], Carlos A. Guerra[6], Guillermo A. García[6], Luca Montemartini[7], Tobias Schindler[7], Joana C. Silva[8,9,10], Claudia Daubenberger[1,2]

1. Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland
2. University of Basel, Basel, Switzerland
3. MCD Global Health, Malabo, Equatorial Guinea
4. Baney Reference Laboratory, Baney, Equatorial Guinea
5. Equatorial Guinea Ministry of Health and Social Welfare, Malabo, Equatorial Guinea
6. MCD Global Health, Silver Spring, Maryland, USA
7. Diaxxo AG, Zurich, Switzerland
8. Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA
9. Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA
10. Global Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical, IHMT, Universidade NOVA de Lisboa, UNL, Lisboa, Portugal

*Corresponding authors

Email: thomas.stabler@swisstph.ch ; claudia.daubenberger@swisstph.ch

Keywords: Malaria, Plasmodium falciparum, PlasmoPod, low density infections, Bioko Island

Abstract:

Background:

Effective malaria control requires accurate identification of *Plasmodium* infections to tailor interventions appropriately. Rapid diagnostic tests (RDTs) are crucial tools for this purpose due to their small size and ease-of-use functionality. These tests typically target the *Plasmodium falciparum* histidine-rich protein 2 (HRP2) antigen. However, some strains of *P. falciparum* have deletions in the *hrp2* and *hrp3* genes, rendering HRP2-based RDTs ineffective. Additionally, RDTs have a detection limit of less than 100 parasites per microliter, insufficient for identifying low-density infections that sustain malaria transmission. This study explores integrating molecular monitoring using a novel qPCR device, PlasmoPod, using samples from a malaria indicator surveys (MIS) on Bioko Island, Equatorial Guinea to enhance detection of low-density infections and inform targeted malaria control strategies.

Methods:

The study utilized a combination of RDTs and the PlasmoPod device for molecular monitoring. The PlasmoPod employs qPCR targeting 18S rDNA/rRNA, capable of detecting parasite densities as low as 0.2 parasites per microliter, significantly more sensitive than HRP2-based RDTs. Samples from the 2023 MIS were extracted from dried blood spots (DBS), qPCR run in duplicate on the PlasmoPod. Data analysis involved linking qPCR data with MIS datasets using unique volunteer ID codes and employing statistical methods to assess the association between various risk factors to malaria infection by qPCR, and risk factors to low density infections.

Results:

The integration of molecular monitoring revealed a proportion of low-density infections that RDTs failed to detect. Notably, individuals in urban communities and those reporting recent fever were more likely to harbor asymptomatic malaria infections. These findings suggest that urban residents, although less associated to malaria infection than rural residents, may be serving as a transmission reservoir. The relationship between low density infections and individuals who recently reported fever may reflect recent antimalarial treatment or natural clearance, and thus have lingering parasites in their blood.

Conclusion:

The study highlights the limitations of HRP2-based RDTs in detecting low-density infections and underscores the potential of molecular tools like PlasmoPod in malaria surveillance. By identifying elusive transmission reservoirs and tracking parasite importation, molecular

monitoring can play a crucial role in achieving malaria elimination. The findings advocate for the broader implementation of molecular diagnostics in malaria programs, especially in areas with low transmission, to enhance the detection and targeting of hidden reservoirs of infection.

Background

Essential to malaria control strategy is efficient, accurate identification of *Plasmodium* infection among human populations at risk to appropriately adjust targeted interventions. In 2022, an estimated 345 million malaria rapid diagnostic tests (RDTs) were directly distributed to national malaria programs (NMPs), mostly in Sub-Saharan Africa (1) with the purpose of providing quick diagnostic capability during clinical visits and field studies. These RDTs have become an essential tool in all NMPs as a means to effectively identify, treat and track cases. The majority of RDTs target the *Plasmodium falciparum* histidine-rich protein 2 (HRP2) antigen, which can cross-react with the closely related *P. falciparum* HRP3 antigen in high-density infections (6). Other malaria proteins targeted by the RDT are lactate dehydrogenase (pLDH) or aldolase to detect non-*falciparum* parasites (7, 8) and are more commonly used in combo RDTs with HRP2. *P. falciparum* accounts for the majority of malaria cases and deaths (1), as such HRP2-targeting RDTs are the most common diagnostic used globally and are much more sensitive than non-HRP2 RDTs (2). However, evidence has been accumulating that some *P. falciparum* strains evade RDT diagnosis when the HRP2 target antigen becomes nonfunctional in the parasite genome as shown in strains that carry *hrp2* and *hrp3* deletions (3, 4). These strains may become especially problematic in low-transmission areas, where selective pressures promote their expansion and compromise diagnosis/monitoring by RDT (5). Additionally, RDT limit of detection (<100 parasites/μl) is insufficient for monitoring low density infections, which allow for continued transmission and hamper elimination efforts (6, 7). Conversely, where false negatives underestimate malaria burden, significant proportions of RDT's results can also be false positive, usually due to the persistence of the HRP2 antigen post-treatment with anti-malarial drugs (8). Individuals may continue to be false RDT positive for up to 15 or more days post treatment (9). Especially for field studies in regions with year-round transmission, if false positive results are significant, this may overestimate malaria prevalence, and bias downstream analysis, potentially affecting control strategy decisions. For NMPs approaching malaria elimination, it is essential to have sensitive, robust, fast and easy to deploy diagnostic methods with a low limit of detection and are not vulnerable to *hrp2* and *hrp3* deletions providing a granular and accurate measure of malaria prevalence.

Molecular malaria surveillance with higher sensitivity can identify transmission reservoirs that go undetected by RDTs (10). Positive samples by molecular techniques can be quantified, unlike RDTs, using the *P. falciparum* WHO International Standard for nucleic

acid amplification technique (NAAT) based assays (11), and assays have also been developed to distinguish between multiple *Plasmodium* species, as some can occur as low density co-infections (12-14). Molecular monitoring further acts as an early-warning tool for detection of increasing prevalence of *P. falciparum* strains able to evade diagnostic tools (15, 16).

Quantitative reverse transcription polymerase chain reaction (RT-qPCR, but referred to as qPCR throughout this manuscript) has a lower limit of detection than RDT and can quantify infections (8, 17, 18) and thus makes it an ideal method for molecular surveillance. Implementation of the qPCR method typically requires specialized equipment, cold-chain sensitive and expensive reagents, advanced laboratory infrastructure and well trained staff. To address these logistical limitations, a portable, cartridge-based, user-friendly qPCR instrument, PlasmoPod (Diaxxo AG, Zurich, Switzerland), without the need for cold-chain or purchase of reagents, and shown to be as sensitive as currently used laboratory-based qPCR, was developed (19). Cartridges are provided with all assay reagents in a lyophilized form within each well so with the addition of extracted nucleic material and immersion oil, qPCR results can be generated within 30 minutes. Advantageous to the PlasmoPod technology platform is its versatility to detect various targets depending on the open access and flexible assay design, including SARS-CoV2 variants of concerns (20). Due to the novelty of PlasmoPod, its utility as a molecular monitoring method on a larger scale incorporated into local malaria surveillance programs has yet to be demonstrated.

Bioko Island, Equatorial Guinea, is a malaria endemic island off the coast of Cameroon in Central-West Africa with a population of approximately 270,000 inhabitants (21). Malaria morbidity and mortality has been significantly reduced since intensive control interventions were implemented starting from 2004 by the Bioko Island Malaria Elimination Project (BIMEP) (21-23). The majority of Bioko's residents live in Malabo city, the capital of Equatorial Guinea, in the northern part of the island. Despite BIMEP's progress, the island is vulnerable to malaria importation from its mainland neighbors (23, 24), and prevalence appears to have plateaued similar to other endemic regions across Africa since 2016 (1).

Here, we describe the outcome of a study of 1,500 dried blood spots (DBS) obtained from the BIMEP 2023 malaria indicator survey (MIS) on Bioko Island. These samples were analyzed with the PlasmoPod device at the local reference laboratory in Baney district. The qPCR results were compared to RDT results from the same sample set to provide insight into the added value of molecular malaria monitoring in the field.

Methods

Sample Collection

During the 2023 MIS conducted between August and October, DBS samples were collected from surveyed volunteers. At the start of each MIS visit to a household, participating adult volunteers provided informed consent and legal guardians gave consent for underage participants (<18 years). Each volunteer then answered a questionnaire similar to previous MIS (25) and provided a blood sample from finger prick for RDT diagnosis (CareStart Malaria HRP2/pLDH Combo RDT) and a DBS on Whatman filter papers (GE healthcare Ltd, Forest farm, Cardiff UK). On each filter paper, four blood spots of 0.5 inches (1.27 cm) in diameter were produced, with each spot representing around 50 µl of blood. Selected survey data was used to investigate epidemiological subpopulations (Supplementary Table S1). Household sampling was based on population density of primary sampling units (PSUs) constructed from 1x1 km map-areas that make up Bioko's mapping grid (26).

DNA extraction

A whole blood DNA extraction method developed by the Malaria Research Program at the University of Maryland, Baltimore, USA, was applied to collected filter papers (27). Briefly, one full blood spot of the DBS was cut (~50µl blood) and then incubated in lysis buffer (Guanidine thiocyanate, Triton X100, 0.5M EDTA, 1M trizma hydrochloride, isopropanol, 6M HCl, and 2-mercaptoethanol) with gentle shaking for 2 hours at 65°C -75°C. Samples underwent two washes (Wash 1: same as lysis buffer without 2-mercaptoethoanol; Wash 2: ethanol and isopropanol (1:1 ratio), 1M trizma hydrochloride, and 5M NaCl) before eluting with TE buffer by centrifugation. Extracted DNA was stored at -80°C until use. Suppliers of lysis and wash reagents have been published previously in supplementary materials (27).

Quantitative Reverse Transcription Polymerase Chain reaction – PlasmoPod

PlasmoPod 20-well cartridges for *Plasmodium* spp. targeting the 18S rDNA/rRNA by RT-qPCR supplied by Diaxxo AG (Zurich, Switzerland) (19) were used. Pspp primers used in the cartridges have been previously published, amplifying highly conserved rDNA and rRNA in chromosomes 1, 5, 7, 11, and 13 found only in the *Plasmodium* genus (28). Briefly, 4.5µl of extracted nucleic acid was added to each well, each sample run in duplicate, and covered in immersing oil, then run on the Diaxxo device, outputting results in approximately 30 minutes. Results were generated automatically by the PlasmoPod software, assigning

positive and negative signals based on Cq curves, where any Cq < 40 was considered positive (19). Serial dilutions of WHO standard were performed with PlasmoPod to quantify parasite densities (13). The total time needed from extraction to results for 96 samples, for example, took 3 days approximately. 12-14 hours for nucleic extraction from DBS, and 4 hours for qPCR with PlasmoPod assuming 10 samples per cartridge. Molecular results for all 1500 samples were generated within 2-3 weeks after receiving DBS papers at the laboratory. Parasite density of infections was quantified by obtaining Cq values using PlasmoPod for known concentrations of serial WHO standard dilutions for DNA from NAAT-based assays from 5 to 500,000 parasites/μl (p/μl) (11). From the resulting standard curve ($R^2 = 0.86$), parasite densities were quantified using slope (m = -3.34) and y-axis intercept (y-int = 36.43) (Supplementary Figure S1). Infections were considered low density if parasite density was under 100 p/μl and considered below the limit of detection for RDTs.

Data Analysis

The generated qPCR data were linked to the MIS dataset using each volunteer's unique ID code. Socioeconomic status (SES) for each individual was generated based on household assets and utilities, scoring each household using principal components analysis (PCA). Assets include ownership of radio, television, VCR/DVD, computer, telephone, clock, watch, sofa, table, armoire, cabinet, fans, air conditioning, refrigerators, stove, washing machine, and car, as well as utilities such as public/private water source, protected/open water source, electricity source, toilet type and sanitation mechanism. Statistical analysis and data visualization was conducted in R (v4.3.2) using the following packages: readxl, ggplot2, dplyr, reshape, pscl and car. T-test p-values were calculated when comparing the statistical significance between two groups by the difference in mean of Cq values. Binomial generalized linear univariate and multi-variate models were explored to produce odds ratios and 95% confidence intervals (CI). Multi-variate models were built from the univariate model, where each subgroup was included if the overall fitness of the model remained or improved. Model fitness was measured using McFadden's pseudo-R^2 values (29). The variance inflation factor (VIF) was calculated between each subpopulation to determine multi-co-linearity, where a subpopulation was excluded from the regression model if it was highly correlated to another subpopulation (30). Wald test p-values were used to measure statistical significance between a subpopulation's reference and comparison group within each regression model.

Results

Data descriptives

From the 2023 MIS, 1,500 DBS filter papers of enrolled volunteers were selected: 1,400 samples from RDT positive (+) individuals and 100 samples from RDT negative (-) individuals sharing a household with an RDT (+) member. PlasmoPod results were generated for all samples, of which 21 (1.4%) were excluded for high variance between Cq results of duplicate wells (> 2 Standard Deviations). A total of 1,479 samples were included, of which recorded RDT diagnosis include 994 (67.2%) HRP2-only positive, 31 (2.1%) pLDH-only positive, 356 (24.1%) HRP2/pLDH positive, and 98 (6.6%) HRP2/pLDH negative, respectively (Table 1). All Bioko districts were represented in the sample set (Malabo = 1034 (69.9%); Baney = 184 (12.4%); Luba = 103 (6.9%); Riaba = 158 (10.7%)), where 51.2% (n=757) were collected in rural communities. Slightly more males were represented (male = 51.5%; female = 48.5%), and the median age of all participants was 16 years (Range: 7 months – 80 years). A number of respondents reported having fever in the previous 14 days (n=223; 15.1%), of which 97.3% (217/223) had a positive RDT result. Travel history was reported within the previous 8 weeks, of which 90 volunteers (6.1%) reported travel activity, with the majority traveling to the continental region of Equatorial Guinea (83/90; 92.2%). Among malaria prevention behaviors, 31.0% (n=459) reported sleeping under a bednet the night before, and 64.8% receiving indoor residual spraying (IRS) within the previous 12 months (n=959). Median parasite density across all samples was approximately 3,862 p/µl (Cq = 24.4) with a range of 0.45 – 959,480 p/µl (Cq 16.2 – 37.8).

Table 1. Distribution and positivity rate of epidemiological subgroups by 18S RT-qPCR (qPCR). Median parasite density (parasite per microliter) is reported.

Subpopulation	Subgroup	N (%)	RT-qPCR (+) %	Median p/µl
RDT	Negative	98 (6.6)	26.5	121
	HRP2	994 (67.2)	71.5	1,176
	pLDH	31 (2.1)	71.0	7,970
	HRP2/pLDH	356 (24.1)	83.4	28,516
Age group	<5	128 (8.7)	66.4	15,333
	5-14	513 (35.9)	69.1	6,045
	15-45	715 (48.3)	73.1	2,467
	45+	105 (7.1)	78.1	1,202
District	Malabo	1034 (69.9)	72.1	4,589
	Baney	184 (12.4)	70.7	2,933
	Luba	103 (7.0)	70.9	1,196
	Riaba	158 (10.7)	69.0	2,467
Sex	Female	718 (48.5)	67.7	3,308
	Male	761 (51.5)	75.0	4,433
SES	Low	654 (44.2)	75.4	4,283
	Middle	313 (21.2)	70.9	5,089
	High	512 (34.6)	66.8	2,346
Community	Rural	757 (51.2)	74.9	5,088
	Urban	722 (48.8)	67.9	3,034
Sick (14 days)	No	1214 (82.1)	71.6	3,731
	Yes	258 (17.4)	70.9	4,588
Fever (14 days)	No	1246 (84.2)	71.3	3,797
	Yes	223 (15.1)	72.6	4,436
Slept under net	No	943 (63.8)	72.6	3,998
	Yes	459 (31.0)	67.5	3,862
Travel (8 weeks)	No	1329 (93.9)	71.6	3,862
	Yes	90 (6.1)	72.2	3,034
IRS	No	520 (35.2)	71.9	4,138
	Yes	959 (64.8)	71.2	3,731
HH density	low (<1 person/room)	701 (47.4)	70.2	5,359
	high (>1 person/room	778 (52.6)	72.6	2,736

Proportion of positivity rate per subgroup

Stratifying by RDT results, 26.5% of RDT (-) were positive by qPCR (i.e. RDT false negatives), and 74.7% of RDT (+) results were positive for qPCR, suggesting 25.3% were false RDT positives. RDT (-) had lower parasite density (median parasite density = 121 p/µl) than all RDT (+) groups (HRP2 median parasite density = 1,176 p/µl; pLDH RDT median parasite density = 7,970 p/µl; and HRP2/pLDH RDT median parasite density = 28,516 p/µl) (Table 1). Among age groups, older adults (45+ years old) had the highest proportion of

qPCR (+) (78.1%) while <5 year-olds had the highest parasite density infections (median parasite density = 15,333 p/µl). By SES, the lowest quintile had the highest proportion of qPCR positives (75.4%), and the highest SES quintile had the lowest proportion of qPCR positives (66.8%).

Parasite density stratified by RDT negative versus RDT positive

Parasite quantification of positive qPCR results stratified by RDT results (negative *versus* positive) illustrate false negative RDTs typically were lower parasite density infections compared to RDT positives (121 p/µl *versus* 4137 p/µl; p = 0.05) (Figure 1). Interesting, this observation was not observed when further stratifying epidemiological subpopulations by RDT results (Supplementary Table S2).

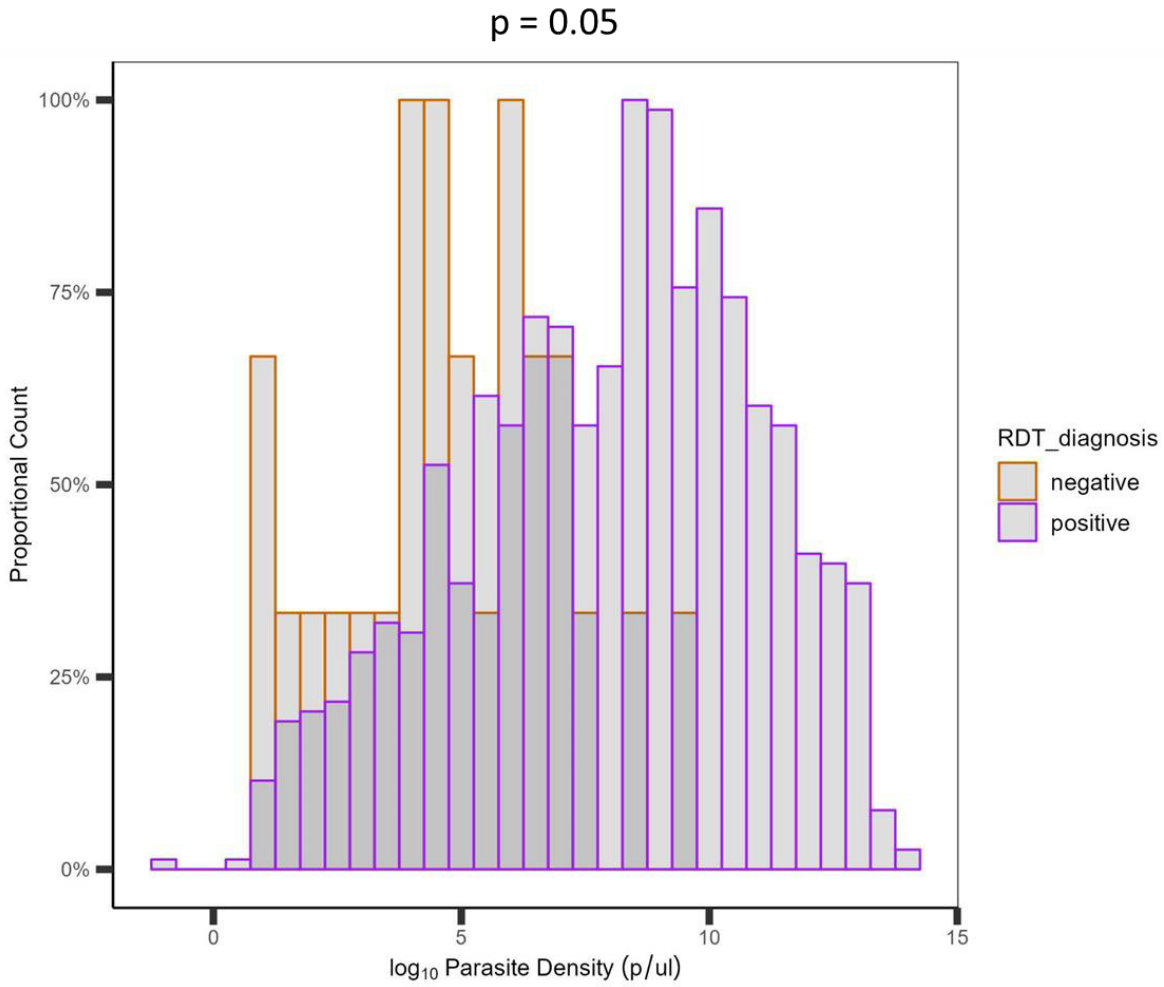


Figure 1. Histogram of log-scaled parasite density by qPCR (p/µl) stratified by RDT results (negative and positive). Orange represents parasite density of false negative RDTs and purple represented parasite density of RDT positives. Y-axis values are scaled to show the proportion count per group. P-values were calculated using t-test method.

Subpopulation groups associated with malaria infection by qPCR

To measure malaria prevalence by qPCR among subpopulations, a multivariate logistic regression model was constructed (Figure 2). To minimize selection bias, all RDT negative results were excluded from the model. The categorical variable for age groups was condensed to optimize the sample sizes of subgroups ($N_{<14}$ = 616; N_{15+} = 765). The subpopulation of reported sickness was excluded from the model due to high multi-co-linearity with reported fever (VIF > 10). Subgroups with a statistically significant association with malaria infection by qPCR compared to their respective reference groups include: individuals over 15 years of age, individuals from Baney, males, high SES, urban communities, and bednet users (Supplementary Table S3). Older individuals (over 15) had higher odds of being qPCR positive compared to younger individuals (OR = 1.31; 95% CI 1.01 – 1.71; p = 0.04). Further, those from Baney had the 53% increased odds of being qPCR positive compared to individuals from Malabo (OR = 1.53; 95% CI: 1.02 – 2.36; p = 0.05). Interestingly, males appeared to have 43% higher odds of being qPCR positive than females (OR = 1.43; 95% CI: 1.11 – 1.85; p = 0.006). Individuals in urban communities were less likely to be qPCR positive compared to rural communities (OR = 0.55; 95% CI: 0.41 – 0.75; p < 0.001). As expected, individuals who slept under a bednet the previous night (i.e. bednet users) had lower odds of being qPCR positive compared to non-bednet users (OR = 0.65; 95% CI: 0.50 – 0.85; p = 0.001).

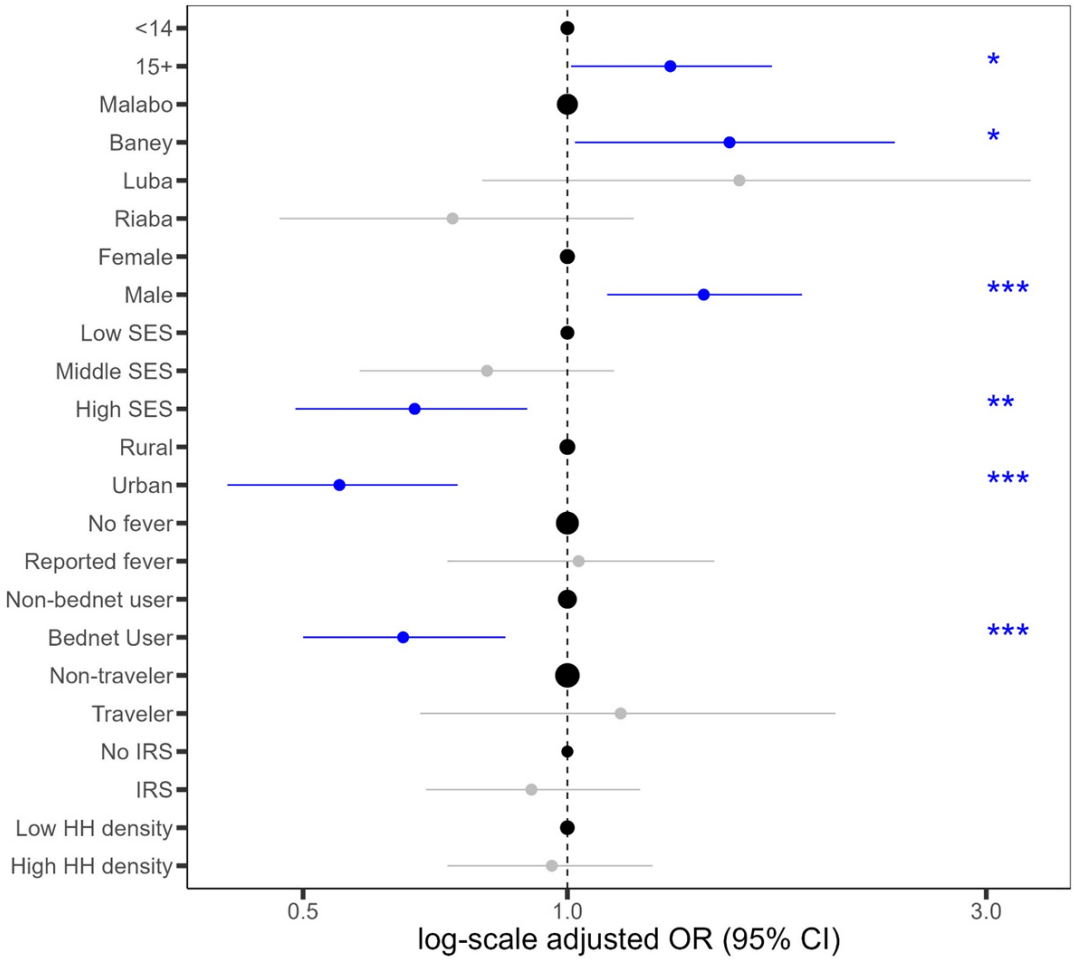


Figure 2. Association of subpopulations to qPCR (+) reported as adjusted odds ratios (OR) estimated from a multivariate logistic regression model among MIS subpopulations on Bioko Island, Equatorial Guinea. OR values are reported on a log scale. Dotted line represents the reference value (OR = 1). Black dots represent reference subgroups and are sized according to sample size. Lines on points represent 95% confidence intervals, and * symbols represent t-test p-values (* < 0.05; ** < 0.01; *** < 0.001). Statistically significant associations are marked in blue.

Subpopulation groups association to low density infections

To measure the association of population subgroups and low density infections (< 100 p/µl), a multivariate logistic regression model was generated with only qPCR (+) samples (n = 1,057) (Figure 3). As observed previously, the subpopulation of reported sickness was excluded from the model due to high multi-co-linearity with reported fever (VIF > 10). Variables with a significant association to low density infections include: false negative RDT results, urban communities, reported fever, and high-density households (Supplementary Table S4). As expected, individuals with a false negative RDT had over 4 times the odds of having a low density infections compared to RDT positive individuals (OR = 4.63; 95% CI: 1.92 – 11.04; p <0.001). Individuals from urban communities had 1.84 times the odds of having a low density infection compared to rural communities (95% CI: 1.22 – 2.80; p =

0.004). Individuals with a reported fever in the previous 14 days had higher odds of having a low density infection compared to individuals with no reported fever (OR = 1.68; 95% CI: 1.08 – 2.56; p = 0.02). Higher density households (>1 person/room) had 64% higher odds of having a low density infections compared to lower density households (OR = 1.64; 95% CI: 1.14 – 2.38; p = 0.008).

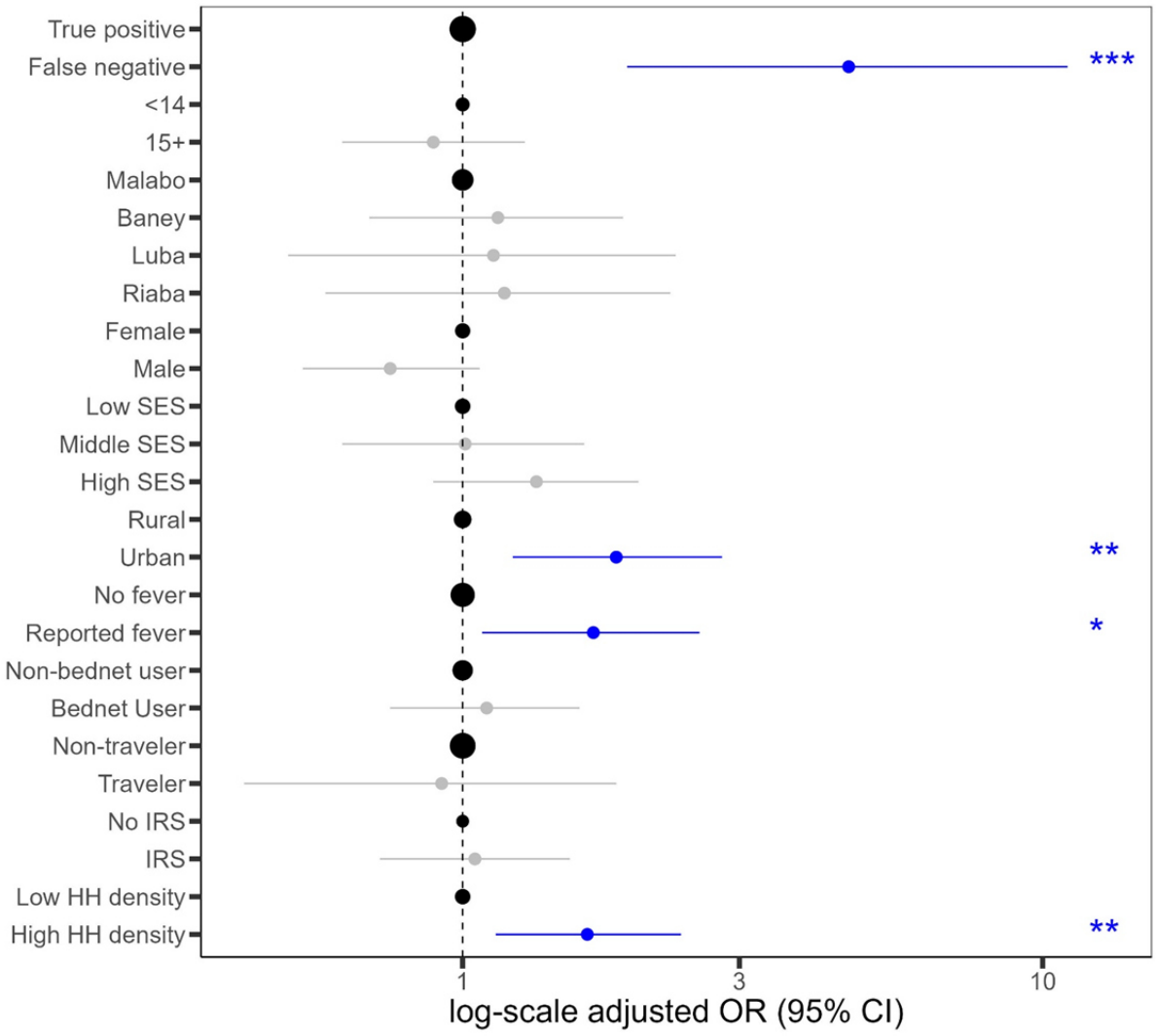


Figure 3. Association of subpopulations to low density infections (less than 500 p/µl) reported as adjusted odds ratios (OR) estimated from a multivariate logistic regression model among MIS subpopulations on Bioko Island, Equatorial Guinea. OR values are reported on a log scale. Dotted line represents the reference value (OR = 1). Black dots represent reference subgroups and are sized according to sample size. Lines on points represent 95% confidence intervals, and * symbols represent t-test p-values (* < 0.05; ** < 0.01; *** < 0.001). Statistically significant associations are marked in blue.

Discussion

The BIMEP aims to eliminate malaria from Bioko Island in the near future, and the continuous surveillance of circulating parasites is essential for prevention of parasite resurgence from low density reservoirs or strains circumventing RDT diagnosis by lacking

functional expression of *hrp2* or *hrp3* (4, 10). When selective pressure is highest (i.e. low transmission settings), these parasites may have an increased evolutionary advantage due to the increased difficulty to detect infections in the general population, leading to their continued contribution to transmission (31, 32). In the event of any interruption in malaria control efforts before complete elimination, these reservoirs can act as a source to undo progress, and lead to a return of pre-control levels in malaria morbidity and mortality (33-35). Molecular monitoring techniques, such as qPCR, is typically a specialized, resource rich laboratory technique. However PlasmoPod was designed as a more rapid, cost-efficient and more accessible qPCR tool for malaria monitoring in the field (19). This study is the first report of PlasmoPod use on a large scale screening of DBS samples for *Plasmodium* spp. from the 2023 MIS on Bioko Island, Equatorial Guinea. The majority of samples were selected from individuals with a positive RDT result to optimize the rate of successful extractions from DBS as it was the first time being performed at the Baney Reference laboratory. RDT (-) samples were selected from households with an RDT (+) member, so the actual rate of false negatives is likely lower. This selection criteria for RDT (-) individuals was done to improve identification of low density infections that cluster at higher rates in such households (36). Although these false negatives were not tested for *hrp2* or *hrp3* deletions, RDT (-) results from Equatorial Guinea have been linked to strains with HRP2 and HRP3 deletions (8, 37, 38). However, most infections are polyclonal on Bioko so their expansion is likely hampered by a co-infection with a non-deletion carrying strain (8, 39). These strains may become a larger issue for BIMEP should Bioko reach low-transmission levels.

The lower proportion of qPCR positives among RDT results highlights the occurrence of RDT false positives, which may overestimate malaria prevalence, and is likely linked to lingering parasite proteins post-clearance of *P. falciparum* (8, 9, 40, 41). This is also reflected in the regression models as individuals reporting a fever in the previous 2 weeks had much higher odds of having a low density infection, but this associative relationship was not observed with qPCR positivity (independent of parasite density). It can be assumed some of these individuals are recovering from a recent *Plasmodium* infection and recently took anti-malaria treatment (40, 41). Also, considering malaria exposure remains high on Bioko, some false positive RDT results may occur among individuals who have spontaneously cleared parasites through naturally acquired immunity (42). Recent anti-malarial treatment is a question included in the MIS, but was not included in the analysis here since the question was

not collected for all age groups and the data had high missingness. In low-transmission areas, differentiating between false positive RDT due to recent anti-malarial treatment or spontaneous clearance may offer greater detail on transmitting parasites as it may be a reflection of parasite virulence. Overall, using only RDT diagnostic methods to measure prevalence during an MIS may overestimate malaria burden, while also underestimating low density infections contributing to transmission (43, 44). Application of a more affordable, rapid, robust and user-friendly qPCR device, like PlasmoPod, into malaria surveillance activities can offer a finer-scaled perspective of malaria burden.

Several subpopulations were associated with infection by qPCR, identifying those groups with higher probability of malaria burden at the molecular level. Older individuals (15+) had strong association to infection by qPCR compared to younger individuals. This is likely due to the successful control measures implemented by the BIMEP, as shifts in age-specific malaria burden can occur by delaying naturally acquired immunity as prevention, surveillance and treatment protect younger at-risk populations (45, 46). Older individuals were not more associated to low density infections as their acquired immunity over decades of exposure may allow for increased tolerability to higher density infections (47, 48). Higher density infections are more likely to result in gametocytemia (49), and this group may be an important reservoir of asymptomatic carriers maintaining transmission. Further, older children (6-15 years) often have low density infections (45) which in this study's regression models was combined with under 5s to optimize the integrity of the model, but may have subsequently diluted the associative measurement between low density infections and age. Of interest, males having a higher likelihood of infection may reflect an increased occupational exposure to malaria-transmitting mosquitos as well as cultural characteristics (50-52), where on Bioko, males more frequently work outside in the jungle and outdoor biting rates of malaria-transmitting *Anopheles* mosquitos have increased (22). This is supported by the significant negative association between urban communities to malaria infection compared to rural communities, as historically there is a higher observed prevalence of malaria in rural areas on Bioko (25). Although individuals in urban communities had lower odds of malaria infection than those in rural settings, they did have a higher probability of having lower density infections. This is somewhat contrary to expectations considering urban environments on Bioko are subject to increased human density and migration to and from the mainland of Equatorial Guinea facilitating malaria transmission (24, 53). The low density infections in urban communities may be a reflection of more accessibility to health clinics and antimalarial

treatment, provided free of charge in Equatorial Guinea (54), subduing prevalence in urban environments despite its potentially higher transmission intensity. In general, low density infections, albeit accounting for a smaller proportion of overall infections on Bioko, likely contribute to transmission (55), and strategies to target these urban subpopulations may need to be expanded once BIMEP achieves low-transmission levels of malaria (56, 57). In addition to adjusting control strategies to account for imported cases, a focus on those hidden malaria transmission reservoirs will be required to effectively eliminate malaria from Bioko Island.

Some limitations exist within this analysis. The majority of samples selected for qPCR were RDT (+) and therefore over-represent higher parasite density infections. To minimize the inherent bias, RDT positives and negatives were assessed separately. Further, RDT (-) were selected from households with a RDT(+) member, increasing the probability of being false negative (36) and may overestimate the rate of false negative RDTs. This selection criteria was used to optimize the probability of identifying positive and low density infections. This study's results complement historical observations of malaria burden on Bioko (22, 25, 36, 53), so although there is the presence of selection bias, we believe the adjustments to the analysis have minimized its effect. Ultimately, this study aimed to illustrate PlasmoPod's effectiveness as an accessible, robust and cost-efficient technique to complement malaria prevalence estimations by RDT in field studies. Here, we focused primarily on RDT (+) results and investigated possible reservoirs of low density infections among the residential population on Bioko Island. Considering the focus of the study, adjusting regression models accordingly, and understanding the inherent biases, the integrity of the overall interpretation of data remains.

Incorporating molecular monitoring in standard NMP surveillance has been recognized and is promoted by the WHO (33, 58). As malaria transmission dynamics change in response to effective interventions, NMPs must continuously adapt their prevention/control strategies and associated resource allocation. Inherent within the successfully reduction in malaria transmission is the subsequent shift in malaria burden (59, 60) and the need to adjust surveillance efforts to find unidentified transmission reservoirs, which can be challenging relying solely on conventional RDTs. Molecular monitoring offers an accurate alternative to RDTs with increased sensitivity, but comes with its challenges (61). PlasmoPod may offer a more accessible substitute to traditional laboratory techniques without significant investment in staff training, supply chain management and equipment but comparable performance. The lower limit of detection by qPCR using PlasmoPod was 0.2 p/µl (19), which is much more

sensitive than HRP2 based RDTs (62). PlasmoPod is open source, and the qPCR assays run on the platform can be adapted to broader research questions like inclusion of drug resistance marker monitoring and contact tracing by targeting predefined SNPs (20). Per 20-well cartridge, the laboratory team ran 8 samples performed in duplicate with accompanying positive and negative controls, also in duplicate. The application of PlasmoPod does not outperform the compatibility or speed of RDTs, but it offers a more sensitive and complementary molecular technique for field studies. Its operation forgoes the need for purchasing the reagents and materials needed for more conventional laboratory protocols. Our data strongly suggest that it is feasible to embed PlasmoPod for regular malaria surveillance in moderate to low transmission settings. The qPCR assay can be run in medium throughput, takes only 30 minutes for obtaining results and in combination with automatized data analyses and uploading to cloud based storage may be very helpful for NMPs to arrive at highly accurate results close to real time for decision making.

Conclusion

To eliminate malaria, NMPs must be able to detect and target malaria transmission reservoirs that evade RDT detection, the most commonly used diagnostic method in Sub-Saharan African countries. PlasmoPod is a novel qPCR device that warrants further investigation using field samples. Within the context of the 2023 MIS conducted by the BIMEP, 18S RT-qPCR highlighted malaria burden on Bioko at the molecular level. By the observed odds ratios, individuals in urban communities or recent reported fever were more positively associated to asymptomatic malaria infection, suggesting that they serve as transmission reservoir. Individuals in urban communities may be under increased exposure, but have improved access to health clinics. For those individuals who reported recent fever events, this may reflect recent anti-malaria treatment or spontaneous clearance, which went undetected by RDT. This may be an interesting avenue for further investigation to differentiate between genomic characteristics of strains needing to be cleared by treatment *versus* those cleared via natural immunity. Ultimately, to address recent plateauing trends, BIMEP may need to rely more heavily on molecular monitoring in its surveillance strategy, identifying more elusive transmission reservoirs, trace parasite importation and the sources and sinks to continue their progress towards malaria elimination.

List of abbreviations

BIMEP – Bioko Island Malaria Elimination Project

CI – confidence intervals

DBS – dried blood spots

HRP2 – histidine-rich protein 2

MIS – malaria indicator survey

NMP – national malaria program

NAAT – nucleic acid amplification technique

OR – odds ratio

PCA – principal components analysis

pLDH – Plasmodium lactate dehydrogenase

PSU – primary sampling units

RDT – rapid diagnostic test

SES – socioeconomic status

VIF - variance inflation factor

Declarations

Ethics approval and consent to participate

Ethics approval for the 2023 MIS was provided by the Equatorial Guinea Ministry of Health and Social Welfare and the ethics committee of the London School of Hygiene and Tropical Medicine (approval number 5556). Written informed consent was sought from each participating adult and on behalf of participating children under 18 years of age.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

Work related to the BIMEP activities was funded by a private sector consortium led by Marathon Oil Corporation and the Government of Equatorial Guinea. The funders had no role in study design, data collection, data analysis, data interpretation, the decision to publish, or preparation of the manuscript.

Author's contributions

TCS and CD conceived and designed the study. TCS trained local laboratory staff, managed laboratory operations, performed data processing, analysis and wrote the manuscript. SH contributed to performing the analyses and writing of the manuscript. EN managed laboratory staff and operations. JNG supported laboratory operations, contributed to data analysis and the writing of the manuscript. MKE, RNB, VMN, IEM, ALB, RMOB, JRB, EQR, AGN, VPI, SB, CNE performed all DBS extractions, and generated PlasmoPod results. MRR sponsored and oversaw the 2023 MIS. DSG managed the 2023 MIS and processed all associated data and contributed to downstream analysis. WPP supervised all BIMEP operations, contributed to supporting the Baney Reference laboratory. CAG contributed to the study design, supported interpretation of 2023 MIS epidemiological data, assisted in generating Bioko Island figures, and edited the paper. GAG contributed fundamental funding and resources to the study, contributed to the study design, and edited the paper. JCS contributed to the interpretation of results and edited the paper. CD conceived of the study, contributed essential resources to the study and wrote the paper. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was supported by the National Malaria Control Programme and the Ministry of Health and Social Welfare of Equatorial Guinea, and MCD Global Health through the Bioko Island Malaria Elimination Project (BIMEP). Specifically, we would like to thank the entire BIMEP team for their efforts conducting the 2023 MIS, and especially the laboratory staff at

the Baney Reference Laboratory. Ultimately, we would like to thank Marathon Oil, Noble Energy, AMPCO (Atlantic Methanol Production Company), and the Ministry of Mines and Energy of Equatorial Guinea for their continued efforts and support for malaria control on Bioko Island.

References

1. WHO. World malaria report 2023: World Health Organization; 2023.
2. WHO. Malaria rapid diagnostic test performance: results of WHO product testing of malaria RDTs: round 6 (2016-2018). 2018.
3. Thomson R, Parr JB, Cheng Q, Chenet S, Perkins M, Cunningham J. Prevalence of Plasmodium falciparum lacking histidine-rich proteins 2 and 3: a systematic review. Bull World Health Organ. 2020;98(8):558-68F.
4. Gendrot M, Fawaz R, Dormoi J, Madamet M, Pradines B. Genetic diversity and deletion of Plasmodium falciparum histidine-rich protein 2 and 3: a threat to diagnosis of P. falciparum malaria. Clinical Microbiology and Infection. 2019;25(5):580-5.
5. Verma AK, Bharti PK, Das A. HRP-2 deletion: a hole in the ship of malaria elimination. The Lancet Infectious Diseases. 2018;18(8):826-7.
6. Kavanaugh MJ, Azzam SE, Rockabrand DM. Malaria Rapid Diagnostic Tests: Literary Review and Recommendation for a Quality Assurance, Quality Control Algorithm. Diagnostics. 2021;11(5):768.
7. Watson OJ, Sumner KM, Janko M, Goel V, Winskill P, Slater HC, et al. False-negative malaria rapid diagnostic test results and their impact on community-based malaria surveys in sub-Saharan Africa. BMJ Global Health. 2019;4(4):e001582.
8. Hosch S, Yoboue CA, Donfack OT, Guirou EA, Dangy J-P, Mpina M, et al. Analysis of nucleic acids extracted from rapid diagnostic tests reveals a significant proportion of false positive test results associated with recent malaria treatment. Malaria Journal. 2022;21(1):23.
9. Dalrymple U, Arambepola R, Gething PW, Cameron E. How long do rapid diagnostic tests remain positive after anti-malarial treatment? Malaria Journal. 2018;17(1):228.
10. Stone W, Gonçalves BP, Bousema T, Drakeley C. Assessing the infectious reservoir of falciparum malaria: past and future. Trends in Parasitology. 2015;31(7):287-96.
11. Padley DJ, Heath AB, Sutherland C, Chiodini PL, Baylis SA, the Collaborative Study G. Establishment of the 1st World Health Organization International Standard for Plasmodium falciparum DNA for nucleic acid amplification technique (NAT)-based assays. Malaria Journal. 2008;7(1):139.
12. Schindler T, Robaina T, Sax J, Bieri JR, Mpina M, Gondwe L, et al. Molecular monitoring of the diversity of human pathogenic malaria species in blood donations on Bioko Island, Equatorial Guinea. Malaria Journal. 2019;18(1):9.
13. Kamau E, Alemayehu S, Feghali KC, Saunders D, Ockenhouse CF. Multiplex qPCR for Detection and Absolute Quantification of Malaria. PLOS ONE. 2013;8(8):e71539.
14. Gimenez AM, Marques RF, Regiart M, Bargieri DY. Diagnostic Methods for Non-Falciparum Malaria. Frontiers in Cellular and Infection Microbiology. 2021;11.
15. Agaba BB, Smith D, Travis J, Pasay C, Nabatanzi M, Arinaitwe E, et al. Limited threat of Plasmodium falciparum pfhrp2 and pfhrp3 gene deletion to the utility of HRP2-based malaria RDTs in Northern Uganda. Malaria Journal. 2024;23(1):3.
16. Rogier E, Battle N, Bakari C, Seth MD, Nace D, Herman C, et al. Plasmodium falciparum pfhrp2 and pfhrp3 gene deletions among patients enrolled at 100 health facilities throughout Tanzania: February to July 2021. Scientific Reports. 2024;14(1):8158.
17. Guirou EA, Schindler T, Hosch S, Donfack OT, Yoboue CA, Krähenbühl S, et al. Molecular malaria surveillance using a novel protocol for extraction and analysis of nucleic acids retained on used rapid diagnostic tests. J Scientific reports. 2020;10(1):1-14.
18. Okell LC, Ghani AC, Lyons E, Drakeley CJ. Submicroscopic Infection in Plasmodium falciparum-Endemic Populations: A Systematic Review and Meta-Analysis. The Journal of Infectious Diseases. 2009;200(10):1509-17.
19. Bechtold P, Wagner P, Hosch S, Gregorini M, Stark WJ, Gody JC, et al. Development and evaluation of PlasmoPod: A cartridge-based nucleic acid amplification test for rapid malaria diagnosis and surveillance. PLOS Global Public Health. 2023;3(9):e0001516.
20. Bechtold P, Wagner P, Hosch S, Siegrist D, Ruiz-Serrano A, Gregorini M, et al. Rapid Identification of SARS-CoV-2 Variants of Concern Using a Portable peakPCR Platform. Analytical Chemistry. 2021;93(49):16350-9.

21. Fries B, Guerra CA, García GA, Wu SL, Smith JM, Oyono JNM, et al. Measuring the accuracy of gridded human population density surfaces: A case study in Bioko Island, Equatorial Guinea. PLOS ONE. 2021;16(9):e0248646.
22. García GA, Janko M, Hergott DEB, Donfack OT, Smith JM, Mba Eyono JN, et al. Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions. Malaria Journal. 2023;22(1):72.
23. Guerra CA, Citron DT, García GA, Smith DL. Characterising malaria connectivity using malaria indicator survey data. Malaria Journal. 2019;18(1):440.
24. Bradley J, Monti F, Rehman AM, Schwabe C, Vargas D, Garcia G, et al. Infection importation: a key challenge to malaria elimination on Bioko Island, Equatorial Guinea. Malaria Journal. 2015;14(1):46.
25. Cook J, Hergott D, Phiri W, Rivas MR, Bradley J, Segura L, et al. Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. Malaria Journal. 2018;17(1):62.
26. García GA, Hergott DEB, Phiri WP, Perry M, Smith J, Osa Nfumu JO, et al. Mapping and enumerating houses and households to support malaria control interventions on Bioko Island. Malaria Journal. 2019;18(1):283.
27. Zainabadi K, Adams M, Han ZY, Lwin HW, Han KT, Ouattara A, et al. A novel method for extracting nucleic acids from dried blood spots for ultrasensitive detection of low-density Plasmodium falciparum and Plasmodium vivax infections. Malaria Journal. 2017;16(1):377.
28. Kamau E, Tolbert LS, Kortepeter L, Pratt M, Nyakoe N, Muringo L, et al. Development of a highly sensitive genus-specific quantitative reverse transcriptase real-time PCR assay for detection and quantitation of plasmodium by amplifying RNA and DNA of the 18S rRNA genes. Journal of clinical microbiology. 2011;49(8):2946-53.
29. McFadden D, Zarembka P. Frontiers in econometrics. Academic Press New York; 1974.
30. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis: John Wiley & Sons; 2021.
31. Schneider KA, Salas CJ. Evolutionary genetics of malaria. Frontiers in Genetics. 2022;13.
32. Schneider KA. Charles Darwin Meets Ronald Ross: A Population-Genetic Framework for the Evolutionary Dynamics of Malaria. In: Teboh-Ewungkem MI, Ngwa GA, editors. Infectious Diseases and Our Planet. Cham: Springer International Publishing; 2021. p. 149-91.
33. WHO. A framework for malaria elimination: World Health Organization; 2017.
34. Nasir SMI, Amarasekara S, Wickremasinghe R, Fernando D, Udagama P. Prevention of re-establishment of malaria: historical perspective and future prospects. Malaria Journal. 2020;19(1):452.
35. Ghani AC, Sutherland CJ, Riley EM, Drakeley CJ, Griffin JT, Gosling RD, Filipe JAN. Loss of Population Levels of Immunity to Malaria as a Result of Exposure-Reducing Interventions: Consequences for Interpretation of Disease Trends. PLOS ONE. 2009;4(2):e4383.
36. Hergott DEB, Balkus JE, García GA, Cruz KR, Seilie AM, Masters H, et al. Clustering of subpatent infections in households with asymptomatic rapid diagnostic test-positive cases in Bioko Island, Equatorial Guinea independent of travel to regions of higher malaria endemicity: a cross-sectional study. Malaria Journal. 2021;20(1):313.
37. Berzosa P, González V, Taravillo L, Mayor A, Romay-Barja M, García L, et al. First evidence of the deletion in the pfhrp2 and pfhrp3 genes in Plasmodium falciparum from Equatorial Guinea. Malaria Journal. 2020;19(1):99.
38. Molina-de la Fuente I, Pacheco MA, García L, González V, Riloha M, Oki C, et al. Evolution of pfhrp2 and pfhrp3 deletions in Equatorial Guinea between the pre– and post–RDT introduction. Malaria Journal. 2024;23(1):215.
39. Sepúlveda N, Phelan J, Diez-Benavente E, Campino S, Clark TG, Hopkins H, et al. Global analysis of Plasmodium falciparum histidine-rich protein-2 (pfhrp2) and pfhrp3 gene deletions using whole-genome sequencing data and meta-analysis. Infection, Genetics and Evolution. 2018;62:211-9.
40. Iqbal J, Siddique A, Jameel M, Hira PR. Persistent Histidine-Rich Protein 2, Parasite Lactate Dehydrogenase, and Panmalarial Antigen Reactivity after Clearance of *Plasmodium falciparum* Monoinfection. Journal of Clinical Microbiology. 2004;42(9):4237-41.
41. Markwalter CF, Gibson LE, Mudenda L, Kimmel DW, Mbambara S, Thuma PE, Wright DW. Characterization of Plasmodium Lactate Dehydrogenase and Histidine-Rich Protein 2 Clearance

Patterns via Rapid On-Bead Detection from a Single Dried Blood Spot. Am J Trop Med Hyg. 2018;98(5):1389-96.
42. Roe MS, O'Flaherty K, Fowkes FJI. Can malaria parasites be spontaneously cleared? Trends in Parasitology. 2022;38(5):356-64.
43. Kojom Foko LP, Moun A, Singh V. Addressing low-density malaria infections in India and other endemic part of the world—the opportune time? Critical Reviews in Microbiology.1-17.
44. Laban NM, Kobayashi T, Hamapumbu H, Sullivan D, Mharakurwa S, Thuma PE, et al. Comparison of a PfHRP2-based rapid diagnostic test and PCR for malaria in a low prevalence setting in rural southern Zambia: implications for elimination. Malaria Journal. 2015;14(1):25.
45. Kigozi SP, Kigozi RN, Epstein A, Mpimbaza A, Sserwanga A, Yeka A, et al. Rapid shifts in the age-specific burden of malaria following successful control interventions in four regions of Uganda. Malaria Journal. 2020;19(1):128.
46. Griffin JT, Ferguson NM, Ghani AC. Estimates of the changing age-burden of Plasmodium falciparum malaria disease in sub-Saharan Africa. Nature Communications. 2014;5(1):3136.
47. Doolan DL, Dobaño C, Baird JK. Acquired Immunity to Malaria. Clinical Microbiology Reviews. 2009;22(1):13-36.
48. Barry A, Hansen D. Naturally acquired immunity to malaria. Parasitology. 2016;143(2):125-8.
49. Gonçalves BP, Kapulu MC, Sawa P, Guelbéogo WM, Tiono AB, Grignard L, et al. Examining the human infectious reservoir for Plasmodium falciparum malaria in areas of differing transmission intensity. Nature Communications. 2017;8(1):1133.
50. Tanner M, Vlassoff C. Treatment-seeking behaviour for malaria: a typology based on endemicity and gender. Social science & medicine. 1998;46(4-5):523-32.
51. Bates I, Fenton C, Gruber J, Lalloo D, Lara AM, Squire SB, et al. Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part 1: determinants operating at individual and household level. The Lancet Infectious Diseases. 2004;4(5):267-77.
52. Okiring J, Epstein A, Namuganga JF, Kamya EV, Nabende I, Nassali M, et al. Gender difference in the incidence of malaria diagnosed at public health facilities in Uganda. Malaria Journal. 2022;21(1):22.
53. Guerra CA, Kang SY, Citron DT, Hergott DEB, Perry M, Smith J, et al. Human mobility patterns and malaria importation on Bioko Island. Nature Communications. 2019;10(1):2332.
54. Rehman AM, Mann AG, Schwabe C, Reddy MR, Roncon Gomes I, Slotman MA, et al. Five years of malaria control in the continental region, Equatorial Guinea. Malaria Journal. 2013;12(1):154.
55. Drakeley C, Gonçalves B, Okell L, Slater H. Understanding the importance of asymptomatic and low-density infections for malaria elimination. Towards Malaria Elimination-A Leap Forward. 2018;18:1-20.
56. Doumbe-Belisse P, Kopya E, Ngadjeu CS, Sonhafouo-Chiana N, Talipouo A, Djamouko-Djonkam L, et al. Urban malaria in sub-Saharan Africa: dynamic of the vectorial system and the entomological inoculation rate. Malaria Journal. 2021;20(1):364.
57. WHO. Global framework for the response to malaria in urban areas: World Health Organization; 2022.
58. WHO. Global technical strategy for malaria 2016–2030, 2021 update. J Geneva: World Health Organization. 2021:1-40.
59. O'Meara WP, Mangeni JN, Steketee R, Greenwood B. Changes in the burden of malaria in sub-Saharan Africa. The Lancet Infectious Diseases. 2010;10(8):545-55.
60. Nkumama IN, O'Meara WP, Osier FHA. Changes in Malaria Epidemiology in Africa and New Challenges for Elimination. Trends in Parasitology. 2017;33(2):128-40.
61. Nsanzabana C. Strengthening Surveillance Systems for Malaria Elimination by Integrating Molecular and Genomic Data. Tropical Medicine and Infectious Disease. 2019;4(4):139.
62. Das S, Peck RB, Barney R, Jang IK, Kahn M, Zhu M, Domingo GJ. Performance of an ultra-sensitive Plasmodium falciparum HRP2-based rapid diagnostic test with recombinant HRP2, culture parasites, and archived whole blood samples. Malaria Journal. 2018;17(1):118.

4 Next Generation malaria vaccine development

This chapter contains the following publications:

Thomson-Luque R, **Stabler TC**, Fürle K, Silva JC, Daubenberger C. **Plasmodium falciparum merozoite surface protein 1 as asexual blood stage malaria vaccine candidate.** Expert Review of Vaccines. 2024 Dec 31;23(1):160-73.

Scalsky R, Dwivedi A, **Stabler TC**, Mbambo G, Ouattara A, Lyke KE, Takala-Harrison S, Silva JC. **Whole-genome sieve analysis: identification of protective malaria antigens by leveraging allele-specific vaccine efficacy.** Manuscript under review at Vaccine.

Dwivedi A, Scalsky RJ, **Stabler TC**, Shrestha B, Joshi S, Harris DG, Gandhi C, Munro JB, Ifeonu OO, Ouedraogo A, Tiono AB, Coulibaly D, Ouattara A, Richie TL, Sim BKL, Plowe CV, Lyke KE, Takala-Harrison S, Hoffman SL, Thera MA, Sirima SB, Laurens MB, Silva JC. **Protective targets of PfSPZ-based whole organism vaccines identified from genome-wide sieve analyses of isolates from field efficacy trials.** Manuscript in preparation. Planned submission to Journal of Translational Medicine.

4.1 Plasmodium falciparum merozoite surface protein 1 as asexual blood stage malaria vaccine candidate.

Published in Review of Vaccines, 2024

Taylor & Francis
Taylor & Francis Group

Expert Review of Vaccines

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ierv20

Plasmodium falciparum merozoite surface protein 1 as asexual blood stage malaria vaccine candidate

Richard Thomson-Luque, Thomas C Stabler, Kristin Fürle, Joana C Silva & Claudia Daubenberger

To cite this article: Richard Thomson-Luque, Thomas C Stabler, Kristin Fürle, Joana C Silva & Claudia Daubenberger (2024) *Plasmodium falciparum* merozoite surface protein 1 as asexual blood stage malaria vaccine candidate, Expert Review of Vaccines, 23:1, 160-173, DOI: 10.1080/14760584.2023.2295430

To link to this article: https://doi.org/10.1080/14760584.2023.2295430

EXPERT REVIEW OF VACCINES
2024, VOL. 23, NO. 1, 160–173
https://doi.org/10.1080/14760584.2023.2295430

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Plasmodium falciparum merozoite surface protein 1 as asexual blood stage malaria vaccine candidate

Richard Thomson-Luque [a,b], Thomas C Stabler[d,f,g], Kristin Fürle[a], Joana C Silva[c,d,e] and Claudia Daubenberger[f,g]

[a]Centre for Infectious Diseases-Parasitology, Heidelberg University Hospital, Heidelberg, Germany; [b]Sumaya-Biotech GmbH & Co. KG Heidelberg, Germany; [c]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA; [d]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA; [e]Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade NOVA de Lisboa (GHTM IHMT, UNL), Lisbon, Portugal; [f]University of Basel Basel, Switzerland; [g]Swiss Tropical and Public Health Institute Allschwil, Switzerland

ABSTRACT

Introduction: Malaria represents a public health challenge in tropical and subtropical regions, and currently deployed control strategies are likely insufficient to drive elimination of malaria. Development and improvement of malaria vaccines might be key to reduce disease burden. Vaccines targeting asexual blood stages of the parasite have shown limited efficacy when studied in human trials conducted over the past decades.

Areas covered: Vaccine candidates based on the merozoite surface protein 1 (MSP1) were initially envisioned as one of the most promising approaches to provide immune protection against asexual blood-stage malaria. Successful immunization studies in monkey involved the use of the full-length MSP1 ($MSP1_{FL}$) as vaccine construct. Vaccines using $MSP1_{FL}$ for immunization have the potential benefit of including numerous conserved B-cell and T-cell epitopes. This could result in improved parasite strain-transcending, protective immunity in the field. We review outcomes of clinical trials that utilized a variety of MSP1 constructs and formulations, including $MSP1_{FL}$, either alone or in combination with other antigens, in both animal models and humans.

Expert opinion: Novel approaches to analyze breadth and magnitude of effector functions of MSP1-targeting antibodies in volunteers undergoing experimental vaccination and controlled human malaria infection will help to define correlates of protective immunity.

ARTICLE HISTORY
Received 25 September 2023
Revised 21 November 2023
Accepted 12 December 2023

KEYWORDS
Plasmodium falciparum; merozoite surface antigen 1; vaccine development; controlled human malaria infection; asexual blood stage

1. Introduction

The malaria disease burden batters particularly impoverished tropical and subtropical regions of the world [1]. The latest estimated number of 247 million annual malaria cases, which resulted in 619,000 deaths globally [2], suggests that the control strategies implemented in the last decades [3], which led to reduction in morbidity and mortality [4], are insufficient for malaria elimination. This holds true especially when facing unexpected challenges, such as the recent SARS-CoV-2 pandemic [5] or outbreaks of armed conflict haunting areas endemic to malaria [6].

Vaccines are considered a potentially impactful, additional tool toward reducing and controlling the malaria disease burden [7], yet the history of malaria vaccine development has been long and tedious. In 1967, Ruth Nussenzweig published an important breakthrough toward a first concept of a pre-erythrocytic vaccine aimed at protecting from malaria infection. She immunized mice with the rodent malaria parasite *Plasmodium berghei* by exposing them to sporozoites that had been attenuated by X-ray irradiation [8]. This approach conferred significant sterile protection in mice against the parasite when later challenged with fully infectious *P. berghei* sporozoites delivered by mosquito bites. Since then, progress toward developing an effective malaria vaccine, and especially against pre-erythrocytic stages (commonly known as sporozoite and/or liver-stage vaccines), has faced many challenges [9]. This is based on a number of factors including the complex biology of *Plasmodium* species and their well-adapted interactions with the mosquito vector and the human host [10,11]. Despite these obstacles, arduous efforts resulted in the development of the first generation pre-erythrocytic malaria vaccine against *P. falciparum* called RTS,S/AS01 [9]. RTS,S/AS01 is the first licensed malaria vaccine (Mosquirix™) and is currently in the implementation phase in three African countries [12]. The WHO-specified malaria vaccine efficacy goal of 75% against malaria disease in African children [13] has recently been reported in a phase II clinical trial evaluating the subunit malaria vaccine candidate R21 [14,15]. Moreover, sterile protection against *P. falciparum* malaria infection has also been

CONTACT Claudia Daubenberger claudia.daubenberger@swisstph.ch University of Basel, Petersplatz 1, Basel 4001, Switzerland; Richard Thomson-Luque thomsonluque@sumayabiotech.com Centre for Infectious Diseases-Parasitology, Heidelberg University Hospital, Im Neuenheimer Feld 324, Heidelberg, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/14760584.2023.2295430

Article highlights

- The critical and life-threatening consequences linked to *Plasmodium falciparum* malaria are directly attributable to the presence of asexual blood-stage parasites.
- Blood-stage malaria vaccines seek to diminish/impede the asexual reproduction of malaria parasites. Various clinical trials centered around *P. falciparum* blood-stage candidate vaccines have been performed but, regrettably, none has demonstrated robust protective effectiveness.
- Ongoing endeavors are concentrated on enhancing liver-stage vaccines which, in the event of not exhibiting sterile immunity, cannot safeguard against the potential advancement of liver-stage merozoites into asexual blood-stage parasites.
- The creation of a multi-stage malaria vaccine, combining liver-stage vaccines with an effective blood-stage vaccine, holds the promise of augmenting vaccine-driven protection against clinical disease.
- The merozoite surface protein 1 (MSP1) of *P. falciparum* is a major target of protective immunity and was initially considered an ideal malaria vaccine candidate.
- In monkey models, employing the complete MSP1 protein ($MSP1_{FL}$) for immunization, synergizing conserved domains from both the N- and C-terminal regions of MSP1, resulted in elevated levels of protection against live parasite challenge.
- MSP1 human trials have employed subunits or composite formulations comprising diverse smaller domains, but never the $MSP1_{FL}$, and exhibited poor performance.
- An $MSP1_{FL}$ vaccine adjuvanted with GLA-SE has recently been tested for safety and immunogenicity in a first in human phase Ia trial involving malaria-naive volunteers.
- Antibodies generated through $MSP1_{FL}$ vaccination demonstrated the ability to stimulate a diverse array of Fc-dependent effector functions *in vitro*.
- The logical progression entails delving into the connection between the spectrum of antibody multi-functionality and their effects on asexual blood-stage parasites in a CHMI model available for use in adult African populations.

accomplished by attenuated, purified sporozoite vaccine approaches (PfSPZ Vaccine) [16,17].

In contrast to pre-erythrocytic stage vaccines, blood-stage malaria vaccines are aimed at reducing or blocking asexual reproduction of the malaria parasite in red blood cells (RBCs). Several clinical trials based on blood-stage parasite antigens of *P. falciparum* have been conducted but, unfortunately, as to date, no vaccine has shown strong protective efficacy in phase II field trials [18,19]. Thus, current efforts are focused on refining liver-stage vaccines [20], while less attention is given to the development of new blood-stage candidates or to the refinement of previously tested antigens [21]. Pre-erythrocytic vaccines targeting the sporozoites and/or liver stage offer the advantage that they act from the moment the limited numbers of parasites transmitted by an infected *Anopheline* mosquito bite enter the host until they exit the liver as merozoites and reach the blood stage. However, if these vaccines are not extremely efficient, they will not prevent potential progression of liver-stage merozoites to asexual blood-stage parasites resulting in malaria disease and potentially death. Life-threatening effects associated with *P. falciparum* malaria, such as severe anemia and cerebral malaria, are caused by asexual blood stages [22].

We provide insight into the potential of the merozoite surface protein 1 (MSP1) to become a component of a multi-stage, multicomponent next-generation malaria vaccine. Development of a multistage malaria vaccine representing a combination of pre-erythrocytic subunit vaccines such as RTS'S/AS01 or R21 and asexual blood-stage proteins like MSP1 might have the potential to further enhance vaccine-induced protection against clinical disease in the field. The biology, genetic diversity, and preclinical studies of MSP1 in animal models are presented. Clinical trials in humans that have employed full length or parts of MSP1 alone or in combination with other malaria antigens and formulations are also reviewed.

2. Biology of the MSP1

In 1981, a novel *P. falciparum* schizont-derived protein with relative molecular mass of around 195 kDa was reported by Anthony Holder [23]. Later, the gene structure and specific fragments resulting from the processing of the encoded protein, what was then called P195 protein, were described [24]. This P195 protein is the most abundant protein on the merozoite surface and is today known as MSP1. MSP1 is synthesized from the onset of schizogony as a 195 kDa precursor that undergoes a series of proteolytic cleavages [25,26]. The first proteolytic processing step is performed by the *P. falciparum* subtilisin-like protease PfSUB1 at the time of merozoite egress from the host cell to yield four fragments, namely the p83, p30, p38, and p42 polypeptides [27]. The multipartite MSP1 complex found on the surface of the free merozoite is shed at the time of RBC invasion in a secondary processing step performed by the subtilisin-like protease PfSUB2 [28]. This step mediates the proteolytic cleavage of p42 into two further fragments, p33 and p19 (Figure 1). The cleavage releases the complex containing p33 as soluble fragment, while the C-terminal, GPI-anchored p19 fragment remains on the parasite surface as the parasite enters the RBC [29,30]. The p19 fragment, comprised of two epidermal growth factor-like domains, is then transferred to the developing food vacuole [31]. MSP1 is thought to play a key role upon the first attachment of the free merozoites to RBCs, although other components of the macromolecular complex anchored to MSP1 such as MSP6 [32], MSP7 [33,34], MSPDBL1, and MSPDBL2 [35] are also functionally involved.

The gene *msp1* (PF3D7_0930300) is located on chromosome 9 of the *P. falciparum* 3D7 reference genome, between the genomic positions 1,201,305–1,207,576. It is a single-copy gene with no introns within its coding region (coordinates 1,201,812–1,206,974) and is divided into 17 blocks with varying levels of genetic diversity, where 7 blocks are highly polymorphic, 5 blocks are semi-conserved, and 5 blocks are highly conserved [37–40]. Evident from its name, *msp1* expression is highest during the asexual blood stage of the parasite (Figure 2a). *msp1* block 2 undergoes frequent reorganization from slipped strand mispairing [41,42] and possibly intragenic recombination [38,40,43]. Differences in fragment size and the presence of unique peptide repeats [44] delineate three major block 2 families: RO33, K1, and MAD20. Among these families, K1 and MAD20 alleles are the most diverse since they have varying numbers of unique tri-peptide repeats. RO33 alleles lack these tri-peptide repeats and are monomorphic. K1 and MAD20 alleles are observed at greater frequency globally than

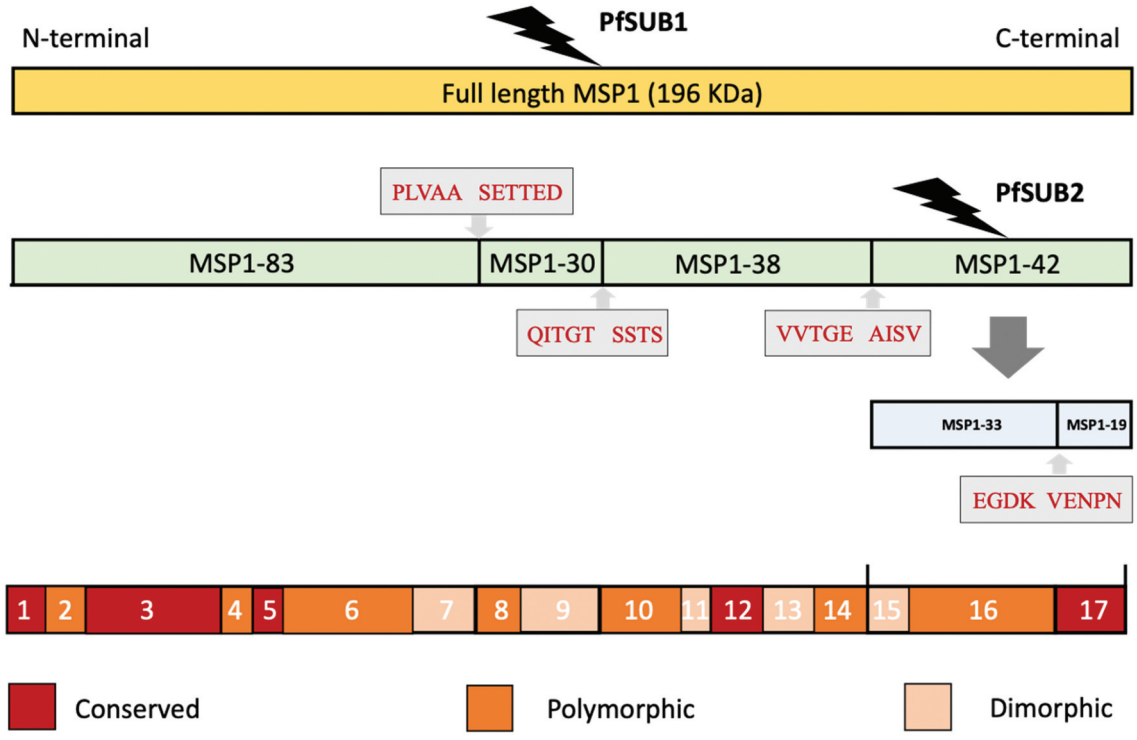
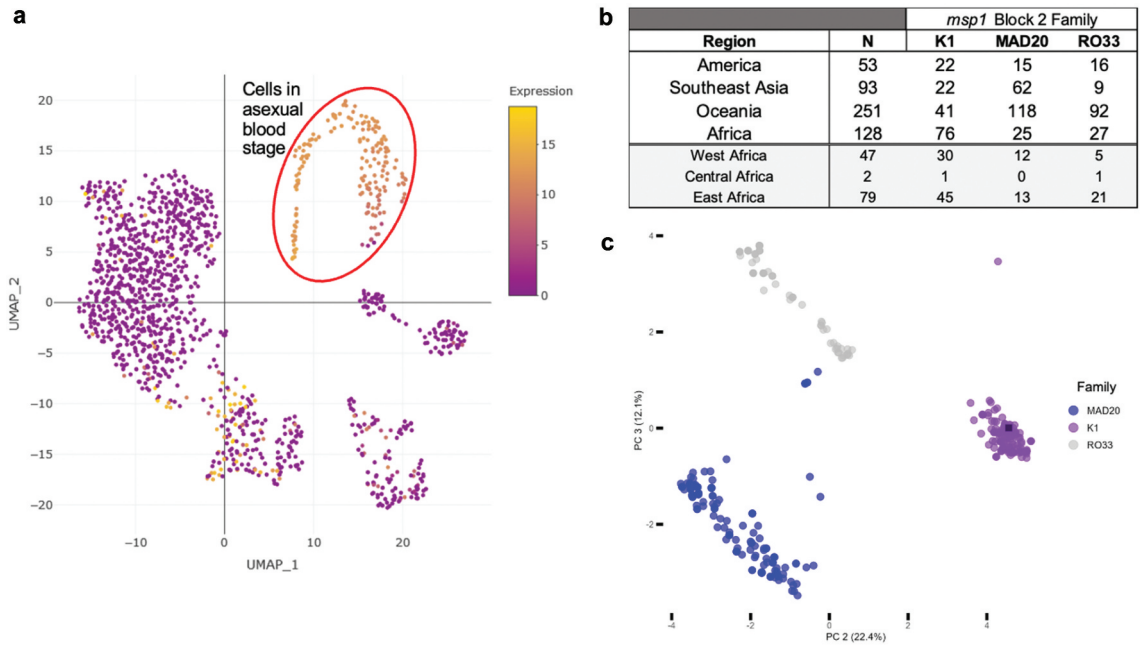


Figure 1. Consecutive enzymatic processing steps of MSP1 protein. MSP1 is produced at the end of schizogony as a precursor protein of around 196 kDa. Once the merozoite reaches maturity, PfSUB1 cleaves MSP1 at specific cleavage sites (shown here in gray squares), leading to formation of four fragments named p83 (MSP1-83), p30 (MSP1-30), p38 (MSP1-38), and p42 (MSP1-42). A second cleavage event of the p42 fragment by PfSUB2 leads to the formation of fragments p33 (MSP1-33) and p19 (MSP1-19). MSP1 is divided in 17 domains [37] based on sequence polymorphisms, whereby 7 domains are highly polymorphic, 5 are semi-conserved, and 5 are highly conserved which are color-marked in red, orange, and pink, respectively. MSP1 has been historically characterized by two prototypic sequences: the MSP1-D from the *P. falciparum* MAD20 strain and the MSP1-F from the WELLCOME strain. These two forms are found in African parasite populations, with distinct distributions observed between East and West Africa [36,148].

RO33 [45–49]. Due to its variable genomic sequence [50], genotyping of *P. falciparum* infections by molecular characterization of *msp1* can differentiate between re-infection and recrudescent infections [51,52] and illuminate genetic diversity and sub-structure in parasite populations [45–47,49,53]. However, also due to regions of high variability in sequence and the presence of insertions/deletions (indels), comprehensive locus genotyping by read mapping of whole-genome sequencing data against a reference genome becomes challenging.

To overcome the challenge of genotyping by read mapping, 37 publicly available or reported *P. falciparum* whole-genome assemblies built from long read sequencing data [54,55] for strains collected from different continents (South East Asia (SEA) = 13; America = 3; Africa = 21) were used to obtain the *msp1* coding region (CDS) and conduct a preliminary analysis of CDS sequence diversity. The average CDS length across all strains was 5092 bp (range: 4884–5166 bp) and nucleotide diversity across the entire. Further, 488 publically available, full-length, *msp1* sequences from *P. falciparum* were downloaded from National Center for Biotechnology Information (NCBI) [39,56–59]. Among all sequences analyzed (n = 525), the average CDS length was 5095 bp (range: 4884–5220 bp), and nucleotide diversity across the entire CDS was relatively low ($\pi_{nucleotide}$ per 100 sites = 1.7%; 95% CI 1.5–1.9) despite only 80% of the nucleotide positions being conserved ($n_{conserved\_sites}$ = 4168). The proportion of each *msp1* block 2 family was consistent with global results obtained from Polymerase Chain Reaction (PCR)-generated allele fragments [45–49], where most strains were classified as K1 (n = 161) or MAD20 (n = 220) (Figure 2b and Supp. Table S1). Across block 2 family types, K1 and MAD20 were observed to have longer sequences on average (K1: 5121 average bp length, range = 4884–5220 bp; MAD20: 5094 average bp length, range = 4484–5172 bp), and RO33 had the shortest average sequence length (RO33: 5068 average bp length, range = 5043–5085 bp). A principal component analysis of the *msp1* sequences shows them clustered according to sequence length and block 2 family designation, illustrating that sequence variance in the dataset is driven in large part by block 2 (Figure 2c and Supp. Fig 1 and 2). PC1 (not shown) accounted for 40% of variance in the sample set, driven by a small subset (n = 10) of the shortest K1 and MAD20 sequences. Variation among all sequences is clearer in PC2 and PC3, where those shorter sequences are still observed as separate from their respective, more frequent, block 2 family clusters. The MAD20 sequences separated from their larger cluster have a larger number of SNPs in downstream *msp1* blocks (Supp. Fig XX). Previous analyses of *msp1* sequence diversity have shown sequence length polymorphism in block 2 as key to *msp1* variation among *P. falciparum* strains, consistent with our results [42,60,61]. Interestingly, most African strains encode K1 variants, while



		msp1 Block 2 Family		
Region	**N**	**K1**	**MAD20**	**RO33**
America	53	22	15	16
Southeast Asia	93	22	62	9
Oceania	251	41	118	92
Africa	128	76	25	27
West Africa	47	30	12	5
Central Africa	2	1	0	1
East Africa	79	45	13	21

Figure 2. Genetic diversity of MSP1 field isolates. (A) UMAP single cell expression profile of the merozoite surface protein gene PF3D7_0930300; https://www.sanger.ac.uk/tool/mca/mca/. Each dot represents a cell. The four quadrants represent the transcriptomes of sporozoites (salivary glands, top left quadrant), sporozoites (hemolymph, lower left quadrant), and ookinetes (lower right quadrant). The top right quadrant, containing the cells in which MSP1 is most highly expressed, is asexual blood-stage parasites. (B) Geographic origin of *P. falciparum* strains from which MSP1 sequences were obtained and frequency of representative alleles based on block 2 classification. Lower rows (in gray) are stratified African samples according to WHO regional delineations. (C) Principal component analysis (PCA) of msp1 sequences aligned with MUSCLE in MegaX v10.0.5. PCA values were calculated using multi-sequence alignment with Rv.4.1.3 library adegenet.

Southeast Asian and Oceania strains encode primarily MAD20 variants. This geographical differentiation is tentative due to the temporal and study differences among analyzed strains.

3. Naturally acquired humoral immune responses against MSP1

Immuno-epidemiological studies have shown that residents from malaria-endemic areas, after repeated exposure to *P. falciparum* infections, are able to attain strain-transcending immunity, allowing for partial protection against clinical disease [64]. One of the key mediators of this acquired immunity seems to be the presence of antibodies targeting asexual blood-stage parasites [65]. Antibodies against MSP1 have been often, but not always, correlated with protection against clinical malaria [66–71]. Contradicting findings might be partially attributed to the organization of the protein, consisting of an alternation between conserved and more polymorphic regions (Figure 1). The early focus of research was set on understanding antibody responses against different subunits of MSP1[72]. MSP1-specific antibody responses in Mali were found to be predominantly directed against the dimorphic parts of MSP1 [72].

One important aspect to clarify was whether antibodies against specific regions, and especially against dimorphic regions of MSP1, have a preponderant role in naturally acquired protection. Using sera from individuals residing in malaria-endemic regions of West Africa, as well as sera from German patients with acute malaria, a correlation between antibodies targeting specific MSP1 regions and a protective immune response against *P. falciparum* when stratified according to age and parasitemia levels was observed [73]. For example, serum samples from adults were more reactive against the dimorphic F5 region of the K1 strain (Figure 3), which is located near the enzymatic processing site between p83 and p30, several months after recovering from an infection when compared to children [73]. Another study in Mali led to more ambiguous results in that no correlation between MSP1 antibody titers and protection against infection was observed in children. Yet, in adolescents, the presence of antibodies to dimorphic fragment M6, pertaining to the N-terminal p83 subunit of MSP1 (MSP1-83) (from MAD20 isolate) correlated with a 50% reduced risk of *P. falciparum* infection and increased ability to control parasitemia [74]. In contrast, in adults living in the same region, the humoral immune response to some of the MSP1 polymorphic regions was associated with an increased risk of infection [74].

In Papua New Guinea, positive correlations of antibody titers against the partially dimorphic C-terminal MSP1-42 (including the conserved MSP1-19) with protection against clinical malaria and severe parasitemia in children were observed [66]. Interestingly, responses to conserved regions such as the C-terminal MSP1-19 subunit only [67] were also shown to correlate with protection when analyzing serum antibodies from Gambia and Sierra Leone in an enzyme-linked immunosorbent assay (ELISA). A systematic review with meta-analysis on the relationship between anti-merozoite antibodies and incidence of *P. falciparum* malaria corroborated that one of the largest protective effects came from IgG against the

conserved MSP1-19 [75]. These findings supported the research shift from focusing mainly on dimorphic regions to also considering conserved regions of the protein.

In The Gambia, the reactivity of antibodies from children aged 3–7 years against the MSP1 block 2 (harbored within the N-terminal MSP1-83 subunit) was tested [76]. MSP1 block 2 is the locus with the lowest inter-population variance in allele frequencies in different regions in Africa and Southeast Asia, indicating immune selection in this part of MSP1 [77]. Antibodies to MSP1 block 2 were strongly associated with protection from *P. falciparum* malaria [76]. Similar results were found when analyzing children from Ghana, with antibodies to MSP1 block 2 being significantly associated with a reduced risk of subsequent clinical malaria [78]. Following data corroborated that the MSP1-83 subunit in the N-terminal end of the protein is a major target of opsonizing antibodies acquired during natural exposure to malaria [79]. In summary, several immune-epidemiological studies suggested that antibodies against C-terminal MSP1-42 subunit, but also to the N-terminal MSP1-83 subunit of MSP1 (both harboring dimorphic and conserved regions), may be relevant to mount a protective humoral response. Yet, despite some initial interest [80–83], the N-terminal part of the protein has not been the focus as a component of an immunization strategy in humans, until recently [84]. Recent findings suggest that MSP1-specific humoral immunity undergoes a progressive maturation process from infants to adults characterized by an augmentation in both quantities but also functional potency as individuals age. This is particularly pronounced in the context of blood-stage antibody-dependent neutrophil functionality [85], which underscores specific antibody Fc-effector profiles correlating with the ability to control malaria.

A common, shared high-throughput immunomonitoring platform using well-defined reagents and read-outs might help in future to better understand the association between anti-MSP1 antibodies in serum samples tested and clinical malaria in different age groups residing in a range of malaria-endemic countries. The availability of ever increasing data on the parasite genotypes circulating in different regions of sub-Saharan Africa will help to improve and adapt these immunomonitoring platforms further by using reagents reflecting the parasite populations better than prime MSP1-specific immunity to be measured.

4. Experimental vaccine studies in non-human primates

Initial experiments using animal models in 1984 consisted in the immunization of *Saimiri sciureus* monkeys with the parasite-purified full-length MSP1 ($MSP1_{FL}$) protein followed by a subsequent intravenous challenge with infected red blood cells (iRBCs) [86]. All immunized monkeys self-resolved the infection following a peak parasitemia of ~4%. However, results were not fully conclusive as two out of four monkeys from the control group spontaneously cleared the infection as well (although their peak parasitemia was much higher at ~11%) [86]. Shortly after, a side-by-side comparison on the protective effect of an immunization with $MSP1_{FL}$ versus rhoptry proteins was performed [87]. Immunization with $MSP1_{FL}$ completely protected *Aotus lemurinus griseimembra* monkeys against *P. falciparum* malaria. Monkeys were immunized either with $MSP1_{FL}$ and subunits or with rhoptry proteins (n = 3) from the Uganda Palo Alto Knob + strain (FUP) versus controls and later challenged with a lethal infection of homologous *P. falciparum* FUP strain 3 weeks after last immunization. No patent parasitemia was detected in all monkeys immunized with MSP1, but only one out of three monkeys immunized with rhoptry proteins was partially protected [87].

Supported by immune-epidemiological data pointing at conserved regions as potentially protection-mediating portions of MSP1, a recombinant vaccine construct was created [88]. It contained two conserved regions of MSP1 involving the N-terminal amino acids 146–312 from MSP1-83, together with amino acids 1059–1196 from the MSP1-38 subunit, covering ~20% of the total length of MSP1. The vaccine was used to immunize *Aotus trivirgatus griseimembra* monkeys resulting in two out of five monkeys being protected against iRBC challenge using the highly virulent *P. falciparum* Vietnam Oak-Knoll (FVO) strain. Control animals immunized with recombinant *P. falciparum* aldolase were not protected. Yet, intriguingly, B-cell and T-cell responses were reported to be comparable in protected versus non-protected animals [88].

To further proof the capacity of conserved regions of MSP1 to induce protective immunity in the monkey model, Etlinger *et al.* [89] used the *Saimiri* non-human primate model. Monkeys were immunized either with native $MSP1_{FL}$ protein or with the so-called 'recombinant protein 1' of MSP1 (representing the N-terminal amino acids 147–321 from the MSP1-83 subunit) belonging to the K1 strain. The subsequent intravenous challenge was performed with iRBC from the Palo Alto strain expressing the MAD20 sequence (heterologous challenge). Whilst all control monkeys required drug treatment, half of the monkeys immunized with 'recombinant protein 1' were protected. Of interest, two out of three *Saimiri* animals immunized with native $MSP1_{FL}$ cleared the heterologous infection [89]. The third monkey was removed from the study 8 days after challenge and, although speculative, could likely be protected as well. All control monkeys required malaria drug therapy 12 days after challenge. These data suggested that the conserved regions from the N-terminal end of MSP1 may be important for heterologous protection in the *Saimiri* monkey model and that the best results in terms of protection were obtained by using MSP_{FL} protein for immunization.

The development of the synthetic peptide-based vaccine named SPf66 by Manuel Patarroyo involved immunizing wild-caught *Aotus trivirgatus* monkeys with merozoite-derived proteins, followed by challenging the monkeys with the asexual blood stages of the *P. falciparum* FVO strain [90,91]. The proteins that provided some levels of protection were partially sequenced, and synthetic peptides were created from these sequences to immunize additional monkeys [92]. The most promising peptides were merged, and the most potent combination was synthesized as a single hybrid polypeptide which was reported to incorporate amino acids from 35 kDa, 55 kDa, and 83 kDa proteins linked by an amino acid sequence reproducing one repeat region of the Circumsporozoite Protein (CSP). The amino acid sequence of the 83 kDa peptide

included in the SPf66 vaccine was demonstrated to be derived from the N-terminal end of MSP1 (MSP1-83). Among the group of six monkeys immunized with the three peptides comprising SPf66, three monkeys did not exhibit complete protection and experienced mild parasitemia reaching a maximum of 5%. Parasitemia peaked 10–15 days later compared to the control group, but the monkeys eventually underwent spontaneous recovery. Conversely, the remaining three monkeys in the same group demonstrated no signs of disease, and no parasites were detected in their blood smear samples up to 180 days of follow-up [90]. Later, SPf66 from two different sources of production (Colombia versus USA) was tested in *Aotus nancymai*, and results underlined that the 83 kDa peptide was one of the less immunogenic of the components of the vaccine [93]. Two separate research teams were later unable to replicate these findings [93,94], but the vaccine proceeded further with human efficacy trials in Latin America, sub-Saharan Africa, and Asia [95–105].

Following experiments in monkeys focused on protein subunits, especially on the C-terminal fragment of MSP1. This was based on immune-epidemiological studies pointing at MSP1-42 and MSP1-19 as immunodominant subunits associated with reduced clinical malaria incidence. Directing the induced immune responses to a specified, immunodominant region would enhance protective responses and might avoid immune evasion mechanisms of the parasite. Moreover, the expression of full-length *Plasmodium* proteins poses technical challenges due to their intricate structural conformations and substantial size. Thus, by shifting from $MSP1_{FL}$ to subunits of MSP1, the large-scale antigen production under good manufacturing practice became feasible. Immunization of *Aotus nancymai* and *Aotus vociferans* monkeys with either recombinant MSP1-42 or MSP1-19 led to distinct protection outcomes [106]. Monkeys immunized with the partially dimorphic MSP1-42 using Freund's adjuvant and later challenged with lethal FVO strain of *P. falciparum* required treatment similarly to the control monkeys. In contrast, monkeys immunized with the conserved MSP1-19 subunit self-resolved an otherwise lethal infection [106]. Unexpectedly, sera from the protected animals had no effect on *in vitro* invasion of *P. falciparum* into RBCs, questioning the dogma considering invasion-blocking antibodies as the main contributors to protection [106]. Yet, opposite results, this time in favor of the role of invasion blocking mechanism in protection, were reported a decade later when *Aotus nancymai* monkeys immunized with recombinant MSP1-42 were protected against the *P. falciparum* FVO strain. Protection strongly correlated with the anti-MSP1-42 antibody titer and with *in vitro* growth inhibition activity (GIA) [107,108]. These findings definitively promoted the combination of ELISA and GIA as key immune-monitoring assays to predict protective immunity induced by blood-stage vaccines in human clinical trials. More recently, a chimeric recombinant PfMSP1/8 vaccine fusing MSP1-19 together with the N-terminus of *P. falciparum* merozoite surface protein 8 (MSP8), lacking its low-complexity Asn/Asp-rich domain, was explored [109,110]. As found in previous mice experimentation [111,112], the vaccine elicited antibody responses to both, MSP8 and the conserved MSP1-19 domains in *Aotus nancymai*. The antibodies generated showed significant cross-reactivity between the FVO and 3D7 alleles of MSP1 and exhibited strong *in vitro* parasite growth inhibitory activity [109,110].

Immunization of *Rhesus* macaques with the polymorphic MSP1 block 2 construct (the MSP1 hybrid antigen) elicited antibodies against epitopes shared between all three block two serotypes, namely K1, MAD20 and RO33 types, but IgG was found to be non-inhibitory in the GIA [113]. Subsequent immunization of *Aotus lemurinus griseimembra* with the same construct, followed by a challenge with *P. falciparum* FVO strain, resulted in 50% of the immunized monkeys protected. Protected animals had four times higher parasite-specific serum antibody titers and different antibody epitope specificities compared to unprotected monkeys [114]. In summary, experimental immunization experiments in monkey animal models suggested that combining conserved domains of both the N- and the C-terminal regions of MSP1, or alternatively using the entire $MSP1_{FL}$ for immunization, promoted

Table 1. Subunit and viral vectored MSP1 vaccine constructs evaluated in clinical trials.

Vaccine candidate	Constructs	Expression system	Adjuvant	Ref.
SPf66	35 kDa, 55 kDa, and 83 kDa (MSP1) of *P. falciparum* blood stages antigens	Synthetic	Aluminum hydroxide	[62-90-95-97-105-107-109-112]
Combination-B	195 L-Block 3 of KI strain MSP1, full-length 3D7 MSP2 and 70% C-terminal of FCQ-27 RESA	*E. coli*	Montanide ISA720	[80–83]
MSP1-19	3D7 or FVO strain MSP1-19 fused to tetanus toxoid epitopes amino acids B-1631 to S-1723 with P30 and P2 epitopes of Th	*S. cerevisiae*	Aluminum hydroxide	[118]
MSP42	3D7 or FVO strain MSP1-42 External domain of MSP1-42	*E. coli*	Alhydrogel	[124]
FMP1	3D7 strain MSP1-42 designed to include T-cell epitopes	*E. coli*	AS02A oil-in-water emulsion, MPL, and immunostimulant QS21	[119–123]
ChAd63-MVA MSP1	MSP1 insert consisting of blocks 1, 3, 5, 12, Wellcome MSP1-33, Wellcome MSP1-19, Mad20 MSP1-33, Mad20 MSP1-19	Cloned into ChAd63 and MVA	None	[126,128]
MSP1-42-C1	MSP1-42 FVO and 3D7 strain mixture (50:50)	*E. coli*	Alhydrogel ±CPG7909 (TLR9)	[125]
BSAM2	3D7 and FVO strain alleles of MSP1-42 and AMA1 mixtures	*E. coli* and *P. pastoris*	Alhydrogel + CPG 7909	[63]
JAIVAC-1	PfF2 (binding domain of EBA175 of pfCAMP strain) MSP1-19 of FVO strain	Not mentioned	Montanide ISA	[130]
SumayaVac-1	3D7 strain full-length MSP1	*E. coli*	GLA-SE (TLR4)	[84,146]

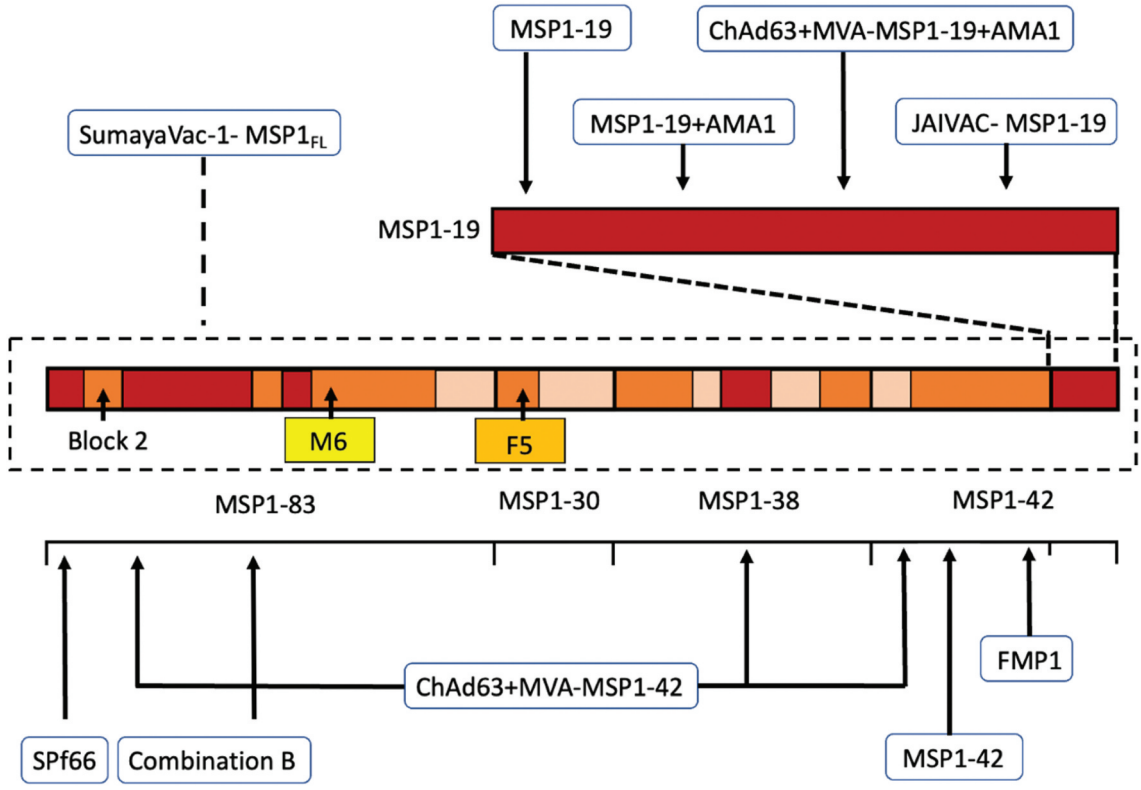


Figure 3. Overview of MSP1-based vaccine constructs evaluated in preclinical and early phase clinical development studies. Arrows point at the MSP1 subunit or epitopes within the subunit contained in the vaccine candidate. Highlighted: block 2, dimorphic fragments M6 (amino acids 671–833 from the K1 strain in yellow) and F5 (amino acids 384–595 from the MAD20 strain in orange).

higher protection levels against asexual blood-stage challenge. These findings essentially corroborated results obtained from immune-epidemiological studies mentioned above.

5. MSP1-based vaccination studies in human populations

MSP1 has undergone a series of clinical trials in different populations, age groups, dosing, and formulations (Table 1, Supp Table 2). An overview of the MSP1 fragments tested in these clinical trials is provided in Figure 3.

The SPf66 vaccine, which includes amino acid sequences from a conserved region of MSP1-83, was the first malaria vaccine to be evaluated in a phase III study. SPf66 was evaluated extensively for safety, immunogenicity, and efficacy in Colombia [95–97], Ecuador [98], Venezuela [99], Tanzania [100–102], The Gambia [103], Thailand [104], and Brazil [105]. Despite a reported estimated vaccine efficacy of 31% (with a 95% confidence interval of 0–52%) [101], overall conclusions of these trials suggested that SPf66 vaccine-induced immunity did not mediate malaria protection [115]. Of interest for this manuscript, knowledge of the detailed immune responses that were induced against the MSP1-83 vaccine component in SPf66-vaccinated individuals remained unclear. Amador et al. [95] reported that sera from vaccines with high antibody titers recognized the 83 kDa protein. However, immune responses against the different components of this synthetic construct have not been described in greater detail [78–88,116,117].

A phase I vaccine trial investigated the safety and immunogenicity of a vaccine candidate called Combination B [80]. This vaccine combined an N-terminal block 3 conserved fragment of MSP1 (190LCS.T3), the full-length MSP2 protein, and a part of the ring-infected erythrocyte surface antigen (RESA) formulated in Montanide ISA720 adjuvant and was tested first in Australian malaria-naïve volunteers. The results of the trial showed a strong T-cell response triggered by the MSP1 fragment and RESA, but generally, the antibody responses were weak. No protection against the *P. falciparum* 3D7 iRBC challenge 4 weeks after immunization was observed [82]. The same vaccine was later explored in a malaria pre-exposed population from Papua New Guinea with similar results [83]. The study suggested that MSP1 was not the component accounting for the observed 62% reduction in parasitemia triggered by Combination B vaccination in the phase I–IIb trial [83].

Another phase I trial evaluated an alum-adsorbed vaccine containing MSP1-19 subunit from either the *P. falciparum* 3D7 or FVO strain fused to tetanus toxoid T-helper epitopes P30 and P2 in healthy adults [118]. Both vaccines were immunogenic with serum antibody responses increasing with the dose administered. Yet, modifications to the formulation were suggested to be necessary to improve safety and immunogenicity profiles. The work of the *MSP1 Malaria Vaccine Working Group* focussed on the development of a MSP1 vaccine called 'falciparum malaria protein 1' (FMP1) derived from the MSP1-42 fragment (3D7 strain) using AS02A as adjuvant [119]. A phase I trial demonstrated the vaccine to be safe and immunogenic

when tested in malaria pre-exposed residents from western Kenya. The elicited antibodies showed GIA activity *in vitro*. Cellular immunity was induced as shown by T-cell proliferation and Enzyme-Linked Immunosorbent Spot (ELISPOT) responses [120]. Thus, this vaccine moved ahead for testing in children aged 12–47 months [121] showing comparable outcomes to adults and with a statistically significant effect of dosage level on immune responses. The FMP1 was further tested under a seasonal malaria setting by the *Mali FMPI Working Group*. FMP1 was also found to be highly immunogenic in Malian adults (n = 40) exposed to intense seasonal malaria transmission, with elicited immune responses binding to genetically diverse parasite clones [122]. The levels of anti-MSP1-42 antibodies exhibited a seasonal pattern that was significantly enhanced and prolonged by the vaccine. Yet, despite being one of the MSP1 candidates with a longer history in human malaria clinical vaccine testing, the vaccine did not progress further [123]. The reason was that in a phase IIb trial in Kenya including children that received three immunizations, and fully completing the follow-up period, the overall vaccine efficacy was 5.1% (95% CI: −26% to +28%), and thus, further development was halted [123] .

To further assess safety and immunogenicity of two additional MSP1-based vaccine candidates encompassing the C-terminal MSP1-42 subunit (FVO vs 3D7 strains) and formulated with Alhydrogel, a phase I clinical trial enrolling malaria-naïve volunteers from Kansas (USA) was conducted [124]. Both vaccines were regarded as safe and well tolerated, but they did not generate a significant functional humoral immune response *in vitro* when tested by GIA. To improve the vaccine effectiveness, the addition of other immunostimulants to the formulation was proposed. Ellis et al. [125] conducted a phase I study with a vaccine formulation of Alhydrogel complemented with the adjuvant CPG 7909. Following the third round of immunization, an *in vitro* GIA was performed. The level of inhibition observed was dependent on the antibody titer, with slightly higher inhibition (14%, range 3–32%) in the CPG 7909 group compared to the non-CPG 7909 group [125].

Other vaccine delivery approaches were developed over time to induce stronger cellular immune responses. A prime-boost approach was followed using the chimpanzee adenovirus 63 (ChAd63) and Modified Vaccinia Ankara (MVA) as delivery vectors including all conserved regions of MSP1 plus the MSP1-42 subunit from the 3D7 and WELLCOME strain [126]. Sixteen malaria-naïve volunteers were immunized, and a strong $CD4^+$ and $CD8^+$ T-cell response was induced [126]. MSP1-specific antibody responses were induced binding as well to native parasites. Antibodies could not inhibit *P. falciparum* growth *in vitro*, and it was speculated that antibody titers were not high enough to mediate growth inhibition, although other antibody functionalities were not tested. Possibly, with the intention of strengthening the ability to block RBC invasion by vaccine-elicited antibodies, a recombinant fusion of the domain III of *P. falciparum* Apical Membrane Antigen 1 (AMA1) together with MSP1-19 was produced, combined with adjuvant Montanide ISA 720 and tested in a phase Ia trial [127]. The vaccine called PfCP2.9/Montanide ISA 720 was evaluated in healthy Chinese malaria-naïve volunteers. The high antibody titers measured to the PfCP2.9 immunogen did not result in *in vitro* inhibition of parasite growth, and there was limited recognition of cultivated asexual blood-stage parasites in an immunofluorescence assay [127]. Later, the same antigenic combination MSP1-AMA1, although not including the conserved regions (as in [126]), and delivered by ChAd63 and MVA in a prime-boost approach, was tested in phase Ia trial involving malaria-naïve volunteers [128,129] and phase IIa trial involving naturally immune Kenyan adults [128,129]. As immunological read out ELISA antibody titers, GIA, IgG avidity, and isotypes were measured. The moderate *in vitro* GIA observed was mainly attributed to responses binding to AMA1 (>20% GIA activity). The phase IIa trial called VAC039 showed limited protection against a natural mosquito bite challenge [128]. Only one out of 38 volunteers showed protection, although some non-protected volunteers presented a longer parasite pre-patent period. Induction of cellular immunity against MSP1 and AMA1 did not appear to impact parasite growth rates *in vivo* [128].

Safety and protective efficacy of the vaccine candidate JAIVAC-1 were also assessed in healthy Indian males aged 18–45 years in a phase Ia trial [130]. JAIVAC-1 is composed of two recombinant proteins formulated with adjuvant Montanide ISA720, namely MSP1-19 and PfF2, representing the amino-terminal, conserved, cysteine-rich region of EBA-175 (with receptor-binding sites for glycophorin A in RBCs). All subjects seroconverted for PfF2, but the immune response to MSP1-19 was poor. A dose–response relationship was observed between the vaccine dose of PfF2 and antibody responses, but the vaccine failed to elicit significant antibody responses against MSP1-19[130].

The latest phase I clinical study evaluating MSP1 took place between 2017 and 2018 in malaria-naïve volunteers. For the first time in humans, the safety and immunogenicity of $MSP1_{FL}$ formulated together with the potent Toll-Like Receptor-4 (TLR4) agonist glucopyranosyl lipid A in an oil-in-water nano-emulsion (GLA-SE) as adjuvant, named SumayaVac-1 was tested [84]. The main difference of the $MSP1_{FL}$ protein included in SumayaVac-1 compared to the $MSP1_{FL}$ tested in monkeys is that while Saimiri or Aotus monkeys were immunized with $MSP1_{FL}$ isolated from parasites [86,87,89,131], SumayaVac-1 is based on $MSP1_{FL}$ produced recombinantly in *E. coli* by a codon optimization of the *msp-1* gene to reduce the AT content. SumayaVac-1 uses pZE expression plasmids [132] to separately express two halves of the MSP1 protein (p83/30 and p38/42) that are refolded together at a later step to reconstitute the full-length protein, structurally resembling the native MSP1 protein [133,134]. In this dose escalation, phase Ia clinical trial, the vaccine was safe and all malaria-naïve vaccinees sero-converted independent of the vaccine dose applied. High MSP1-specific antibody titers of IgG and IgM isotype peaked at 4 weeks after last vaccination, exceeding the antibody level seen in semi-immune populations from Kenya [84].

6. Humoral immunity against MSP1 and antibody effector function

Antibody-mediated effector functions against merozoites have proven to be highly correlating with protection against clinical

disease [135–141]. Fc receptor-dependent functionality is the result of interaction of antibody-opsonized merozoites with the complement system [136,137], as well as with a variety of circulating immune effector cells, namely, neutrophils [135,142,143], monocytes [138,139], and Natural Killer (NK) cells [141]. In a controlled human malaria infection study in Kenya, the breadth of Fc-mediated effector functions against merozoites, but not GIA, delineated grades of clinical manifestations and parasite replication rates [144]. Importantly, effector functions against $MSP1_{FL}$ recapitulated the results achieved when these effector mechanisms were tested against purified merozoites, indicating that $MSP1_{FL}$ is indeed a key target of naturally acquired opsonizing antibodies [145]. To the best of our knowledge, serum samples collected neither in MSP1 trials in monkeys nor in humans have tested for Fc-mediated functionality except recently [84,146]. In the SumayaVac-1 trial (EudraCT 2016–002463-33), antibodies from malaria-naïve vaccinees promoted activation of the classical complement pathway, opsonic phagocytosis by both neutrophils and monocytes, as well as the production of IFNγ and degranulation of NK cells. These multi-functional antibodies were sustained above baseline level months after immunization [146]. Interestingly, antibodies against both, the C-terminal MSP1-42 and the N-terminal MSP1-83, displayed the highest levels of opsonization-mediated functions. The use of $MSP1_{FL}$ could aid in the design of future MSP1-based vaccine constructs through systematic identification of epitopes that are targets of functional, protective antibodies when assessed in Controlled Human Malaria Infection (CHMI) studies in malaria pre-exposed volunteers. However, $MSP1_{FL}$ could divert the immune response leading to insufficient achievement of functional antibody titers, a possibility that will be explored in upcoming studies. MSP1-83 harbors the highly conserved MSP1 blocks 1, 3, and 5, of great importance for strain-transcending immunity [147]. The two prototypic MSP1 sequences, namely MSP1-D and MSP1-F, are recognized by antibodies produced after immunization with $MSP1_{FL}$ and antibody-mediated effector functions showed cross-strain variability [84]. This is of paramount interest for trials in African populations exposed to a variety of parasite strains [84,148].

7. Conclusion

For MSP1, as for other malaria blood-stage vaccine candidates, the key mechanism mediating immune protection has been attributed to elucidation of high titers of antibodies capable of inhibiting in a Fab-mediated manner merozoite invasion into RBCs [67,149]. Thus, the first *in vitro* surrogate of blood-stage vaccine efficacy has been testing of serum antibodies for mediation of GIA. This is the case for the most advanced blood-stage vaccine candidate PfRH5 [150] that leads to induction of high levels of parasite neutralizing serum antibodies in the GIA [151–153], although no sterile protection was found in a recently conducted phase I trial followed by CHMI [154].

Immune-epidemiological data regarding MSP1 immunity suggest that both the N- and C-terminal ends of MSP1 play an important role in protection. The technical possibility of Good Manufacturing Practices (GMP) production of $MSP1_{FL}$ preserving the natural protein conformation of both the N- and C-terminal end will be essential for generating protective humoral immune responses based on growth inhibitory antibodies or Fc-mediated effector functions [146]. The MSP1 structure of the Sumaya-Vac1 has recently been re-solved [133], highlighting that MSP1 forms dimers in a concentration-dependent manner and that dimerization is affected by the presence of the erythrocyte cytoskeleton protein spectrin [133].

Conserved regions of MSP1-19 and MSP1-83 were included in an MSP_{FL} vaccine construct, which was proven to be highly protective in monkey trials. Recently, antibodies elicited by MSP_{FL} vaccination in malaria-naïve humans were shown to promote a wide range of antibody Fc-dependent effector function *in vitro*. The availability of a well-characterized CHMI model in adult African populations will allow as next step to dissect the correlation between the breath of antibody effector functions induced by $MSP1_{FL}$ vaccination, their interaction with preexisting MSP1-specific immunity, and impact on asexual blood-stage parasite growth rates and pre-patent period under highly controlled conditions.

Testing of $MSP1_{FL}$ in humans has also the potential to significantly improve the development of future potentially preventive or even therapeutic anti-MSP1 monoclonal antibodies (mAbs) isolated from vaccinated and CHMI protected volunteers similar to already established antibodies against CSP [155,156]. Some well-described mAbs against MSP1 are known to react with epitopes on the MSP1-42 and MSP1-19 C-terminal processing fragments and to recognize the first growth factor-like domain of MSP1 [157,158]. Yet, not all mAbs have been characterized as inhibitory and some of them deploy an interfering or blocking activity [159,160]. Of interest, some mAbs able to inhibit *P. falciparum* growth *in vitro* reacted with epitopes on the MSP1-83 N-terminal processing fragment [161] included in $MSP1_{FL}$ vaccines [84].

8. Expert opinion

Our main existing deployed tools for malaria control, including drugs and insecticides, lose their activity to increasingly resistant parasites and mosquitoes and raise the challenge of further improving other tools to avoid an increase in morbidity and mortality after decades of successful efforts. The CSP-based pre-erythrocytic RTS,S/AS01 malaria vaccine or its R21 biosimilar has shown that vaccines targeting the major surface protein of *P. falciparum* sporozoites do significantly reduce clinical malaria especially in children living in sub-Saharan Africa. Building also on evidence obtained from naturally acquired protective humoral immunity, further improvement of the protection level of malaria vaccines might depend on the development of a multistage, multicomponent vaccine targeting both, sporozoites and asexual blood-stage parasites, as these latest stages are the responsible for the clinical manifestations, and tackling the blood stage contributes to the reduction of the reservoir of transmissible sexual stages.

Until a multi-stage vaccine against malaria is generated, the addition of an anti-merozoite blood-stage vaccine based on the recombinantly expressed $MSP1_{FL}$ to vaccination strategies already in place using pre-erythrocytic vaccines is a valuable way to go. $MSP1_{FL}$ delivered with safe adjuvants like GLA-SE has proven to be safe and highly immunogenic in a recently conducted first in human phase 1a clinical trial in naïve volunteers. The capacity of $MSP1_{FL}$ to elicit functional and strain-transcending antibodies is worth exploring in future clinical studies in combination with pre-erythrocytic vaccines such as RTS,S/AS01 or R21.

Antibodies play a central role in protection against malaria infection and clinical malaria. The dissection of antibody Fc-mediated effector functions has shed more light on the range of effector functions mediated by parasite-binding antibodies and their interaction with immune players such as neutrophils, monocytes, NK-cells, and the complement system. These Fc-mediated effector mechanisms are dependent on opsonization in order to trigger inhibition of parasite invasion, antibody-dependent cellular cytotoxicity, respiratory burst of neutrophils, complement fixation, deposition, and membrane attack complex formation, as well as antibody-mediated phagocytosis by neutrophils and monocytes.

Malaria pre-exposure levels impact on the quality, quantity, and duration of malaria vaccine-induced immunity and protection. The availability of clinical research facilities able to perform CHMI studies in malaria pre-exposed populations based on the application of defined dosages of live, non-attenuated purified, cryopreserved *P. falciparum* sporozoites or iRBC has opened the opportunity to evaluate under highly controlled conditions the protective potential of novel vaccine combinations in the relevant target population. The deployment of system serology studies as immune-monitoring tool in serum samples collected from these clinical trial participants with known protection status against CHMI will help to unravel the functionally most relevant antibody-binding sites on their target antigens and the effector functions mediated. Dissection of vaccine-induced antibody fine specificity and their association to preexisting antibody levels and fine specificity will shed much needed insight on the impact of levels of malaria pre-exposure on vaccine-induced immunity and the resulting immune protection in the different age groups. This knowledge will help to understand how to improve already existing malaria vaccines for different malaria pre-exposed populations.

Acknowledgments

The authors extend their heartfelt gratitude to the dedicated researchers, institutions, and study volunteers participating in malaria vaccine studies who have relentlessly pursued the formidable objective of jointly creating a malaria vaccine candidate centered around the merozoite surface antigen 1. The collective endeavors of the malaria research community have profoundly driven scientific advancements presented in this manuscript.

Funding

Richard Thomson Luque is employed by Sumaya-Biotech GmbH & Co. KG that received support by the EU Malaria Fund. This work was supported in part by the U.S. National Institutes of Health (NIH) through awards R01 AI141900 and U19 AI110820 to JCS.

Declaration of interests

R Thomson Luque is employee of Sumaya-Biotech GmbH & Co. KG. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions

All authors have contributed to the conception and design of the review article and interpreting the relevant literature and have been involved in writing the review article or revised it for intellectual content.

Dedication

This review is dedicated to Prof. Herman Bujard, an inspiring scientist who, as a force of nature, drew everyone who interacted with him to join his passion to develop an MSP1-based malaria vaccine to protect vulnerable people living in malaria-endemic countries.

Data availability statement

The *Plasmodium falciparum* whole-genome assemblies used for this work are available from the NCBI (https://www.ncbi.nlm.nih.gov/gene/).

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of other entities.

ORCID

Richard Thomson-Luque http://orcid.org/0000-0001-8378-9497

References

1. World Health Organization. Malaria vaccine: WHO position paper - January 2016. Weekly Epidemiol Rec. 2016;91:33–52.
2. Organization, W.H. World malaria report 2021. (2021).
3. Hemingway J, Ranson H, Magill A, et al. Averting a malaria disaster: will insecticide resistance derail malaria control? Lancet. 2016;387:1785–1788.
4. Organization, W.H. World malaria report 2019. (2019).
5. Achan J, Serwanga A, Wanzira H, et al. Current malaria infection, previous malaria exposure, and clinical profiles and outcomes of COVID-19 in a setting of high malaria transmission: an exploratory cohort study in Uganda. Lancet Microbe. 2022;3:e62–e71.
6. Cervellati M, Esposito E, Sunde U, et al. Malaria risk and civil violence. Munich, Germany: CESifo; 2017.
7. Penny MA, Camponovo F, Chitnis N, et al. Future use-cases of vaccines in malaria control and elimination. Parasite Epidemiol Control. 2020;10:e00145.
8. Nussenzweig RS, Vanderberg J, Most H, et al. Protective immunity produced by the injection of x-irradiated sporozoites of plasmodium berghei. Nature. 1967;216:160–162.

9. Keating C. The history of the RTS,S/AS01 malaria vaccine trial. Lancet. 2020;395:1336–1337.
10. Crompton PD, Moebius J, Portugal S, et al. Malaria immunity in man and mosquito: insights into unsolved mysteries of a deadly infectious disease. Annu Rev Immunol. 2014;32:157–187.
11. Beeson JG, Kurtovic L, Valim C, et al. The RTS,S malaria vaccine: current impact and foundation for the future. Sci Transl Med. 2022;14:eabo6646.
12. Adepoju P. RTS,S malaria vaccine pilots in three African countries. Lancet. 2019;393:1685.
13. World Health Organization. World malaria report 2018. Geneva: World Health Organization; 2018.
14. Datoo MS, Natama MH, Somé A, et al. Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. Lancet. 2021;397:1809–1818.
15. Datoo MS, Natama HM, Somé A, et al. Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial. Lancet Infect Dis. 2022;22. doi: 10.1016/S1473-3099(22)00442-X
16. Epstein JE, Paolino KM, Richie TL, et al. Protection against *Plasmodium falciparum* malaria by PfSPZ vaccine. JCI Insight. 2017;2:e89154.
17. Richie TL, Church LWP, Murshedkar T, et al. Sporozoite immunization: innovative translational science to support the fight against malaria. Expert Rev Vaccines. 2023;22:964–1007.
18. Thera MA, Doumbo OK, Coulibaly D, et al. A field trial to assess a blood-stage malaria vaccine. N Engl J Med. 2011;365:1004–1013.
19. Genton B, Reed ZH. Asexual blood-stage malaria vaccine development: facing the challenges. Curr Opin Infect Dis. 2007;20:467–475.
20. Duffy PE, Patrick Gorres J. Malaria vaccines since 2000: progress, priorities, products. NPJ Vaccines. 2020;5:48.
21. Cai J, Chen S, Zhu F, et al. Whole-killed blood-stage vaccine: is it worthwhile to further develop it to control malaria? Front Microbiol. 2021;12:670775.
22. Dunachie SJ, Berthoud T, Keating SM, et al. MIG and the regulatory cytokines IL-10 and TGF-β1 correlate with malaria vaccine immunogenicity and efficacy. PLoS One. 2010;5:e12557.
23. Holder AA, Freeman RR. Immunization against blood-stage rodent malaria using purified parasite antigens. Nature. 1981;294:361–364.
24. Holder AA, Lockyer MJ, Odink KG, et al. Primary structure of the precursor to the three major surface antigens of *Plasmodium falciparum* merozoites. Nature. 1985;317:270–273.
25. Holder AA. The precursor to major merozoite surface antigens: structure and role in immunity. Prog Allergy. 1988;41:72–97.
26. Holder AA, Blackman MJ, Burghaus PA, et al. A malaria merozoite surface protein (MSP1)-structure, processing and function. Mem Inst Oswaldo Cruz. 1992;87 Suppl 3:37–42.
27. Koussis K, Withers-Martinez C, Yeoh S, et al. A multifunctional serine protease primes the malaria parasite for red blood cell invasion. EMBO J. 2009;28:725–735.
28. Harris PK, Yeoh S, Dluzewski AR, et al. Molecular identification of a malaria merozoite surface sheddase. PLoS Pathog. 2005;1:241–251.
29. Blackman MJ, Holder AA. Secondary processing of the *Plasmodium falciparum* merozoite surface protein-1 (MSP1) by a calcium-dependent membrane-bound serine protease: shedding of MSP133 as a noncovalently associated complex with other fragments of the MSP1. Mol Biochem Parasitol. 1992;50:307–315.
30. Child MA, Epp C, Bujard H, et al. Regulated maturation of malaria merozoite surface protein-1 is essential for parasite growth. Mol Microbiol. 2010;78:187–202.
31. Dluzewski AR, Ling IT, Hopkins JM, et al. Formation of the food vacuole in *Plasmodium falciparum*: a potential role for the 19 kDa fragment of merozoite surface protein 1 (MSP1(19)). PLoS One. 2008;3:e3085.
32. Trucco C, Fernandez-Reyes D, Howell S, et al. The merozoite surface protein 6 gene codes for a 36 kDa protein associated with the *Plasmodium falciparum* merozoite surface protein-1 complex. Mol Biochem Parasitol. 2001;112:91–101.
33. Stafford WHL, Günder B, Harris A, et al. A 22 kDa protein associated with the *Plasmodium falciparum* merozoite surface protein-1 complex. Mol Biochem Parasitol. 1996;80:159–169.
34. Pachebat JA, Ling IT, Grainger M, et al. The 22 kDa component of the protein complex on the surface of *Plasmodium falciparum* merozoites is derived from a larger precursor, merozoite surface protein 7. Mol Biochem Parasitol. 2001;117:83–89.
35. Chiu CY, Hodder AN, Lin CS, et al. Antibodies to the *Plasmodium falciparum* proteins MSPDBL1 and MSPDBL2 opsonize merozoites, inhibit parasite growth, and predict protection from clinical malaria. J Infect Dis. 2015;212:406–415.
36. Babiker HA, Creasey AM, Fenton B, et al. Genetic diversity of *Plasmodium falciparum* in a village in eastern Sudan. 1. Diversity of enzymes, 2D-PAGE proteins and antigens. Trans R Soc Trop Med Hyg. 1991;85:572–577.
37. Tanabe K, Mackay M, Goman M, et al. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J Mol Biol. 1987;195:273–287.
38. Takala S, Branch O, Escalante AA, et al. Evidence for intragenic recombination in *Plasmodium falciparum*: identification of a novel allele family in block 2 of merozoite surface protein-1: Asembo Bay Area Cohort Project XIV. Mol Biochem Parasitol. 2002;125:163–171.
39. Tanabe K, Sakihama N, Walliker D, et al. Allelic dimorphism-associated restriction of recombination in *Plasmodium falciparum* msp1. Gene. 2007;397:153–160.
40. Kiwanuka GN. Genetic diversity in *Plasmodium falciparum* merozoite surface protein 1 and 2 coding genes and its implications in malaria epidemiology: a review of published studies from 1997-2007. J Vector Borne Dis. 2009;46:1–12.
41. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 1987;4:203–221.
42. Ferreira MU, Ribeiro WL, Tonon AP, et al. Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*. Gene. 2003;304:65–75.
43. Gentz R, Certa U, Takacs B, et al. Major surface antigen p190 of *Plasmodium falciparum*: detection of common epitopes present in a variety of plasmodia isolates. Embo J. 1988;7:225–230.
44. Jongwutiwes S, Tanabe K, Nakazawa S, et al. Sequence variation in the tripeptide repeats and T cell epitopes in P190 (MSA-1) of *Plasmodium falciparum* from field isolates. Mol Biochem Parasitol. 1992;51:81–89.
45. Agonhossou R, Akoton R, Lagnika H, et al. P. falciparum msp1 and msp2 genetic diversity in P. falciparum single and mixed infection with P. malariae among the asymptomatic population in Southern Benin. Parasitol Int. 2022;89:102590.
46. Mwingira F, Nkwengulila G, Schoepflin S, et al. *Plasmodium falciparum* msp1, msp2 and glurp allele frequency and diversity in sub-Saharan Africa. Malar J. 2011;10:79.
47. Chen J-T, Li J, Zha G-C, et al. Genetic diversity and allele frequencies of *Plasmodium falciparum* msp1 and msp2 in parasite isolates from Bioko Island, Equatorial Guinea. Malar J. 2018;17:458.
48. Apinjoh TO, Tata RB, Anchang-Kimbi JK, et al. *Plasmodium falciparum* merozoite surface protein 1 block 2 gene polymorphism in field isolates along the slope of mount Cameroon: a cross – sectional study. BMC Infect Dis. 2015;15:309.
49. Somé AF, Bazié T, Zongo I, et al. *Plasmodium falciparum* msp1 and msp2 genetic diversity and allele frequencies in parasites isolated from symptomatic malaria patients in Bobo-Dioulasso, Burkina Faso. Parasit Vectors. 2018;11:323.
50. Miller LH, Roberts T, Shahabuddin M, et al. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol Biochem Parasitol. 1993;59:1–14.
51. Snounou G, Beck HP. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. Parasitol Today. 1998;14:462–467.
52. Cattamanchi A, Kyabayinze D, Hubbard A, et al. Distinguishing recrudescence from reinfection in a longitudinal antimalarial drug efficacy study: comparison of results based on genotyping of msp-1, msp-2, and glurp. Am J Trop Med Hyg. 2003;68:133–139.

53. Mohammed H, Mindaye T, Belayneh M, et al. Genetic diversity of *Plasmodium falciparum* isolates based on MSP-1 and MSP-2 genes from Kolla-Shele area, Arbaminch Zuria District, southwest Ethiopia. Malar J. 2015;14:73.
54. Otto TD, Böhme U, Sanders MJ, et al. Long read assemblies of geographically dispersed *Plasmodium falciparum* isolates reveal highly structured subtelomeres. Wellcome Open Res. 2018;3(52). doi: 10.12688/wellcomeopenres.14571.1
55. Moser KA, Dwivedi A, Stucke EM, et al. New *Plasmodium falciparum* genome assemblies from diverse endemic regions enables the comprehensive genomic and genetic characterization of clinical isolates. Am J Trop Med Hyg. 2017;97:511–511. AMER SOC TROP MED & HYGIENE 8000 WESTPARK DR, STE 130, MCLEAN, VA 22101 USA
56. Tanabe K, Zollner G, Vaughan JA, et al. *Plasmodium falciparum*: genetic diversity and complexity of infections in an isolated village in western thailand. Parasitol Int. 2015;64:260–266.
57. Tanabe K, Mita T, Palacpac NMQ, et al. Within-population genetic diversity of *Plasmodium falciparum* vaccine candidate antigens reveals geographic distance from a Central sub-Saharan African origin. Vaccine. 2013;31:1334–1339.
58. White WL. Erratum to: why I hate the index finger. Hand (N Y). 2011;6:233.
59. Tanabe K, Sakihama N, Kaneko A. Stable SNPs in malaria antigen genes in isolated populations. Science. 2004;303:493.
60. Takala SL, Escalante AA, Branch OH, et al. Genetic diversity in the Block 2 region of the merozoite surface protein 1 (MSP-1) of *Plasmodium falciparum*: additional complexity and selection and convergence in fragment size polymorphism. Infect Genet Evol. 2006;6:417–424.
61. Rich SM, Ayala FJ. Population structure and recent evolution of *Plasmodium falciparum*. Proc Natl Acad Sci U S A. 2000;97:6994–7001.
62. Aponte JJ, Menendez C, Schellenberg D, et al. Age interactions in the development of naturally acquired immunity to *Plasmodium falciparum* and its clinical presentation. PLoS Med. 2007;4:e242.
63. Ellis RD, Wu Y, Martin LB, et al. Phase 1 study in malaria naïve adults of BSAM2/Alhydrogel®+CPG 7909, a blood stage vaccine against P. falciparum malaria. PLoS One. 2012;7:e46094.
64. Cohen S, Mc GI, Carrington S. Gamma-globulin and acquired immunity to human malaria. Nature. 1961;192:733–737.
65. Hviid L, Lopez-Perez M, Larsen MD, et al. No sweet deal: the antibody-mediated immune response to malaria. Trends Parasitol. 2022;38:428–434.
66. al-Yaman F, Genton B, Kramer KJ, et al. Assessment of the role of naturally acquired antibody levels to *Plasmodium falciparum* merozoite surface protein-1 in protecting Papua New Guinean children from malaria morbidity. Am J Trop Med Hyg. 1996;54:443–448.
67. Egan AF, Morris J, Barnish G, et al. Clinical immunity to *Plasmodium falciparum* malaria is associated with serum antibodies to the 19-kDa C-terminal fragment of the merozoite surface antigen, PfMSP-1. J Infect Dis. 1996;173:765–769.
68. Dodoo D, Theander TG, Kurtzhals JAL, et al. Levels of antibody to conserved parts of *Plasmodium falciparum* merozoite surface protein 1 in Ghanaian children are not associated with protection from clinical malaria. Infect Immun. 1999;67:2131–2137.
69. Nebie I, Diarra A, Ouedraogo A, et al. Humoral responses to *Plasmodium falciparum* blood-stage antigens and association with incidence of clinical malaria in children living in an area of seasonal malaria transmission in Burkina Faso, West Africa. Infect Immun. 2008;76:759–766.
70. Osier FH, Fegan G, Polley SD, et al. Breadth and magnitude of antibody responses to multiple *Plasmodium falciparum* merozoite antigens are associated with protection from clinical malaria. Infect Immun. 2008;76:2240–2248.
71. Bowman NM, Moormann AM, Bowman NM, et al. Longevity of genotype-specific immune responses to *Plasmodium falciparum* merozoite surface protein 1 in Kenyan children from regions of different malaria transmission intensity. Am J Trop Med Hyg. 2016;95:580–587.
72. Früh K, Doumbo O, Müller HM, et al. Human antibody response to the major merozoite surface antigen of *Plasmodium falciparum* is strain specific and short-lived. Infect Immun. 1991;59:1319–1324.
73. Muller HM, Früh K, von Brunn A, et al. Development of the human immune response against the major surface protein (gp190) of *Plasmodium falciparum*. Infect Immun. 1989;57:3765–3769.
74. Tolle R, Früh K, Doumbo O, et al. A prospective study of the association between the human humoral immune response to *Plasmodium falciparum* blood stage antigen gp190 and control of malarial infections. Infect Immun. 1993;61:40–47.
75. Fowkes FJ, Richards JS, Simpson JA, et al. The relationship between anti-merozoite antibodies and incidence of *Plasmodium falciparum* malaria: a systematic review and meta-analysis. PLoS Med. 2010;7:e1000218.
76. Conway DJ, Cavanagh DR, Tanabe K, et al. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nat Med. 2000;6:689–692.
77. Noranate N, Prugnolle F, Jouin H, et al. Population diversity and antibody selective pressure to *Plasmodium falciparum* MSP1 block2 locus in an African malaria-endemic setting. BMC Microbiol. 2009;9:219.
78. Cavanagh DR, Dodoo D, Hviid L, et al. Antibodies to the N-terminal block 2 of *Plasmodium falciparum* merozoite surface protein 1 are associated with protection against clinical malaria. Infect Immun. 2004;72:6492–6502.
79. Jaschke A, Coulibaly B, Remarque EJ, et al. Merozoite surface protein 1 from *Plasmodium falciparum* is a major target of opsonizing antibodies in individuals with acquired immunity against malaria. Clin Vaccine Immunol. 2017;24. DOI:10.1128/CVI.00155-17
80. Saul A, Lawrence G, Smillie A, et al. Human phase I vaccine trials of 3 recombinant asexual stage malaria antigens with Montanide ISA720 adjuvant. Vaccine. 1999;17:3145–3159.
81. Genton B, Betuela I, Felger I, et al. A recombinant blood-stage malaria vaccine reduces *Plasmodium falciparum* density and exerts selective pressure on parasite populations in a phase 1–2b trial in Papua New Guinea. J Infect Dis. 2002;185:820–827.
82. Lawrence G, Cheng Q, Reed C, et al. Effect of vaccination with 3 recombinant asexual-stage malaria antigens on initial growth rates of *Plasmodium falciparum* in non-immune volunteers. Vaccine. 2000;18:1925–1931.
83. Genton B, Al-Yaman F, Anders R, et al. Safety and immunogenicity of a three-component blood-stage malaria vaccine in adults living in an endemic area of Papua New Guinea. Vaccine. 2000;18:2504–2511.
84. Blank A, Fürle K, Jäschke A, et al. Immunization with full-length *Plasmodium falciparum* merozoite surface protein 1 is safe and elicits functional cytophilic antibodies in a randomized first-in-human trial. NPJ Vaccines. 2020;5:10.
85. Nziza N, Tran TM, DeRiso EA, et al. Accumulation of neutrophil phagocytic antibody features tracks with naturally acquired immunity against malaria in children. J Infect Dis. 2023;228:759–768.
86. Perrin LH, Merkli B, Loche M, et al. Antimalarial immunity in Saimiri monkeys. Immunization with surface components of asexual blood stages. J Exp Med. 1984;160:441–451.
87. Siddiqui WA, Tam LQ, Kramer KJ, et al. Merozoite surface coat precursor protein completely protects Aotus monkeys against *Plasmodium falciparum* malaria. Proc Natl Acad Sci U S A. 1987;84:3014–3018.
88. Herrera S, Herrera MA, Perlaza BL, et al. Immunization of Aotus monkeys with *Plasmodium falciparum* blood-stage recombinant proteins. Proc Natl Acad Sci U S A. 1990;87:4017–4021.
89. Etlinger HM, Caspers P, Matile H, et al. Ability of recombinant or native proteins to protect monkeys against heterologous challenge with *Plasmodium falciparum*. Infect Immun. 1991;59:3498–3503.
90. Patarroyo ME, Romero P, Torres ML, et al. Induction of protective immunity against experimental infection with malaria using synthetic peptides. Nature. 1987;328:629–632.
91. Rodriguez R, Moreno A, Guzman F, et al. Studies in owl monkeys leading to the development of a synthetic vaccine against the asexual blood stages of *Plasmodium falciparum*. Am J Trop Med Hyg. 1990;43:339–354.

92. Moreno A, Patarroyo ME. Development of an asexual blood stage malaria vaccine. Blood. 1989;74:537–546.
93. Ruebush TKN, Campbell GH, Moreno A, et al. Immunization of owl monkeys with a combination of *Plasmodium falciparum* asexual blood-stage synthetic peptide antigens. Am J Trop Med Hyg. 1990;43:355–366.
94. Herrera S, Guerrero R, Clavijo C, et al. Failure of a synthetic vaccine to protect Aotus lemurinus against asexual blood stages of *Plasmodium falciparum*. Am J Trop Med Hyg. 1992;47:682–690.
95. Amador R, Moreno A, Murillo LA, et al. Safety and immunogenicity of the synthetic malaria vaccine SPf66 in a large field trial. J Infect Dis. 1992;166:139–144.
96. Amador R, Moreno A, Valero V, et al. The first field trials of the chemically synthesized malaria vaccine SPf66: safety, immunogenicity and protectivity. Vaccine. 1992;10:179–184.
97. Valero MV, Amador LR, Galindo C, et al. Vaccination with SPf66, a chemically synthesised vaccine, against *Plasmodium falciparum* malaria in Colombia. Lancet. 1993;341:705–710.
98. Sempértegui F, Estrella B, Moscoso J, et al. Safety, immunogenicity and protective effect of the SPf66 malaria synthetic vaccine against *Plasmodium falciparum* infection in a randomized double-blind placebo-controlled field trial in an endemic area of Ecuador. Vaccine. 1994;12:337–342.
99. Noya O, Berti YG, Noya BAD, et al. A population-based clinical trial with the SPf66 synthetic *Plasmodium falciparum* malaria vaccine in Venezuela. J Infect Dis. 1994;170:396–402.
100. Teuscher T, Armstrong Schellenberg JRM, Bastos de Azevedo I, et al. SPf66, a chemically synthesized subunit malaria vaccine, is safe and immunogenic in Tanzanians exposed to intense malaria transmission. Vaccine. 1994;12:328–336.
101. Alonso P, Smith T, Armstrong Schellenberg JRM, et al. Randomised trial of efficacy of SPf66 vaccine against *Plasmodium falciparum* malaria in children in southern Tanzania. Lancet. 1994;344:1175–1181.
102. Acosta C, Galindo CM, Schellenberg D, et al. Evaluation of the SPf66 vaccine for malaria control when delivered through the EPI scheme in Tanzania. Trop Med Int Health. 1999;4:368–376.
103. D'alessandro U, Leach A, Olaleye BO, et al. Efficacy trial of malaria vaccine SPf66 in Gambian infants. Lancet. 1995;346:462–467.
104. Nosten F, Luxemburger C, Kyle DE, et al. Randomised double-blind placebo-controlled trial of SPf66 malaria vaccine in children in northwestern Thailand. Lancet. 1996;348:701–707.
105. Urdaneta M, Struchiner CJ, Boulos M, et al. Evaluation of SPf66 malaria vaccine efficacy in Brazil. The American Journal of Tropical Medicine and Hygiene. 1998;58:378–385.
106. Kumar S, Yadava A, Keister DB, et al. Immunogenicity and in vivo efficacy of recombinant *Plasmodium falciparum* merozoite surface protein-1 in Aotus monkeys. Mol Med. 1995;1:325–332.
107. Darko CA, Angov E, Collins WE, et al. The clinical-grade 42-kilodalton fragment of merozoite surface protein 1 of *Plasmodium falciparum* strain FVO expressed in *Escherichia coli* protects Aotus nancymai against challenge with homologous erythrocytic-stage parasites. Infect Immun. 2005;73:287–297.
108. Singh S, Miura K, Zhou H, et al. Immunity to recombinant *Plasmodium falciparum* merozoite surface protein 1 (MSP1): protection in Aotus nancymai monkeys strongly correlates with anti-MSP1 antibody titer and in vitro parasite-inhibitory activity. Infect Immun. 2006;74:4573–4580.
109. Alaro JR, Partridge A, Miura K, et al. A chimeric *Plasmodium falciparum* merozoite surface protein vaccine induces high titers of parasite growth inhibitory antibodies. Infect Immun. 2013;81:3843–3854.
110. Burns JM Jr, Miura K, Sullivan J, et al. Immunogenicity of a chimeric *Plasmodium falciparum* merozoite surface protein vaccine in Aotus monkeys. Malar J. 2016;15:159.
111. Shi Q, Lynch MM, Romero M, et al. Enhanced protection against malaria by a chimeric merozoite surface protein vaccine. Infect Immun. 2007;75:1349–1358.
112. Alaro JR, Lynch MM, Burns JM Jr. Protective immune responses elicited by immunization with a chimeric blood-stage malaria vaccine persist but are not boosted by Plasmodium yoelii challenge infection. Vaccine. 2010;28:6876–6884.
113. Cowan GJ, Creasey AM, Dhansarnsombut K, et al. A malaria vaccine based on the polymorphic block 2 region of MSP-1 that elicits a broad serotype-spanning immune response. PLoS One. 2011;6:e26616.
114. Cavanagh DR, Kocken CHM, White JH, et al. Antibody responses to a novel *Plasmodium falciparum* merozoite surface protein vaccine correlate with protection against experimental malaria infection in Aotus monkeys. PLoS One. 2014;9:e83704.
115. Graves P, Gelband H. Vaccines for preventing malaria (SPf66). Cochrane Database Syst Rev. 2006;2006:Cd005966.
116. Daubenberger CA, Nickel B, Ciatto C, et al. Amino acid dimorphism and parasite immune evasion: cellular immune responses to a promiscuous epitope of *Plasmodium falciparum* merozoite surface protein 1 displaying dimorphic amino acid polymorphism are highly constrained. Eur J Immunol. 2002;32:3667–3677.
117. Pöltl-Frank F, Zurbriggen R, Helg A, et al. Use of reconstituted influenza virus virosomes as an immunopotentiating delivery system for a peptide-based vaccine. Clin Exp Immunol. 1999;117:496–503.
118. Keitel WA, Kester KE, Atmar RL, et al. Phase I trial of two recombinant vaccines containing the 19kd carboxy terminal fragment of *Plasmodium falciparum* merozoite surface protein 1 (msp-1(19)) and T helper epitopes of tetanus toxoid. Vaccine. 1999;18:531–539.
119. Stoute JA, Gombe J, Withers MR, et al. Phase 1 randomized double-blind safety and immunogenicity trial of *Plasmodium falciparum* malaria merozoite surface protein FMP1 vaccine, adjuvanted with AS02A, in adults in western Kenya. Vaccine. 2007;25:176–184.
120. Ockenhouse CF, Angov E, Kester KE, et al. Phase I safety and immunogenicity trial of FMP1/AS02A, a *Plasmodium falciparum* MSP-1 asexual blood stage vaccine. Vaccine. 2006;24:3009–3017.
121. Withers MR, McKinney D, Ogutu BR, et al. Safety and reactogenicity of an MSP-1 malaria vaccine candidate: a randomized phase Ib dose-escalation trial in Kenyan children. PLoS Clin Trials. 2006;1:e32.
122. Thera MA, Doumbo OK, Coulibaly D, et al. Safety and allele-specific immunogenicity of a malaria vaccine in Malian adults: results of a phase I randomized trial. PLoS Clin Trials. 2006;1:e34.
123. Ogutu BR, Apollo OJ, McKinney D, et al. Blood stage malaria vaccine eliciting high antigen-specific antibody concentrations confers no protection to young children in Western Kenya. PLoS One. 2009;4:e4708.
124. Malkin E, Long CA, Stowers AW, et al. Phase 1 study of two merozoite surface protein 1 (MSP1(42)) vaccines for *Plasmodium falciparum* malaria. PLoS Clin Trials. 2007;2:e12.
125. Ellis RD, Martin LB, Shaffer D, et al. Phase 1 trial of the *Plasmodium falciparum* blood stage vaccine MSP1(42)-C1/Alhydrogel with and without CPG 7909 in malaria naive adults. PLoS One. 2010;5:e8787.
126. Sheehy SH, Duncan CJ, Elias SC, et al. Phase Ia clinical evaluation of the *Plasmodium falciparum* blood-stage antigen MSP1 in ChAd63 and MVA vaccine vectors. Mol Ther. 2011;19:2269–2276.
127. Malkin E, Hu J, Li Z, et al. A phase 1 trial of PfCP2.9: an AMA1/MSP1 chimeric recombinant protein vaccine for *Plasmodium falciparum* malaria. Vaccine. 2008;26:6864–6873.
128. Sheehy SH, Duncan CJ, Elias SC, et al. ChAd63-MVA-vectored blood-stage malaria vaccines targeting MSP1 and AMA1: assessment of efficacy against mosquito bite challenge in humans. Mol Ther. 2012;20:2355–2368.
129. Biswas S, Choudhary P, Elias SC, et al. Assessment of humoral immune responses to blood-stage malaria antigens following ChAd63-MVA immunization, controlled human malaria infection and natural exposure. PLoS One. 2014;9:e107903.
130. Chitnis CE, Mukherjee P, Mehta S, et al. Phase I clinical trial of a recombinant blood stage vaccine candidate for *Plasmodium falciparum* malaria based on MSP1 and EBA175. PLoS One. 2015;10:e0117820.
131. Hall R, Hyde JE, Goman M, et al. Major surface antigen gene of a human malaria parasite cloned and expressed in bacteria. Nature. 1984;311:379–382.
132. Epp C, Kauth CW, Bujard H, et al. Expression and purification of *Plasmodium falciparum* MSP-1(42): a malaria vaccine candidate. J Chromatogr B Analyt Technol Biomed Life Sci. 2003;786:61–72.

133. Dijkman PM, Marzluf T, Zhang Y, et al. Structure of the merozoite surface protein 1 from *Plasmodium falciparum*. Sci Adv. 2021;7. doi: 10.1126/sciadv.abg0465
134. Kauth CW, Epp C, Bujard H, et al. The merozoite surface protein 1 complex of human malaria parasite *Plasmodium falciparum*: interactions and arrangements of subunits. J Biol Chem. 2003;278:22257–22264.
135. Joos C, Marrama L, Polson HEJ, et al. Clinical protection from falciparum malaria correlates with neutrophil respiratory bursts induced by merozoites opsonized with human serum antibodies. PLoS One. 2010;5:e9871.
136. Boyle MJ, Reiling L, Feng G, et al. Human antibodies fix complement to inhibit *Plasmodium falciparum* invasion of erythrocytes and are associated with protection against malaria. Immunity. 2015;42:580–590.
137. Reiling L, Boyle MJ, White MT, et al. Targets of complement-fixing antibodies in protective immunity against malaria in children. Nat Commun. 2019;10:610.
138. Hill DL, Eriksson EM, Li Wai Suen CSN, et al. Opsonising antibodies to P. falciparum merozoites associated with immunity to clinical malaria. PLoS One. 2013;8:e74627.
139. Osier FH, Feng G, Boyle MJ, et al. Opsonic phagocytosis of *Plasmodium falciparum* merozoites: mechanism in human immunity and a correlate of protection against malaria. BMC Med. 2014;12:108.
140. Kana IH, Garcia-Senosiain A, Singh SK, et al. Cytophilic antibodies against key *Plasmodium falciparum* blood stage antigens contribute to protection against clinical malaria in a high transmission region of Eastern India. J Infect Dis. 2018;218:956–965.
141. Odera DO, Tuju J, Mwai K, et al. Antibodies targeting merozoites induce natural killer cell degranulation and interferon gamma secretion and are associated with immunity against malaria. Sci Transl Med. 2023;151:eabn5993. doi: 10.1126/scitranslmed.abn5993.
142. Feng G, Boyle MJ, Cross N, et al. Human immunization with a polymorphic malaria vaccine candidate induced antibodies to conserved epitopes that promote functional antibodies to multiple parasite strains. J Infect Dis. 2018;218:35–43.
143. Garcia-Senosiain A, Kana IH, Singh S, et al. Neutrophils dominate in opsonic phagocytosis of P. falciparum blood-stage merozoites and protect against febrile malaria. Commun Biol. 2021;4:984.
144. Nkumama IN, Odera D, Musasia F, et al. Breadth of Fc-mediated effector function delineates grades of clinical immunity following human malaria challenge. bioRxiv. 2022;2022:2010.2011.511755.
145. Rosenkranz M, Nkumama IN, Kraker S, et al. Full-length merozoite surface protein 1 of *Plasmodium falciparum* is a major target of protective immunity following controlled human malaria infections. medRxiv. 2022;2022.2010.2012.22280947.
146. Rosenkranz M, Fürle K, Hibbert J, et al. Multifunctional IgG/IgM antibodies and cellular cytotoxicity are elicited by the full-length MSP1 SumayaVac-1 malaria vaccine. NPJ Vaccines. 2023;8:112.
147. Marsh K, Kinyanjui S. Immune effector mechanisms in malaria. Parasite Immunol. 2006;28:51–60.
148. Conway DJ, Greenwood BM, McBride JS. Longitudinal study of *Plasmodium falciparum* polymorphic antigens in a malaria-endemic population. Infect Immun. 1992;60:1122–1127.
149. Riley EM, ALLEN SJ, Wheeler JG, et al. Naturally acquired cellular and humoral immune responses to the major merozoite surface antigen (Pf MSP1) of *Plasmodium falciparum* are associated with reduced malaria morbidity. Parasite Immunol. 1992;14:321–337.
150. Volz JC, Yap A, Sisquella X, et al. Essential role of the PfRh5/PfRipr/CyRPA complex during *Plasmodium falciparum* invasion of erythrocytes. Cell Host Microbe. 2016;20:60–71.
151. Douglas AD, Baldeviano G, Lucas C, et al. A PfRH5-based vaccine is efficacious against heterologous strain blood-stage *Plasmodium falciparum* infection in Aotus monkeys. Cell Host Microbe. 2015;17:130–139.
152. Payne RO, Silk SE, Elias SC, et al. Human vaccination against RH5 induces neutralizing antimalarial antibodies that inhibit RH5 invasion complex interactions. JCI Insight. 2017;2. doi: 10.1172/jci.insight.96381
153. Douglas AD, Baldeviano GC, Jin J, et al. A defined mechanistic correlate of protection against *Plasmodium falciparum* malaria in non-human primates. Nat Commun. 2019;10:1953.
154. Minassian AM, Silk SE, Barrett JR, et al. Reduced blood-stage malaria growth and immune correlates in humans following RH5 vaccination. Med (N Y). 2021;2:701–719.e719.
155. Tan J, Sack BK, Oyen D, et al. A public antibody lineage that potently inhibits malaria infection through dual binding to the circumsporozoite protein. Nat Med. 2018;24:401–407.
156. Kayentao K, Ongoiba A, Preston AC, et al. Safety and efficacy of a monoclonal antibody against malaria in Mali. N Engl J Med. 2022;387:1833–1842.
157. Chappel JA, Holder AA. Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion in vitro recognise the first growth factor-like domain of merozoite surface protein-1. Mol Biochem Parasitol. 1993;60:303–311.
158. Uthaipibull C, Aufiero B, Syed SEH, et al. Inhibitory and blocking monoclonal antibody epitopes on merozoite surface protein 1 of the malaria parasite *Plasmodium falciparum*11Edited by J. A. Wells. J Mol Biol. 2001;307:1381–1394.
159. Locher CP, Tam LQ, Chang SP, et al. *Plasmodium falciparum*: gp195 tripeptide repeat-specific monoclonal antibody inhibits parasite growth in vitro. Exp Parasitol. 1996;84:74–83.
160. Patel PN, Dickey TH, Hopp CS, et al. Neutralizing and interfering human antibodies define the structural and mechanistic basis for antigenic diversion. Nat Commun. 2022;13:5888.
161. Lin CS, Uboldi AD, Epp C, et al. Multiple *Plasmodium falciparum* merozoite surface protein 1 complexes mediate merozoite binding to human erythrocytes. J Biol Chem. 2016;291:7703–7715.

4.2 Whole-genome sieve analysis: identification of protective malaria antigens by leveraging allele-specific vaccine efficacy.

Manuscript under review at Vaccine.

Whole-genome sieve analysis: identification of protective malaria antigens by leveraging allele-specific vaccine efficacy.

Ryan Scalsky[a], Ankit Dwivedi[a,b], Thomas Stabler[c,d], Gillian Mbambo[a], Amed Ouattara[e], Kirsten E. Lyke[e], Shannon Takala-Harrison[e], Joana C. Silva[a,b,f]

[a] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, USA

[b] Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, USA

[c] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland

[d] University of Basel, Basel, Switzerland

[e] Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, USA

[f] Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade NOVA de Lisboa (NOVA), Lisbon, Portugal

Corresponding Author: Joana C. Silva (jcsilva@som.umaryland.edu)

Full postal address:

University of Maryland School of Medicine
670 W. Baltimore St., Bldg HSF-3, Room 3106
Baltimore, MD 21201
USA

Abstract

Discovery of new protective malaria antigens will enable the development of novel vaccine formulations with potentially higher efficacy. While several high-throughput experimental approaches enable the identification of novel immunogens, none so far has been designed to selectively identify protective antigens. Here, we propose that whole-genome sieve analysis (wgSA) can be used specifically for this purpose. We review available high-throughput methods for antigen identification and contextualize the need for the identification of protective antigens. We then provide the rationale for why wgSA is ideally suited for the identification of protective antigens in recombining pathogens with large genome size, describe necessary conditions for optimal use, and discuss potential pitfalls. Most importantly, this approach can be applied to the discovery of new protective targets in any recombining organism for which there is a whole organism-based vaccine that can be safely deployed in a disease-endemic region.

Keywords: Whole-genome sieve analysis; wgSA; Malaria; vaccine; *P. falciparum*; Protective antigen identification

Introduction

The roadmap to malaria eradication requires vaccines with a minimum of 75% protective efficacy [1]. Eukaryotic pathogens in the genus *Plasmodium* are the deadliest mosquito-borne infectious pathogens infecting humans. *Plasmodium falciparum* exerts an immense burden globally, causing the vast majority of the 249 million malaria cases and resulting in 619,000 deaths estimated to have occurred worldwide in 2022 [2]. After over half a century of malaria vaccinology research, the World Health Organization (WHO) has now recommended two vaccines against *P. falciparum*, RTS,S/AS01 (commercialized as Mosquirix™) and R21/Matrix-M® [3]. Notwithstanding this momentous development, the formulation of both vaccines is based on *P. falciparum* circumsporozoite protein (PfCSP), which has shown allele-specific efficacy as part of RTS,S field trials [4, 5], demonstrating susceptibility to vaccine escape. This raises the possibility that widespread distribution of these vaccines may lead to selection and concomitant increase in frequency of strains capable of evading the cross-protection induced by these vaccines, and leading to decreased vaccine efficacy --a threat that has been discussed for well over a decade [6]. Therefore, it is imperative to identify additional parasite proteins that can impart effective protection, to enable the development of the next generation of malaria vaccines with broader protective efficacy. Here, we argue that allele-specific vaccine efficacy, while a detrimental phenomenon for any one vaccine, can be leveraged to identify new antigens that contribute to protection, in infectious organisms that satisfy specific conditions. These conditions include a pathogen that undergoes recombination, a vaccine based on the whole organism, testing conducted in high-transmission field settings, and against natural exposure, and where there is evidence of allele-specific vaccine efficacy. *P. falciparum* is a pathogen that satisfies all these criteria; therefore, we describe the use of allele-specific efficacy for the identification of protective antigens in the context of malaria.

Allele-specific vaccine efficacy in advanced malaria vaccines

Malaria vaccines can be broadly classified into pre-erythrocytic, blood stage, and transmission-blocking vaccines based upon their mechanism of action, including, respectively, the prevention of infection within the liver, the blocking of erythrocytic infection, or the prevention of successful transmission within or by the mosquito vector [7, 8]. Of the many malaria vaccines currently undergoing Phases 1 - 4 clinical trials [9], so far two have been endorsed by the WHO for use in malaria-endemic regions with moderate to high malaria transmission. These are sub-unit vaccines targeting the most abundant protein on the surface of the *P. falciparum* sporozoite, circumsporozoite protein (PfCSP). The first to be recommended for implementation in children aged 5-17 months was Mosquirix™ (RTS,S/AS01) in 2021, followed recently by R21/Matrix-M in 2023 [10, 11]. Both are pre-erythrocytic-stage vaccines, targeting the estimated 5 to 100 sporozoites transferred into the human skin during mosquito feeding [12, 13]. The most developed alternatives to RTS,S/AS01 and R21/Matrix-M are two pre-erythrocytic whole-organism vaccines based on sporozoites (SPZ) of the *P. falciparum* strain NF54: PfSPZ Vaccine, based on radiation-attenuated, metabolically active, non-replicative sporozoites, and PfSPZ CVac, which uses fully infectious, non-attenuated sporozoites administered under the cover of a chemoprophylactic. These two SPZ-based vaccines are now in late development stages with promising results [14].

RTS,S has been shown to induce allele-specific protection, the phenomenon by which vaccines are most efficacious against pathogen strains that encode an antigen identical to that used in the vaccine formulation but have lower efficacy against strains that encode distinct variants of the same protein [4, 15]. Allele-specific efficacy, by preferentially preventing infection by strains with vaccine-like genotypes (immunologically homologous strains), provide a selective advantage to strains that escape vaccine-induced protection, leading to their increase in frequency and, presumably, an eventual decrease in vaccine efficacy. In malaria, this phenomenon has been most extensively characterized in the context of an apical membrane

antigen 1 (AMA)-based vaccine, AMA1/AS02A [16, 17]. This vaccine uses the AMA1 variant encoded by Pf3D7, the reference *P. falciparum* strain, adjuvanted to GlaxoSmithKline's AS0 adjuvant platform, and was shown to have >60% efficacy against homologous parasite strains. However, this locus is highly variable in *P. falciparum* populations, and the vaccine showed minimal protection against most circulating Malian *P. falciparum* strains encoding variants of AMA1 that differed from Pf3D7 at key residues within B cell epitopes [18-20].

Unlike sub-unit vaccines, the whole organism vaccine approach provides greater breadth in antigenic presentation, characterized by numerous antigens for immune priming and a much larger B and T cell epitope pool. This diversity of immunogens within whole organism vaccines may overcome two key potential challenges: genetic differences between the vaccine strain and all circulating strains against which the vaccine is intended to protect (which, in the presence of allele-specific vaccine efficacy, result in vaccine escape by a subset of strains), as well as differences in HLA-based antigen recognition, which result in host-related variability in protection [21]. By addressing these challenges, whole organism vaccines may have a higher likelihood of protection against the diversity of pathogen strains circulating in natural populations. Identification of allele-specific efficacy for whole organism malaria vaccines is not straightforward, since the sporozoite expressed over a thousand proteins [22], and yet it is unclear which subset of those is recognized by the immune system and is responsible for the protective response. Initial clinical trials where efficacy was assessed by exposure to *P. falciparum* laboratory strains via controlled human malaria infection (CHMI), found protection to be lower in heterologous CHMI (where the strain used for challenge differs from the one in the vaccine) than in homologous CHMI, where a strain immunologically identical to the one in the vaccine is used for challenge [23]. These results are consistent with protection that is contingent on genotype, i.e., allele-specific vaccine efficacy. We note, however, that this relationship is nuanced. More recent CHMI studies have been performed protection to be in part dose-

dependent. These variable effects of dose were attributed to the increasing dose eliciting a broader T-cell response which may overcome a degree of allele-specific efficacy; however, protection against heterologous strains still did not meet the 75% durable efficacy predicted to allow malaria eradication through vaccination. Additionally, the authors clarify that all participants in these CHMI studies were malaria-naïve US adults not representative of individuals living in malaria endemic regions [24]. Multiple prior studies have demonstrated that the dynamics of immune response are heavily modified by age, ethnicity, prior exposure and hemoglobinopathies, and thus further investigation is required to better understand if traditional methods can overcome allele-specific efficacy in endemic regions [25, 26]. Considering this, to overcome allele specificity, the next generation of vaccines can combine multiple variants of several protective antigens, as either multivalent or multi-strain vaccines [15, 27]. To this end, the identification of novel, protective antigens is essential.

Experimental approaches to identify novel malaria antigens

The development of highly efficacious vaccines is contingent on the identification of antigens that are both immunogenic and, principally, protective against infection. A variety of experimental approaches currently exist to identify and validate immunogenic targets. Here, we briefly describe recent, medium- or high-throughput approaches for the rapid, simultaneous experimental screening of a large number of potential candidates. These approaches include peptide/protein-based assays and novel *in vitro* and *ex vivo* techniques. Both humoral and cell-mediated immunity are important in the context of malaria; however, antibody responses are particularly pertinent during blood stage malaria as well as against sporozoites, prior to hepatocyte invasion, while T cell responses are key during the liver stages of infection.

Assessment of Humoral Immunity

Protein arrays have allowed the validation and discovery of novel immunogens, while peptide

arrays have enabled more granular knowledge such as the confirmation of immunogenic sites and the mapping of epitopes within known antigens. High-density peptide arrays used for the discovery and characterization of immunogenic parasite targets were described in 2016, though peptide arrays have been used to investigate humoral response to malaria infection as early as 2008 [28, 29]. This method involves the development of peptide arrays containing 10^3 to $>10^5$ individual peptides per array that represent known or suspected antigens or epitopes involved in immune response to infection [30]. These arrays are then probed with serum from study participants, at least a fraction of whom have had natural infections or have been immunized. Detection of peptide-antibody binding is usually done in a label-dependent manner, most commonly by addition of a secondary antibody with a fluorescent tag [30, 31]. This process allows the interrogation of responses to immunogenic proteins involved in naturally acquired or vaccine-induced humoral (B-cell) immunity. Additionally, these methods can incorporate antibodies specific to IgG, IgM or IgA coupled to fluorescent dyes to investigate binding and prevalence of antibody classes in a participant's serum [28]. Peptide arrays have been used to successfully validate epitopes, determine differences in the strength of epitope recognition and location of epitopes between variants of the same protein, and compare the immune response elicited by infection *vs*. vaccination, as well as to identify novel antigens [32-35].

A novel, peptide-based method known as PepSeq combines in-solution antibody recognition of DNA-barcoded peptides with high-throughput sequencing to identify epitopes in a semi-quantitative, highly-multiplexed approach [36]. In the PepSeq approach, nucleotide sequences that encode up to hundreds of thousands of peptides of interest are transcribed and translated *en masse*, with each newly synthesized peptide being covalently bound to the cDNA (which will serve as a specific barcode) which is reverse transcribed from its encoding mRNA, to form the individual PepSeq probes. Incubation of these barcoded PepSeq probes with serum samples to be queried results in the recognition of specific peptides by the antibodies present, in a process

that is dependent both on the frequency of antibodies and their affinity for their target peptides. PepSeq probes that remained unbound, in solution, are removed, and the remaining cDNAs, barcoding the antibody-bound PepSeq probes, can subsequently be PCR amplified and high-throughput sequenced. In this way, the resulting DNA sequences are a semi-quantitative measure of peptide immunogenicity [36]. PepSeq is a highly specific and customizable approach which can be applied to investigate humoral responses in a myriad of infectious targets, with the tremendous advantage that antibody responses to $>10^5$ specific peptides can be assayed jointly, using as little as <1 μL of serum [36]. Its application in the context of SARS-CoV-2, and to characterize infection history and antibody dynamics across several dozen viral species simultaneously have already been reported [37, 38]. Applications to the study of responses to malaria infection and immunization are underway (A. Berry, personal comm.).

Assessment of Cell-Mediated Immunity

Despite an established and critical role of T cell-mediated protection following immunization with *P. falciparum* sporozoites [39-41], significantly less investigation has been done using proteins or peptide pools to identify *P. falciparum* targets of cell-mediated immunity (CMI) owing to the requirement of large amounts of peripheral blood mononuclear cells (PBMC) from human participants. Additional challenges posed by assays to query CMI include the requirement for antigen-presenting cells (APCs), primed T cells and detection of cytokines as markers of activation [42]. Despite these limitations, it is still possible to identify T-cell antigen targets through several platforms including flow cytometry and mass spectroscopy; however, the most sensitive of methodologies remains the T cell ELISpot assay [43-46]. This assay is a modified version of traditional ELISpot that involves incubating T cells, APCs and proteins or peptides of interest in wells that are coated with capture antibodies which can selectively bind secreted cytokines of interest. These cytokines are usually markers of activation, which indicate successful binding and presentation of antigens to T cells following cellular stimulation, and

include interferon gamma (IFN-γ) and tumor necrosis factor alpha (TNF-α) for CD8+ T cells and IFN-γ, interleukin-2 (IL-2), and interleukin-4 (IL-4) for CD4+ T helper type 1 and type 2 cells [47]. After binding of the cytokine of interest to the capture antibody, a secondary biotinylated detection antibody is added which binds the capture antibody-cytokine complex. Lastly, streptavidin-enzyme conjugate which selectively binds the secondary antibody and converts the substrate in each well to countable spots [48].

Preliminary studies investigating cell-mediated immunological responses in malaria have been performed [49, 50]. In one study the authors generated peptides corresponding to epitopes within malaria antigens previously known to elicit an HLA A2-specific cytotoxic CD8+ T lymphocyte responses, and stimulated PBMCs from malaria-positive Malian children aged 3 months to 14 years and healthy controls (age- and residence-matched, asymptomatic, aparasitemic individuals), for T cell activation using ELISpot [49]. This study provided evidence of functional CMI targeting pre-erythrocytic malaria antigens, and that HLA A2 supertypes were defined as such to mitigate the limitations of HLA restriction. A more recent study assessed the reactivity of CD8+ T cells in PBMCs from malaria naïve adults after administration of a recombinant adenovirus vector expressing *P. falciparum* AMA1. This was achieved by creating pools of overlapping peptides, each of 15 amino acid residues in length (15-mers), covering the entire length of AMA1, to stimulate participants PBMCs and then assessed reactivity using IFN-γ ELISpot. This study investigated seven HLA alleles within four HLA supertypes [51]. Additionally, the authors predicted MHC-I epitopes, bioinformatically, using NetMHCpan [52], which were then confirmed via ELISpot [51]. This study expanded upon previous work by combining peptide pools with ELISpot and *in silico* prediction methods to investigate CD8+ T cell responses across multiple HLA types. While ELISpot enables accurate epitope validation, it is not amenable, logistically or cost-wise, to antigen discovery in a meaningful scale.

Though in preliminary stages of development, recent advancements may enable the high-throughput investigation of $CD8^+$ T cell responses using peptide microarrays in a not-too-distant future. In 2023, Kramer and colleagues published an exciting, novel approach to evaluate T cell receptor (TCR) binding to potentially thousands of peptide-human leukocyte antigen (pHLA) complexes [53]. This method involves printing peptides and spotting them, is suspension, across distinct wells in a microarray. This peptide suspension can be allowed to dry and stored for future use. Once specific HLAs are introduced in each well, in a small volume, the pHLA complexes form. These pHLAs are immobilized onto a slide, applied on top of the chip such that it closes the wells, using a biotin-streptavidin conjugation system. After the slide is removed and washed, the binding dynamics of each fixed pHLA to a specific TCR to be queried is quantified, using single color reflectometry, by flowing the TCR in suspension over the slide [53]. This and similar approaches may pave the way for broad and agnostic querying of $CD8^+$ T cell responses to target peptides.

Trogocytosis, the process by which cells share membrane and membrane-associated proteins during binding, has been leveraged to identify target cells which have been engineered to present epitopes specific to TCR-transduced T cells [54, 55]. This process allowed researchers to generate libraries of known epitopes from public databases, transduce them into artificial antigen presenting cells, and then expose these to T cells transduced with a known TCR to assess antigen-TCR specific trogocytosis via fluorescent activated cell sorting (FACS). This process allowed for TCR-specific ligand discovery in a high throughput fashion that did not require prior knowledge of epitope structure. This process was able to assess MHC I interactions but could be adapted to MHC II – a considerably understudied aspect of malaria immunity. The approach offers the potential to identify protective epitopes at a higher frequency compared to alternative methods, given the increased occurrence of trogocytosis in high-avidity

interactions between MHC/ligands pairs and TCRs, thereby facilitating enhanced CD8 T cell cytolytic activity [56].

A recently proposed method, known as T-scan, provides a higher throughput approach to T cell targeted epitopes relative to trogocytosis, with less cell culture requirements. It consists of co-culturing a participant's T cells with antigen presenting cells that have been engineered using lentiviral-transfected proteins to display the proteome of a target pathogen on endogenous MHC I. Antigen binding by cognate TCRs on the participant's T cells are then monitored through a fluorescent reporter activated by granzyme B in target cells expressing a given antigen. These are then sorted and isolated by FACS. Lastly, polymerase chain reaction (PCR) and next generation sequencing are used to identify the antigens that the respective target, sorted cell is expressing. This method provides a high throughput agnostic approach for the identification of antigens eliciting responses in participant's samples [57].

Limitations of existing approaches

The peptide-based approaches discussed are considered medium- to high-throughput, in that they allow the screening of immune responses to many dozens to hundreds of thousands of potential epitopes. Nevertheless, those still represents only a limited number of all potential targets, given all existing protein variants encoded by different alleles of each of the ~5,500 protein-coding loci in the genome of *P. falciparum*.

To investigate B cell responses, protein and peptide arrays are currently limited to linear B cell epitopes. Methods do exist to investigate discontinuous, or conformational, B cell epitopes using structures derived from crystallographic analysis, but these are expensive and time consuming, though studies have demonstrated that discontinuous B-cell epitopes do exist and warrant investigation [58].

Additionally, in the context of T cell responses, there is the tremendous diversity of HLA alleles to which different pathogen epitopes are restricted, such that different HLA alleles may recognize distinct epitopes in each antigen as well as in each of their respective variants. Given this, for any studies investigating *ex vivo* T cell responses, the HLA genotype of participants involved must be obtained so that responses can be interpreted which mandates additional financial and ethical considerations. Of note is that HLA loci are some of the most diverse in the human genome and their frequency distribution, particularly in developing countries, is not well characterized [59]. In this context, the approaches discussed above for T cells are not entirely agnostic, as a set of potential HLA allele(s) and targets are selected *a priori*. Additionally, the high throughput peptide microarray method for investigating pHLA-TCR binding is limited to investigating individual TCRs. In investigating natural $CD8^+$ T cell responses it is essential that a broad variety of TCRs be able to be screened in combination with an even broader repertoire of pHLA complexes. This profoundly diverse combinatorial binding continues to pose serious challenges in investigating $CD8^+$ mediated immunity using peptide microarrays.

Above all, and despite the exciting technological progress in throughput, the acute need for agnostic approaches that identify antigens are both immunogenic and, most of all, protective, remains.

Leveraging allele-specific vaccine efficacy to identify new protective antigens

Sieve analysis and allele-specific vaccine efficacy

In 2001, Gilbert and colleagues proposed an approach, which was coined “sieve analysis”, to determine the extent to which the genotype of the infecting pathogen impacts vaccine efficacy. This is achieved by quantifying the relative risk of infection, post-vaccination, associated with pathogens encoding each variant of the vaccine immunogen [60]. In other words, sieve analysis quantifies the risk of vaccine evasion by each antigenic variant relative to the vaccine variant.

When that risk differs significantly from 1, for one or more variants, the vaccine is said to have allele-specific efficacy. This quantification relies on the characterization of the frequency of each variant of the target antigen among unvaccinated participants (i.e., placebo recipients), which is assumed to reflect the frequency of these variants in the pathogen population. Then quantifying how those frequencies are altered in vaccinees who get infected, under the assumption that observed differences in the two study arms are due to a sieve-like process effected by the primed immune system, which “catches” variants immunologically similar to the vaccine variant but has more limited protection against other variants (**Figure 1**). The authors applied sieve analysis to studies testing the efficacy of vaccines against cholera, HIV-1, hepatitis B, rotavirus and pneumococcus, demonstrating broad applicability to pathogens despite vast taxonomic distances, as well as differences in etiology and protection mechanism [60].

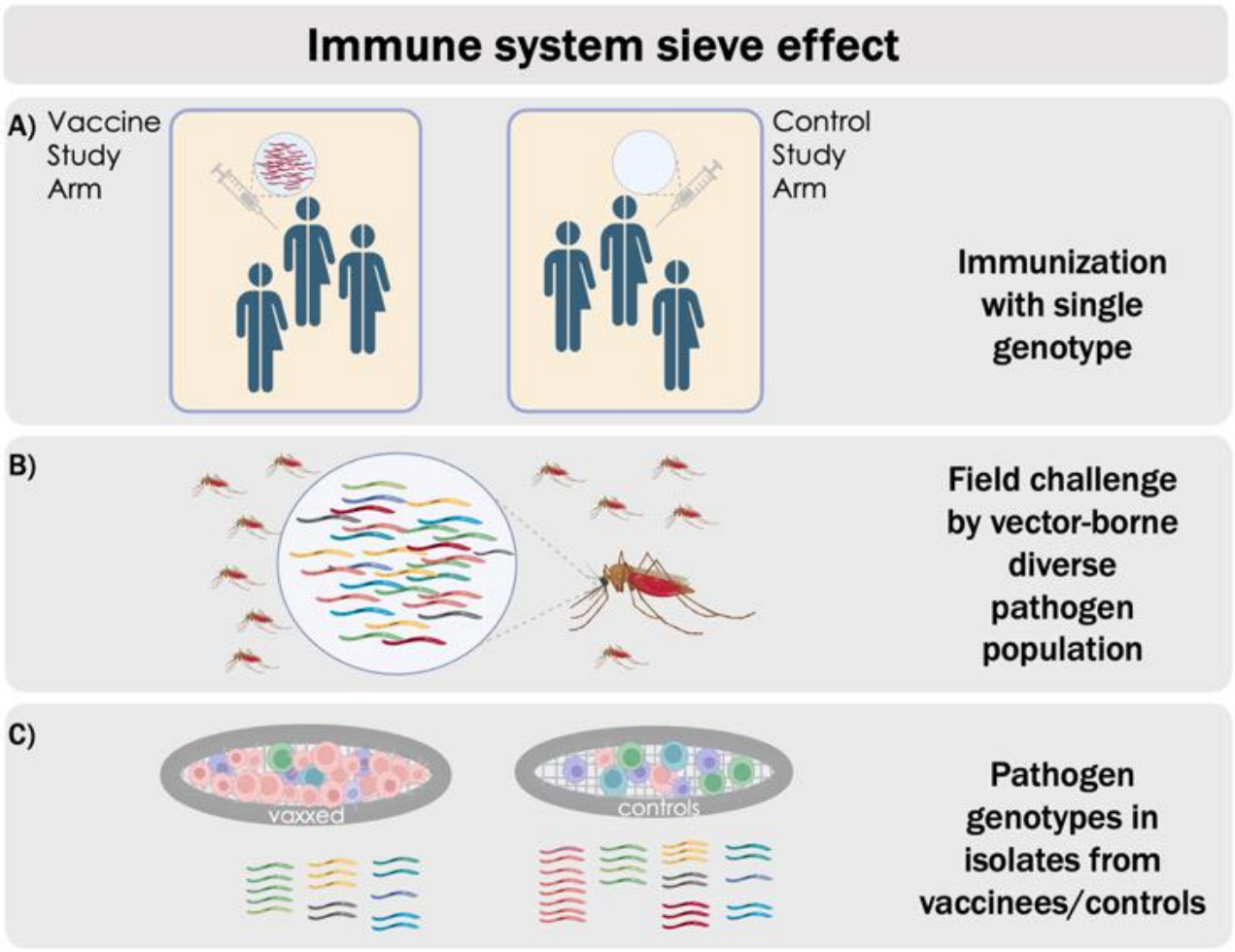


Figure 1. Sieve effect associated with vaccination. A) Participants in a malaria vaccine trial are either immunized with a "red genotype" or receive a placebo, such as saline solution. B) Vaccine trial participants are exposed to vector transmitting a genotypically diverse parasite population, represented by multiple colors. C) The frequency of different variants differs in infections from vaccinees and controls, due to the sieve effect associated with the immune system, shown as immune cells on a sieve, which was trained through vaccination to recognize a "red" genotype and others immunologically similar. Therefore, "red" and "deep red" genotypes are absent from malaria isolates collected from vaccinees. Figure prepared with BioRender.

Modifications to the initial sieve analysis have been introduced, aimed at testing hypotheses related to the mechanism of vaccine-induced protection, or host factors that contribute to modify vaccine efficacy [5, 61-64]. Sieve analyses are now regularly conducted on vaccine efficacy data for a variety of viral pathogens, including most recently HIV-1 and SARS-CoV-2 vaccines. These analyses help interpret vaccine efficacy outcomes by determining if vaccines are allele/genotype-specific, how pathogen genetic diversity impacts efficacy, and to shed light on epitopes or specific amino acid residues that are important for protection [65-68].

In malaria, sieve analysis was applied to study the cross-protective efficacy of a malaria vaccine, FMP2.1/AS02$_A$, based upon a single variant of the apical membrane antigen-1, AMA1. The locus encoding AMA1 can exhibit extreme variability in malaria-endemic regions with hyper-endemic transmission, making this vaccine susceptible to vaccine escape [20]. The authors observed that while the vaccine exhibited >60% efficacy against parasites that encoded an AMA1 variant identical to the vaccine strain at immunologically relevant loci, overall efficacy was very low due to the limited frequency of the vaccine variant in the region where the vaccine was deployed. Key amino acid residues and epitopes were identified as important for protection [16, 17, 69]. More recently, the sieve analysis approach has been applied to samples from field trials of the RTS,S/AS01 malaria vaccine, demonstrating that the vaccine has allele-specific protection, with efficacy being highest against *P. falciparum* strains with alleles identical to the vaccine. Efficacy decreased with increasing genetic distance, between the vaccine and challenge strains, in amino acid residues in the Th3R T cell epitope region of PfCSP, but found no modification of the sieve effect by host age [4, 5, 61, 62].

These observations have led to a large number of molecular epidemiological studies in *P. falciparum* in recent years, which aim to characterize genetic diversity in, and the selective pressures acting on, *Plasmodium* vaccine candidates to assess the possibility of vaccine escape, and to attempt to identify conserved antigens – or, at minimum, the most abundant and geographically distributed variants of each antigen [18, 70-72]. Due to the advanced stage of several CSP-based vaccines, many studies have focused on the CSP encoding gene, assessing genetic diversity in parasite populations around the globe, not only *P. falciparum*, but also other species of human-infecting *Plasmodium* [73-76].

Identification of protective targets of vaccination thought whole-genome sieve analysis (wgSA)

Traditional sieve analysis, where the immunogen is known, enables the assessment of how parasite genotype and host factors contribute to modify vaccine efficacy. In 2015, sieve analysis

was extended to the whole genome of HIV-1, termed whole-genome sieve analysis, wgSA [64]. A wgSA consists in the identification of "*target sites*", i.e., *genomic sites that differed significantly in genotype between vaccinees and controls and in which the vaccine genotype is depleted among vaccinees*. Target sites were identified both in regions contained in the vaccine immunogen and in regions that were not. The signal from the former regions was likely associated with vaccine-induced protection; the difference in allele frequency in vaccinees *vs*. controls, with the reduction of the vaccine variant among vaccinee infections, was likely due to the immune system's sieve effect, which blocks infection by strains encoding residues similar to those in the vaccine. Conversely, "target sites" (or perceived positives) in regions that were not part of the vaccine immunogen could be due to genetic linkage between those and variants in truly protective sites [64]. Therefore, genetic linkage can be a source of noise in a wgSA and should be evaluated in wgSA applications.

Here, we contend that wgSA can be applied successfully to pathogens of large genome size, provided recombination is high to minimize linkage disequilibrium, to identify genomic sites associated with vaccine-induced protection and, in this way, enable the identification of protective loci. In particular, when the vaccine immunogen is the whole organism, when protection is assessed against natural challenge in a region of high attack rate (i.e., rate of challenge is high), when disease transmission intensity is high (resulting in high pathogen genetic diversity), and when rate of effective recombination is high, such that genetic linkage is low, the "target sites" identified are likely to be enriched for protective sites (**Figure 2**). *P. falciparum*, a parasite with high effective recombination rate in geographic regions where malaria transmission is high [77-79], and for which there are whole-organism vaccines [80], is an ideal pathogen with which to test wgSA.

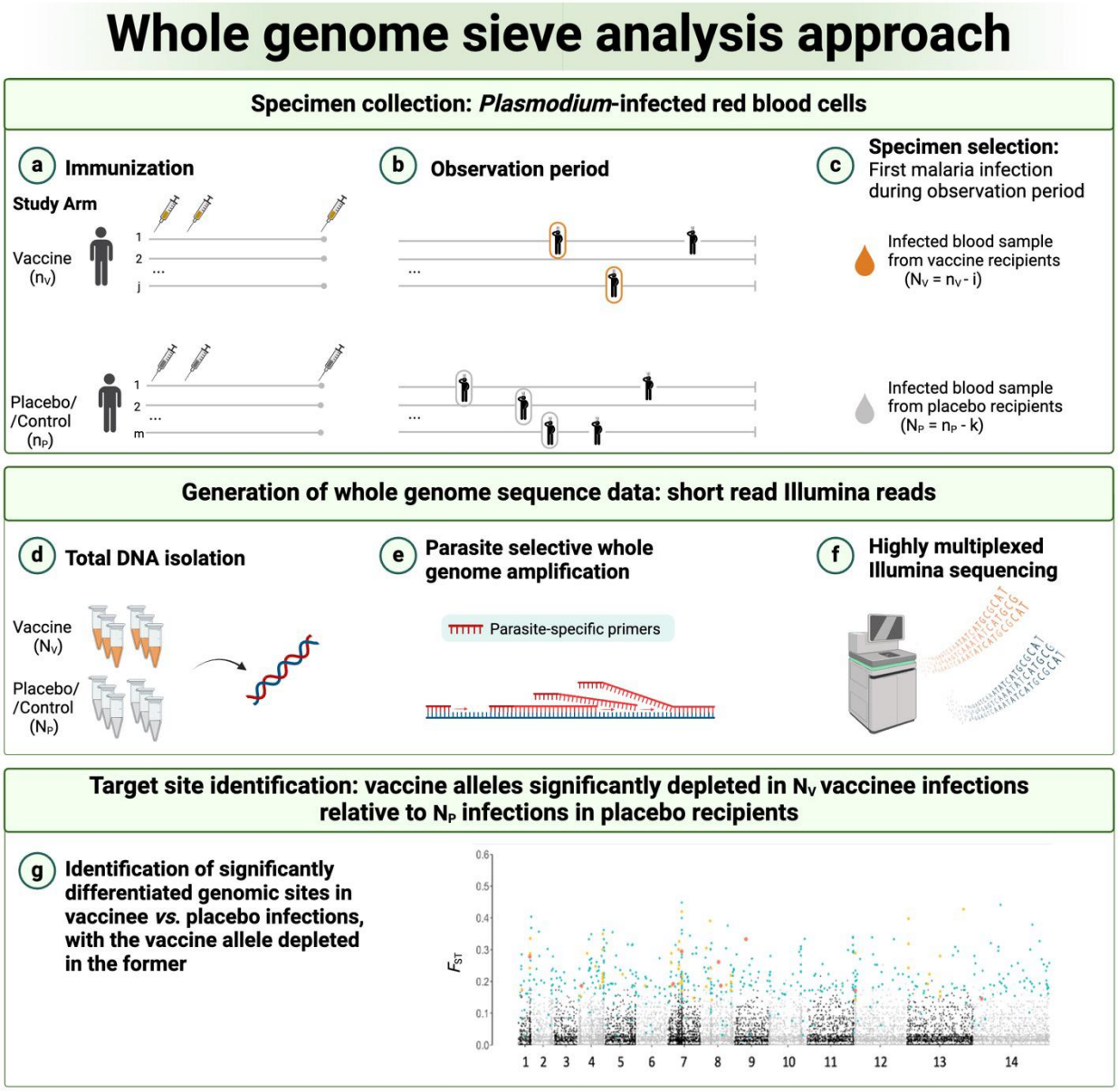


Figure 2. Schematic of a whole-genome sieve analysis (wgSA) study. a) Participants in a field trial to study efficacy of a whole-organism vaccine receive a vaccine (nV) or a placebo (nP), according to study arm. b) During the observation period, participants are challenged by natural infection and while some are never infected (i vaccinees and k controls), others become infected once or more during this time (NV vaccinees and NP placebo recipients). c) A blood sample, corresponding to first infection post immunization, is obtained from each infected participant and (c) processed, to obtain whole genomic DNA. e) If leukocytes were not removed during blood collection, total DNA can be subjected to selective (parasite) whole genome amplification to enrich the sample for the desired DNA. f) A library is prepared for each sample, and libraries are multiplexed for high-throughput sequencing. g) wgSA is conducted on whole genome sequencing data to identify significantly differentiated sites, genome-wide, between infections from vaccinees and placebo recipients, in which the vaccine allele is depleted in vaccinee samples. Figure prepared with BioRender.

By screening the entire genome, wgSA provides a high-throughput approach to identify potentially protective sites. It requires no prior knowledge of antigen targets and, by surveying

the entire genome, it is both agnostic and, in principle, able to identify a full list of protective targets. To add power to genome-wide statistical analysis, wgSA can be limited to genomic sites likely to be causative, such as protein-coding regions, non-synonymous sites within these regions, or even coding regions with predicted epitopes, much like proposed before with the analysis of HIV-1 sequences [64]. For experimental validation, any comprehensive list of "target sites" can then be prioritized according to parameters known to be associated with immunogenicity, such as target sites that fall in known antigen-encoding loci, loci encoding proteins with predicted subcellular localization that exposes them to the host's immune system, or loci in which target sites fall preferentially in epitope-encoding regions. Lastly, because wgSA is entirely agnostic, it can investigate non-coding regions, and uncover regulatory mechanisms that influence vaccine escape. While most vaccinology research focuses on antigens, evidence suggests that genetic regulatory mechanisms can confer fitness advantages associated with immune evasion [81].

Limitations of whole-genome sieve analysis

While potentially powerful, wgSA is limited by features of the data that can lead to both false positive and false negative target sites. Target sites and "target loci" (the loci containing one or more target sites) correspond to genomic sites or loci with significantly different allele or variant frequencies in samples from vaccine and placebo recipients, and in which the vaccine allele is depleted in vaccinees. For the purpose of this discussion, alleles (nucleotide-based) and variants (amino acid-based) are broadly defined and can vary in length from individual genomic sites to long haplotypes and from individual amino acid residues, to epitopes, or complete polypeptides.

False positives results can emerge when alleles or variants are present in significantly different frequencies between the two study arms of a vaccine trial simply due to chance. This can easily occur in small studies. Other potential sources of false positive target sites include: (i) alleles or variants that are not protective but that are in linkage disequilibrium with true protective sites; (ii) genotyping errors in members of multigene families (e.g., when loci across strains do not have an orthologous relationship and sequence reads map across paralogs); or (iii) differential assortment of alleles or variants by study arm due to their association with an unknown covariate (e.g., time of sample collection).

Conversely, false negative results can also occur. The most obvious example are loci that confer partial protection and that are invariant in sequence, either due to strong selective constraints or due to lack of time for mutations to accrue. If all sequences are identical, they cannot be differentiated between study arms. False negative results can also emerge (i) when genetic variation in the pathogen population is very high relative to the number of study participants; or (ii) due to disconnect between host immune system and the sequence variants circulating in the pathogen population (e.g., when protection is T cell-dependent and the host populations HLA alleles cannot functionally distinguish between circulating epitope variants).

Concluding remarks

The identification of protective antigens represents a holy grail in vaccinology research, particularly for parasitic infectious diseases. The application of wgSA presents a unique and potentially powerful approach to identify new candidate pathogen targets that induce a protective response, resulting in new vaccine formulations with improved efficacy. The definite risk for false negatives implies that true protective targets can be missed; more importantly, the presence of potential false positives necessitates the experimental validation of target loci using immunological assays and eventually *in vivo* pre-clinical studies in non-human animal models. Malaria caused by *P. falciparum* represents an excellent test case for the application of wgSA,

given the presence of whole-organism vaccines such as PfSPZ Vaccine, and the availability of samples from efficacy field trials of these vaccines in malaria-endemic areas with high, seasonal transmission.

Acknowledgements

We thank all colleagues of the International PfSPZ Consortium (iPfSPZ), conducting research in sporozoite-based malaria vaccines, for the many insightful discussions, and in particular Drs. Stephen Hoffman and Robert Seder, who were early and enthusiastic supporters of the application of wgSA to samples from field trials of PfSPZ vaccines.

Funding

This work was funded in part by National Institutes of Health (NIH) award R01AI141900 to JCS.

ICMJE Authorship criteria statement: All authors attest they meet the ICMJE criteria for authorship.

Conflict of Interest declaration: All authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Bibliography

1. Group, M.V.F., *Malaria Vaccine Technology Roadmap*. 2013.
2. Organization, W.H., *World Malaria Report 2021*. 2021.
3. Vogel, G., *Second malaria vaccine gets WHO green light.* Science, 2023. **382**(6666): p. 16-17.
4. Neafsey, D.E., et al., *Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine.* N Engl J Med, 2015. **373**(21): p. 2025-37.
5. Benkeser, D., P.B. Gilbert, and M. Carone, *Estimating and Testing Vaccine Sieve Effects Using Machine Learning.* J Am Stat Assoc, 2019. **114**(527): p. 1038-1049.
6. Takala, S.L. and C.V. Plowe, *Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming 'vaccine resistant malaria'.* Parasite Immunol, 2009. **31**(9): p. 560-73.
7. Laurens, M.B., *Novel malaria vaccines.* Hum Vaccin Immunother, 2021. **17**(11): p. 4549-4552.
8. Duffy, P.E., *Current approaches to malaria vaccines.* Curr Opin Microbiol, 2022. **70**: p. 102227.
9. P.E. Duffy, J.P.G., *Malaria vaccines since 2000: progress, priorities, products.* npj Vaccines 2020
10. RTS, S.C.T.P., et al., *A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants.* N Engl J Med, 2012. **367**(24): p. 2284-95.
11. Datoo, M.S., et al., *Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial.* Lancet Infect Dis, 2022. **22**(12): p. 1728-1736.

12. Cook, G.C., *Manson's Tropical Infectious Diseases*. 23 ed. Malaria, ed. J. Farrar. 2014.
13. Ramasubramanian, M.K., O.M. Barham, and V. Swaminathan, *Mechanics of a mosquito bite with applications to microneedle design.* Bioinspir Biomim, 2008. **3**(4): p. 046001.
14. Sulyok, Z., et al., *Heterologous protection against malaria by a simple chemoattenuated PfSPZ vaccine regimen in a randomized trial.* Nat Commun, 2021. **12**(1): p. 2518.
15. Ouattara, A., et al., *Designing malaria vaccines to circumvent antigen variability.* Vaccine, 2015. **33**(52): p. 7506-12.
16. Ouattara, A., et al., *Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications.* J Infect Dis, 2013. **207**(3): p. 511-9.
17. Thera, M.A., et al., *A field trial to assess a blood-stage malaria vaccine.* N Engl J Med, 2011. **365**(11): p. 1004-13.
18. Naung, M.T., et al., *Global diversity and balancing selection of 23 leading Plasmodium falciparum candidate vaccine antigens.* PLoS Comput Biol, 2022. **18**(2): p. e1009801.
19. Nirmolia, T., et al., *Genetic diversity of Plasmodium falciparum AMA-1 antigen from the Northeast Indian state of Tripura and comparison with global sequences: implications for vaccine development.* Malar J, 2022. **21**(1): p. 62.
20. Takala, S.L., et al., *Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development.* Sci Transl Med, 2009. **1**(2): p. 2ra5.
21. Nielsen, C.M., et al., *RTS,S malaria vaccine efficacy and immunogenicity during Plasmodium falciparum challenge is associated with HLA genotype.* Vaccine, 2018. **36**(12): p. 1637-1642.
22. Florens, L., et al., *A proteomic view of the Plasmodium falciparum life cycle.* Nature, 2002. **419**(6906): p. 520-6.
23. Epstein, J.E., et al., *Protection against Plasmodium falciparum malaria by PfSPZ Vaccine.* JCI Insight, 2017. **2**(1): p. e89154.
24. Lyke, K.E., et al., *Attenuated PfSPZ Vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection.* Proc Natl Acad Sci U S A, 2017. **114**(10): p. 2711-2716.
25. Travassos, M.A., et al., *Hemoglobin C Trait Provides Protection From Clinical Falciparum Malaria in Malian Children.* J Infect Dis, 2015. **212**(11): p. 1778-86.
26. Bejon, P., et al., *Efficacy of RTS,S malaria vaccines: individual-participant pooled analysis of phase 2 data.* Lancet Infect Dis, 2013. **13**(4): p. 319-27.
27. Trieu, A., et al., *Sterile protective immunity to malaria is associated with a panel of novel P. falciparum antigens.* Mol Cell Proteomics, 2011. **10**(9): p. M111 007948.
28. F.L. Loeffler, J.P., K. Heiss, *High-Density Peptide Arrays for Malaria Vaccine Development.* Vaccine Design 2016. **1403**: p. 569 - 582.
29. Doolan, D.L., et al., *Profiling humoral immune responses to P. falciparum infection with protein microarrays.* Proteomics, 2008. **8**(22): p. 4680-94.
30. Jenne, F., et al., *Resemblance-Ranking Peptide Library to Screen for Binders to Antibodies on a Peptidomic Scale.* Int J Mol Sci, 2022. **23**(7).
31. Forsstrom, B., et al., *Proteome-wide epitope mapping of antibodies using ultra-dense peptide arrays.* Mol Cell Proteomics, 2014. **13**(6): p. 1585-97.
32. Jaenisch, T., et al., *High-density Peptide Arrays Help to Identify Linear Immunogenic B-cell Epitopes in Individuals Naturally Exposed to Malaria Infection.* Mol Cell Proteomics, 2019. **18**(4): p. 642-656.
33. Bailey, J.A., et al., *Microarray analyses reveal strain-specific antibody responses to Plasmodium falciparum apical membrane antigen 1 variants following natural infection and vaccination.* Sci Rep, 2020. **10**(1): p. 3952.

34. Camponovo, F., et al., *Proteome-wide analysis of a malaria vaccine study reveals personalized humoral immune profiles in Tanzanian adults.* Elife, 2020. **9**.
35. DeAnna J. Friedman-Klabanoff, M.A.T., Olukemi O Ifeonu, Sonia Agrawal, Amed Ouattara, Jason A Bailey, Matthew Adams, Drissa Coulibaly, Kirsten E Lyke, Andrea Berry, *Epitope-Specific Antibody Responses to a Plasmodium falciparum Subunit Vaccine Target in a Malaria-Endemic Population.* The Journal of Infectious Disease, 2020. **223** (11).
36. Henson, S.N., et al., *PepSeq: a fully in vitro platform for highly multiplexed serology using customizable DNA-barcoded peptide libraries.* Nat Protoc, 2023. **18**(2): p. 396-423.
37. Ladner, J.T., et al., *Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-reactivity with endemic human coronaviruses.* Cell Rep Med, 2021. **2**(1): p. 100189.
38. Kelley, E.J., et al., *Virome-wide detection of natural infection events and the associated antibody dynamics using longitudinal highly-multiplexed serology.* Nat Commun, 2023. **14**(1): p. 1783.
39. Kurup, S.P., N.S. Butler, and J.T. Harty, *T cell-mediated immunity to malaria.* Nat Rev Immunol, 2019. **19**(7): p. 457-471.
40. Epstein, J.E., et al., *Live attenuated malaria vaccine designed to protect through hepatic CD8(+) T cell immunity.* Science, 2011. **334**(6055): p. 475-80.
41. Schofield, L., et al., *Interferon-gamma inhibits the intrahepatocytic development of malaria parasites in vitro.* J Immunol, 1987. **139**(6): p. 2020-5.
42. C.A. Janeway, P.T., M. Walport *T Cell-Mediated Immunity*, in *Immunobiology: The Immune System in Health and Disease*. 2001, Garland Science: NCBI.
43. Tobery, T.W., et al., *A simple and efficient method for the monitoring of antigen-specific T cell responses using peptide pool arrays in a modified ELISpot assay.* J Immunol Methods, 2001. **254**(1-2): p. 59-66.
44. Ioannidis, L.J., et al., *High-dimensional mass cytometry identifies T cell and B cell signatures predicting reduced risk of Plasmodium vivax malaria.* JCI Insight, 2021. **6**(14).
45. Carralot, J.P., et al., *Mass spectrometric identification of an HLA-A*0201 epitope from Plasmodium falciparum MSP-1.* Int Immunol, 2008. **20**(11): p. 1451-6.
46. Costa, R.M., et al., *Immunoproteomic analysis of Plasmodium falciparum antigens using sera from patients with clinical history of imported malaria.* Malar J, 2013. **12**: p. 100.
47. Seder, R.A. and R. Ahmed, *Similarities and differences in CD4+ and CD8+ effector and memory T cell generation.* Nat Immunol, 2003. **4**(9): p. 835-42.
48. E. Girmanova, P.H., O. Vicklicky, A. Slavcev, *ELISpot assay and prediction of organ transplant rejection.* International Journal of Immunogenetics, 2021. **49**(1).
49. Lyke, K.E., et al., *HLA-A2 supertype-restricted cell-mediated immunity by peripheral blood mononuclear cells derived from Malian children with severe or uncomplicated Plasmodium falciparum malaria and healthy controls.* Infect Immun, 2005. **73**(9): p. 5799-808.
50. Bruder, J.T., et al., *Profiling the Targets of Protective CD8(+) T Cell Responses to Infection.* Mol Ther Methods Clin Dev, 2017. **7**: p. 20-31.
51. M. Sedegah, Y.K., B. Peters, S. McGrath, H. Ganeshan, J. Lejano, E. Abot, G. Banania, M. Belmonte, R. Sayo, F. Farooq, D.L. Doolan, M.R. Hollingdale, T.L. Richie, A. Sette, *Identification and localization of minimal MHC-restricted CD8+ T cell epitopes within the Plasmodium falciparum AMA1 protein.* Malaria Journal 2010.
52. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.* Nucleic Acids Res, 2020. **48**(W1): p. W449-W454.
53. Kramer, S., et al., *An ultra-high-throughput screen for the evaluation of peptide HLA-Binder interactions.* Sci Rep, 2023. **13**(1): p. 5290.

54. S. Zhao, L.Z., S. Xiang, Y. Hu, Z. Wu, J. Shen, *Gnawing Between Cells and Cells in the Immune System: Friend or Foe? A Review of Trogocytosis.* Frontiers in Immunology 2022.
55. Li, G., et al., *T cell antigen discovery via trogocytosis.* Nat Methods, 2019. **16**(2): p. 183-190.
56. Nakayama, M., et al., *Shaping of T Cell Functions by Trogocytosis.* Cells, 2021. **10**(5).
57. Kula, T., et al., *T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes.* Cell, 2019. **178**(4): p. 1016-1028 e13.
58. Igonet, S., et al., *Cross-reactivity studies of an anti-Plasmodium vivax apical membrane antigen 1 monoclonal antibody: binding and structural characterisation.* J Mol Biol, 2007. **366**(5): p. 1523-37.
59. Gonzalez-Galarza, F.F., et al., *A snapshot of human leukocyte antigen (HLA) diversity using data from the Allele Frequency Net Database.* Hum Immunol, 2021. **82**(7): p. 496-504.
60. Gilbert, P., Self, S., Roa, M., Naficy A., Clemens, J. , *Sieve analysis methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation.* Journal of Clinical Epidemiology 2001. **54**(1): p. 68 - 85.
61. Yang, G., L.B. Balzer, and D. Benkeser, *Causal inference methods for vaccine sieve analysis with effect modification.* Stat Med, 2022. **41**(8): p. 1513-1524.
62. Benkeser, D., M. Juraska, and P.B. Gilbert, *Assessing trends in vaccine efficacy by pathogen genetic distance.* J Soc Fr Statistique (2009), 2020. **161**(1): p. 164-175.
63. Dai, J.Y., S.S. Li, and P.B. Gilbert, *Case-only method for cause-specific hazards models with application to assessing differential vaccine efficacy by viral and host genetics.* Biostatistics, 2014. **15**(1): p. 196-203.
64. Edlefsen, P.T., et al., *Comprehensive sieve analysis of breakthrough HIV-1 sequences in the RV144 vaccine efficacy trial.* PLoS Comput Biol, 2015. **11**(2): p. e1003973.
65. Amin, A.B., et al., *Monovalent Rotavirus Vaccine Efficacy Against Different Rotavirus Genotypes: A Pooled Analysis of Phase II and III Trial Data.* Clin Infect Dis, 2023. **76**(3): p. e1150-e1156.
66. Juraska, M., et al., *Prevention efficacy of the broadly neutralizing antibody VRC01 depends on HIV-1 envelope sequence features.* Proc Natl Acad Sci U S A, 2024. **121**(4): p. e2308942121.
67. Magaret, C., et al., *Quantifying how single dose Ad26.COV2.S vaccine efficacy depends on Spike sequence features.* Res Sq, 2023.
68. Cao, L., et al., *Rapid evaluation of COVID-19 vaccine effectiveness against symptomatic infection with SARS-CoV-2 variants by analysis of genetic distance.* Nat Med, 2022. **28**(8): p. 1715-1722.
69. Ouattara, A., et al., *Epitope-based sieve analysis of Plasmodium falciparum sequences from a FMP2.1/AS02(A) vaccine trial is consistent with differential vaccine efficacy against immunologically relevant AMA1 variants.* Vaccine, 2020. **38**(35): p. 5700-5706.
70. Ouattara, A., et al., *An In Silico Analysis of Malaria Pre-Erythrocytic-Stage Antigens Interpreting Worldwide Genetic Data to Suggest Vaccine Candidate Variants and Epitopes.* Microorganisms, 2022. **10**(6).
71. Ajibola, O., et al., *In silico characterisation of putative Plasmodium falciparum vaccine candidates in African malaria populations.* Sci Rep, 2021. **11**(1): p. 16215.
72. Patel, P., et al., *Genetic diversity and antibody responses against Plasmodium falciparum vaccine candidate genes from Chhattisgarh, Central India: Implication for vaccine development.* PLoS One, 2017. **12**(8): p. e0182674.
73. Gandhi, K., et al., *Variation in the circumsporozoite protein of Plasmodium falciparum: vaccine development implications.* PLoS One, 2014. **9**(7): p. e101783.
74. Almeida-de-Oliveira, N.K., et al., *Balancing selection and high genetic diversity of Plasmodium vivax circumsporozoite central region in parasites from Brazilian Amazon and Rio de Janeiro Atlantic Forest.* PLoS One, 2020. **15**(11): p. e0241426.

75. Chong, E.T.J., et al., *Genetic diversity of circumsporozoite protein in Plasmodium knowlesi isolates from Malaysian Borneo and Peninsular Malaysia.* Malar J, 2020. **19**(1): p. 377.
76. Zeeshan, M., et al., *Genetic variation in the Plasmodium falciparum circumsporozoite protein in India and its relevance to RTS,S malaria vaccine.* PLoS One, 2012. **7**(8): p. e43430.
77. Conway, D.J., et al., *High recombination rate in natural populations of Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4506-11.
78. Su, X., et al., *A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum.* Science, 1999. **286**(5443): p. 1351-3.
79. Miles, A., et al., *Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum.* Genome Res, 2016. **26**(9): p. 1288-99.
80. Richie, T.L., et al., *Sporozoite immunization: innovative translational science to support the fight against malaria.* Expert Rev Vaccines, 2023. **22**(1): p. 964-1007.
81. Dieng, M.M., et al., *Integrative genomic analysis reveals mechanisms of immune evasion in P. falciparum malaria.* Nat Commun, 2020. **11**(1): p. 5093.

Figure Legends

Figure 1. Sieve effect associated with vaccination. A) Participants in a malaria vaccine trial are either immunized with a “red genotype” or receive a placebo, such as saline solution. B) Vaccine trial participants are exposed to vector transmitting a genotypically diverse parasite population, represented by multiple colors. C) The frequency of different variants differs in infections from vaccinees and controls, due to the sieve effect associated with the immune system, shown as immune cells on a sieve, which was trained through vaccination to recognize a “red” genotype and others immunologically similar. Therefore, “red” and “deep red” genotypes are absent from malaria isolates collected from vaccinees. Figure prepared with BioRender.

Figure 2. **Schematic of a whole-genome sieve analysis (wgSA) study**. a) Participants in a field trial to study efficacy of a whole-organism vaccine receive a vaccine (n_V) or a placebo (n_P), according to study arm. b) During the observation period, participants are challenged by natural infection and while some are never infected (i vaccinees and k controls), others become

infected once or more during this time (N_V vaccinees and N_P placebo recipients). c) A blood sample, corresponding to first infection post immunization, is obtained from each infected participant and (c) processed, to obtain whole genomic DNA. e) If leukocytes were not removed during blood collection, total DNA can be subjected to selective (parasite) whole genome amplification to enrich the sample for the desired DNA. f) A library is prepared for each sample, and libraries are multiplexed for high-throughput sequencing. g) wgSA is conducted on whole genome sequencing data to identify significantly differentiated sites, genome-wide, between infections from vaccinees and placebo recipients, in which the vaccine allele is depleted in vaccinee samples. Figure prepared with BioRender.

4.3 Protective targets of PfSPZ-based whole organism vaccines identified from genome-wide sieve analyses of isolates from field efficacy trials.

Manuscript in preparation. Planned submission to Journal of Translational Medicine.

Protective targets of PfSPZ vaccines identified from whole-genome sieve analyses of isolates from malaria vaccine efficacy trials

Ankit Dwivedi[1,2], Ryan J. Scalsky[1], Thomas C. Stabler[3], Biraj Shrestha[4], Sudhaunshu Joshi[4], David G. Harris[5], Chakshu Gandhi[1], James B. Munro[1], Olukemi O. Ifeonu[1], Alphonse Ouedraogo[6], Alfred B. Tiono[6], Drissa Coulibaly[7], Amed Ouattara[4], Thomas L. Richie[8], B. Kim Lee Sim[8], Christopher V. Plowe[4], Kirsten E. Lyke[4], Shannon Takala-Harrison[4], Stephen L. Hoffman[8], Mahamadou A. Thera[7], Sodiomon B. Sirima[6], Matthew B. Laurens[4], Joana C. Silva[1,2,9*]

[1] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
[2] Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA
[3] Swiss Tropical Public Health Institute, and University of Basel, Basel, Switzerland
[4] Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA
[5] Department of Computer Science, University of Maryland College Park, MD, USA
[6] Groupe de Recherche Action en Santé, Ouagadougou, Burkina Faso
[7] Malaria Research and Training Center, University of Sciences, Techniques and Technologies, Bamako, Mali
[8] Sanaria Inc., Rockville, Maryland, USA
[9] Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade NOVA de Lisboa (NOVA), Lisbon, Portugal

***Corresponding Author:** Joana C. Silva (jcsilva@som.umaryland.edu)

Abstract (250 words)

Background. Identification of protective antigens is a central quest in malaria vaccinology. Whole-genome sieve analysis (wgSA) in samples collected from placebo-controlled field trials of *Plasmodium falciparum* (Pf) sporozoite (SPZ) whole organism vaccines may enable their identification.

Methods. We applied wgSA to parasite genomic data generated from Pf parasites collected during two field trials measuring the efficacy in malaria-exposed African adults of two PfSPZ vaccines. These randomized, double-blind, placebo-controlled trials were conducted in regions of Mali and in Burkina Faso characterized by high seasonal transmission, where parasite genetic diversity is high. Genomic sites in which the vaccine allelic state was significantly underrepresented among breakthrough infections in vaccinees relative to placebo recipients were termed “target sites”. Protein-coding loci containing target sites that changed amino acids were termed “target loci”.

Findings. The wgSAs of genomic data from clinical trial samples collected in Burkina Faso and Mali resulted in the identification of 165 and 93 target loci, respectively, with 18 loci present in both data sets (expectation, $E = 5{\cdot}0$, with 99% confidence interval between [0, 10]). Among all target loci, 138 and 80 were single-copy protein-coding genes in the Burkinabe and Malian data sets, respectively, with twelve common to both ($E = 3{\cdot}9$; 99%CI = [0, 9]). Among the target genes common to both studies was the thrombospondin-related anonymous protein locus, which encodes PfSSP2/TRAP, one of the most well-characterized pre-erythrocytic stage antigens.

Interpretation. wgSA is a potentially powerful tool for the identification of protective vaccine antigens in recombining pathogens with large genome size.

Funding. This work was supported by the National Institutes of Health through grants R01AI141900, U01AI112367, U19AI110820, and HHSN272201300022.

Research in context.

Currently, most antigen discovery tools utilize immune responses to identify targets. Medium- and high-throughput approaches utilize anti-sera to identify targets in peptide or protein microarrays, and pools of peptides are used to identify targets of T cell responses. However, there has been little progress in utilizing an agnostic, high-throughput approach to identify protective targets of immunization. Whole-genome (or comprehensive) sieve analysis (wgSA) was applied previously to samples collected from placebo-controlled efficacy trials of vaccines against HIV-1 to discern potential protective mechanisms of vaccine-induced protection. The study identified target sites in vaccine immunogens (antibody or T cell epitopes) as well as some non-immunogenic targets that were possibly in linkage disequilibrium with protective sites. These studies were based on vaccines containing a known immunogen and were designed against a pathogen with a small genome (~15 kb) encoding a small number of well-

characterized protein-coding genes (~10 genes). The present study represents the first application of wgSA with the intent of identifying protective novel targets of immunization and its first application to samples from a vaccine trial against a eukaryotic parasite, with a much larger genome size (~23.3 Mb) than the previously studied viruses, and with a potentially large number (several hundreds to a few thousand) unknown immunogens. We previously applied a similar rationale to identify putative targets of natural acquired immunity to malaria. This promising approach requires the availability of a whole organism-based vaccine for a pathogen with high genetic diversity, to provide abundant genetic markers throughout the genome, and a high rate of effective recombination, such that those markers segregate independently. *Plasmodium falciparum*, the deadliest eukaryotic pathogen of humans and the most virulent agent of human malaria, satisfies these conditions.

Evidence before this study

In 2001, Gilbert and colleagues proposed an approach, which was coined “sieve analysis”, to determine the extent to which the infecting pathogen is susceptible to vaccine-induced protection, given its genotype. This is achieved by quantifying the relative risk of infection, post-vaccination, associated with pathogens encoding different variants of the vaccine immunogen. The authors applied sieve analysis to studies testing the efficacy of vaccines against viral and bacterial species, demonstrating broad applicability to pathogens despite vast taxonomic distances, as well as differences in etiology and protection mechanism. We and others applied this approach to genotype data of additional organisms including those that cause malaria, in the context of single antigen vaccine efficacy (VE) studies, to identify key antigenic residues involved in protection. Modifications to the initial sieve analysis have been introduced, aimed at testing hypotheses related to the mechanism of vaccine-induced protection, or host factors that contribute to modify VE. Literature searches on PubMed revealed no articles in which sieve analysis was applied to the entire genome as an agnostic tool that leverages allele-specific VE to identify targets of vaccine-induced protection. Earlier this year we wrote a review (Scalsky et al.; under review in *Vaccine*) in which we proposed that this might be a productive approach to identify novel targets of vaccine-induced protection in parasitic diseases.

Added value of this study

In addition to establishing whole-genome sieve analysis as a new approach to identify novel protective targets of whole organism-based immunization for pathogens of genome size in the order of megabases, this study identified novel potential protective antigens. These candidates should be confirmed with wgSA in independent data sets and validated experimentally to show that they are both immunogenic and protective.

Implications of all the available evidence

The results of this study support the often-stated hypothesis that multiple parasite proteins contribute to the protective immune responses induced by vaccination with PfSPZ vaccines. The results also indicate that, as expected for T-cell mediated immunity, these responses are, in many cases, allele/epitope-specific. Thus, one approach to optimizing the protection induced by PfSPZ vaccines would be to include multiple variants of protective antigens by (i) mixing complementary strains of Pf in the vaccine, (ii) creating hybrid parasites through crosses of

different strains of Pf, and/or (iii) expressing the sequences of multiple variants of key proteins in the same PfSPZ to overcome the antigenic diversity of *P. falciparum* strains circulating in malaria-endemic areas.

Background.

The World Health Organization (WHO) recommended the very first malaria vaccine, RTS,S/AS01, in October 2021, reflecting decades of research. The formulation of RTS,S/AS01 is based on the variant of *Plasmodium falciparum* (Pf) circumsporozoite protein (CSP) encoded by the 3D7 strain. In 2023, a second malaria vaccine, R21/Matrix-M, was endorsed by the WHO that is based on the same variant of CSP. While this momentous advance will reduce malaria morbidity and mortality, it is clear that neither vaccine will meet the current WHO strategic goal of achieving >90% efficacy against Pf infection.[1] Furthermore, RTS,S/AS01 has been shown to induce allele-specific protection[2] and is, therefore, susceptible to vaccine escape, highlighting the need for additional vaccines to protect against breakthrough infections.

Whole organism vaccines based on Pf sporozoites (PfSPZ) are in late stage of development,[3] and are highly promising as pre-erythrocytic candidate vaccines that can prevent malaria disease and interrupt malaria transmission by preventing parasites from ever leaving the liver. They possess highly desirable features,[4] including (i) the stimulation of human immunity to multiple Pf antigens present during the pre-erythrocytic phase of parasite development; ii) 100% protection against controlled human malaria infection (CHMI) with parasites homologous to the vaccine and >75% efficacy against CHMI with heterologous parasites; and (iii) durable efficacy without boosting in malaria-experienced adults in highly endemic settings in sub-Saharan Africa. These features represent an extremely high achievement for a malaria vaccine that must overcome a lifetime of immune exhaustion resulting from repeat malaria infection to stimulate a protective antimalarial human immune response. Efficacy in adults living in a high transmission setting with limited ability to respond strongly to a malaria vaccine suggests that PfSPZ vaccines may be able to achieve higher efficacy in pediatric populations with less preexisting immunity to malaria. Nevertheless, the generally lower protection against heterologous relative to homologous challenge is also consistent with genotype- (or allele)-specific efficacy.

Knowledge of Pf antigens that induce a protective response is required to advance malaria subunit vaccine development and could be used to enhance PfSPZ vaccines. New multivalent subunit vaccines or multi-strain PfSPZ vaccines, where the multiple components complement each other, may reduce or eliminate vaccine escape and provide broad, strain-transcending protection against malaria. As we recently outlined,[5] identification of putative protective targets may be achieved with a whole-genome sieve analysis (wgSA) performed on parasite isolates collected from participants in randomized, placebo-controlled field trials of a whole-organism vaccine, when several conditions are met: in particular, i) when trials are conducted in a geographic region of high disease transmission, such that genetic diversity in the parasite population is high; ii) when the pathogen has a high rate of effective recombination, such that genetic linkage is limited to only a few hundred base pairs; and (iii) vaccine efficacy (VE) is genotype-specific.

Two such clinical trials were completed in regions with intense, seasonal malaria transmission of sub-Saharan Africa. One assessed the efficacy of PfSPZ Vaccine, a radiation-attenuated whole sporozoite vaccine *versus* saline placebo in 80 adults living in Burkina Faso, West Africa, who were cleared of parasites pre-vaccination.[6] Efficacy against infection was determined by thick blood smear microscopy during illness and every 4 weeks for two malaria seasons (72 weeks total). VE (1 – hazard ratio) was 48% and 46% after 6 and 18 months of follow-up, respectively (p = 0·061 and 0·018). A second study tested PfSPZ-CVac (CQ), a chemoattenuated PfSPZ vaccine in which live, infectious PfSPZ are administered to participants who are receiving chloroquine to eliminate blood stage infection.[7] Participants were not cleared of parasites prior to immunization. VE in this study was assessed by thick blood smear during illness and every 4 weeks for a single malaria transmission season (24 weeks). VE compared to saline placebo in 62 participants in this study was 34% (p = 0·21). Both studies randomized participants 1:1 to receive either the malaria vaccine or placebo. The PfSPZ in both vaccines were from the PfNF54 isolate, the parent stock from which the reference 3D7 was cloned.[8] In protein-coding regions, PfNF54 and 3D7 differ in fewer than 100 non-synonymous sites, mostly in members of multigene families.[9]

Here, we report the outcome of wgSAs performed on Pf isolates collected from vaccinees and control subjects (placebo recipients) from these two vaccine clinical trials when they first became positive. We identified, in each trial, a set of parasite genes in which vaccine genotypes were significantly under-represented in infections from vaccinees relative to controls and prioritize genes identified in both studies. We describe these genes and show enrichment for features expected from protective antigens, consistent with identification of true protective candidates by wgSA.

Methods.

Samples

Samples originated from two clinical trials assessing the efficacy of malaria vaccines against natural infection in West Africa, one in Burkina Faso (NCT02663700) and one in Mali (NCT02996695). Samples analyzed correspond to the timing of the first thick blood smear-positive malaria infection occurring at least two weeks after the last immunization dose, corresponding to 33 and 40 participants, respectively, in the Burkinabe and Malian trials. Samples consisted of a 2 mL venous blood draw which was leukocyte-depleted, pelleted and stored at -20°C or -80°C. DNA was extracted from each sample, using the Qiagen Blood DNA Midi Kit (Valencia, CA, USA). Total DNA extracted varied between 8·0-1031·2 ηg (median: 75·9 ηg) for Burkinabe samples and between 0·0-78043·3 ηg (median: 10·3 ηg) for Mali samples. Nine out of 40 samples from Mali had unusually high DNA amounts, likely due to contamination. All samples with <100 ηg of total DNA or those with likely contamination underwent selective whole genome amplification (sWGA) with Pf-specific primers to increase the relative amount of parasite DNA. Three out of 40 Malian samples had insufficient DNA for library construction.

Sequencing and variant identification

DNA sequencing and sequence variant identification were done as before[9] and details are presented in the Appendix. For each sample, sequencing data was generated to reach a minimum of 75% breadth of coverage of the 3D7 genome with ≥5X depth of coverage. When that coverage level was unfeasible, due to a high proportion of human DNA, samples were removed from downstream analysis (**Appendix Tab1**). In addition to data from samples from the two clinical trials described, whole genome shotgun sequencing (WGS) data for 1,289 Pf isolates collected from other 20 malaria-endemic countries was used in joint SNP calling (**Appendix Tab2**).

Multiplicity of infection.
Multiplicity of infection was assessed with the statistic F_{WS}, estimated with the R package moimix.[10] Difference in F_{WS} difference between study arms was tested with a Wilcoxon Rank Sum test. The number of clones per infection was estimated with dEploid-IBD (v0.5)[11] with default parameters, and significance tested with a chi-square test.

Whole genome sieve analysis (wgSA).
A wgSA was conducted separately using WGS data generated from isolates collected in each clinical trial, and in each case was based on all polymorphic sites that passed filters, as described above. A wgSA consists of two steps: *i*) identification of parasite genomic sites significantly differentiated between study arms (vaccinees and controls), and (*ii*) identification of sites from (*i*) where the vaccine allele is depleted among vaccinee infections. For each genomic site, difference in allele frequency distribution between study arms (step *i*) was quantified using Wright's Fixation Index, F_{ST} (Weir and Cockerham's implementation in vcftools v0.1015[12]). For each analysis, F_{ST} value significance was determined by randomization of samples across study arms, with 5000 replicates. Sites found to be significantly differentiated between vaccinee and control infections, and in which the vaccine (PfNF54) allele was underrepresented in vaccinee infections, were termed "*target sites*". Deviation of expected frequency of target sites per study arm was tested with a chi-square test.

Loci were termed "*target loci*" if they contained non-synonymous target sites. The 99% confidence interval and p-value for the number of target loci found, by chance, in both wgSA (given *m* and *n* targets genes, one set in each study), given the total number of genes *g* with variable non-synonymous sites across studies, were both determined by Fisher's exact conditional test for 2x2 independence (fixed row and column sums, *m* and *n*), which is a conservative version of a multinomial (unconditional) test. The multinomial test is more appropriate (fixed *g* but variable number of target loci, *m* and *n*) but it is computationally expensive. The expected number of target loci that overlap between studies, *E*, is given by $E = m*n/g$.

Feature Identification and enrichment analyses
CD8+ T cell epitopes were predicted across target loci using netMHCpan (v4.1)[13] using default parameters. Only strong binding epitopes were considered in downstream analysis. The HLA alleles used for epitope prediction were the 22 most common among those identified in Mali and Burkina Faso.[14] Enrichment analyses of gene ontology (GO) terms among different gene sets

were done using ShinyGo (v0.741).[15] In target loci, enrichment of target sites in strong binding epitope-encoding regions was tested using a chi-square test.

Results.

Thirty-three participants (13 and 20 in the vaccine and placebo study arms, respectively) in the Burkinabe PfSPZ Vaccine trial were infected at least once with Pf during the follow up period starting two weeks after the last immunization. Total DNA was isolated for each participant's first infection. An average of 43 million Illumina reads were generated per sample (**Appendix Tab1**). On average, 18 million reads per sample mapped to the parasite genome (range: 65,600 – 30,890,811). Two samples, one each in the vaccine and placebo arms did not reach desired coverage threshold and were not used further. Complexity of infection, measured by number of clones per infection or by F_{WS}, did not differ significantly between study arms (**Appendix F1)**. In the Malian PfSPZ-CVac (CQ) trial, 40 participants (17 vaccinees and 23 who received a placebo) experienced one or more Pf infections during follow up. An average of 60·5 million sequence reads were obtained per sample, and 23 million mapped to the parasite genome (range: 85,469 – 43,912,950) (**Appendix Tab1**). Four samples from vaccinees and five from placebo recipients did not pass the coverage threshold; 13 and 18 isolates from vaccinees and placebo recipients, respectively, were used in downstream analyses. Once more complexity of infection did not differ significantly between study arms (**Appendix F2)**. A total of 31,122 and 31,659 SNPs in the Pf nuclear genome passed quality filters, respectively, for the Burkinabe and Malian samples (**Appendix Tab3**) and were used for downstream analyses.

Since the efficacy of PfSPZ Vaccine is partly genotype-specific,[16] we conducted a whole-genome sieve analysis (wgSA) on Pf isolates from each trial, to identify putative protective targets. A wgSA, by revealing genomic sites in which the vaccine allele (here, the allele in PfNF54) is significantly underrepresented among vaccinee infections, identifies "putative protective sites", or "target sites".[5,17] Also, we term "target loci" (encoding putative protective proteins) the protein-coding genes that contain non-synonymous target sites.

Among the samples from the Burkinabe PfSPZ Vaccine trial, 684 significantly differentiated SNPs were identified between vaccinees and controls, 508 of which were target sites (**Table 1**). Of those, 202 were non-synonymous SNPs, which mapped to 165 protein-coding loci or "target loci" (**Table 1; Appendix Tab4**). These loci included several that encode well-established sporozoite antigens, most notably PF3D7_0304600, which encodes CSP, and PF3D7_1335900, which encodes the thrombospondin-related anonymous protein (PfSSP2/TRAP). In Mali, 314 target sites were identified. Among those were 110 non-synonymous target sites, mapping to 93 target loci (**Table 1; Appendix Tab5**), including sporozoite antigens PfSSP2/TRAP and E140 (encoded by PF3D7_0104100), among others. The target sites identified in each wgSA were distributed across all 14 chromosomes (**Figure 1**).

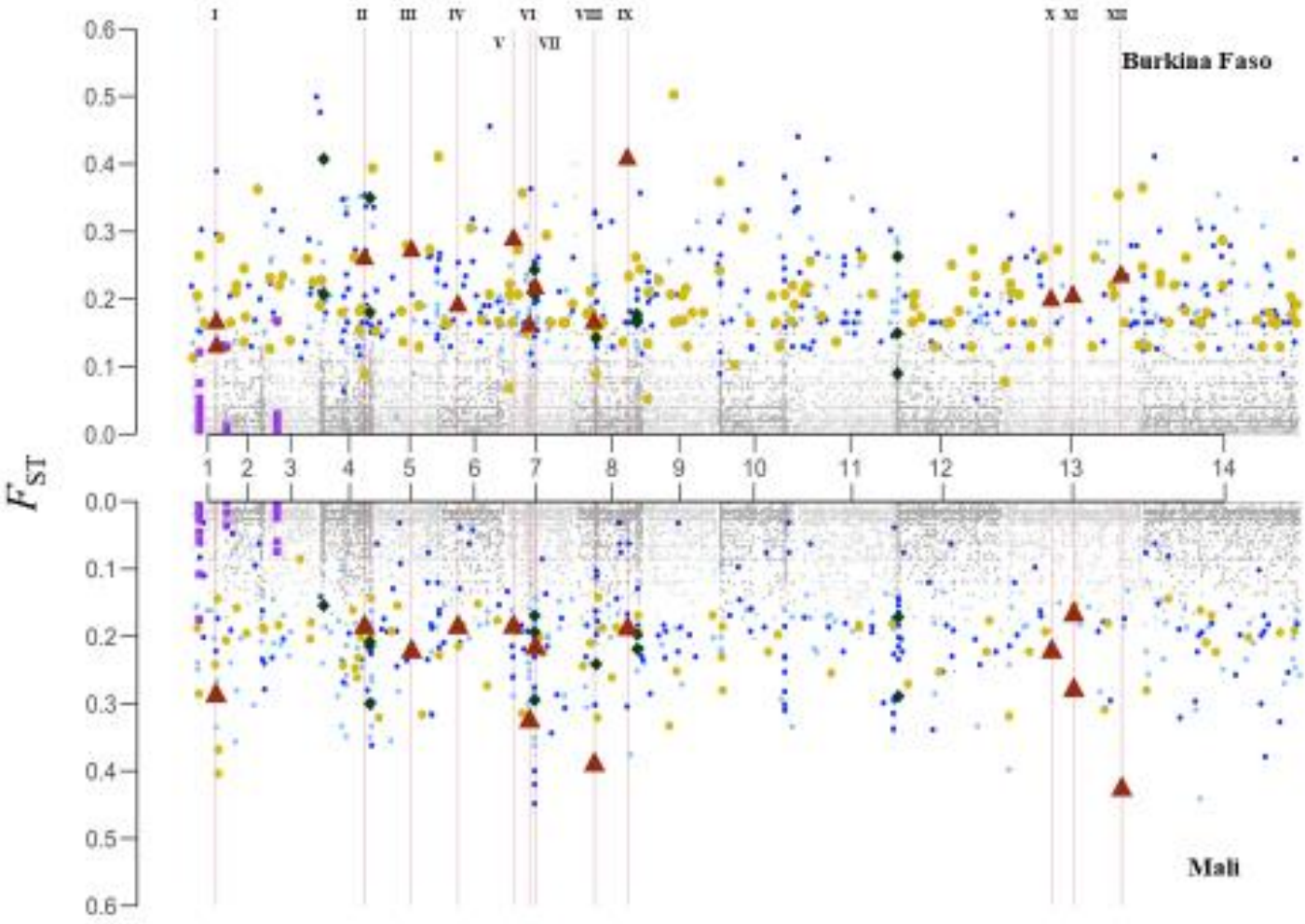


Figure 1. **Whole-genome sieve analysis of P. falciparum isolates from field trials of two PfSPZ vaccines, in Burkina Faso and in Mali.** Site-wise FST values between placebo and vaccinee parasites are shown for Burkina Faso and Mali vaccine trials, separately. Significantly differentiated genomic sites (permutation method; k=5000; p-value<0□05) are shown for each trial (non-grey), including sites where the vaccine allele was underrepresented in isolates in controls (light blue) and those where the vaccine allele was underrepresented among vaccinee isolates (target sites; dark blue, and non-blue). Non-synonymous target sites (tan) are further highlighted they were observed in the same target locus in both studies, and the locus is a member of a multigene family (dark green diamond); and when they are observed in same single-copy target locus in both studies (brown triangle). The latter (n=12; shown in roman numerals) are shown across the genome (red, vertical line): PF3D7_0113800 (I), PF3D7_0421700 (II), PF3D7_0518700 (III), PF3D7_0609600 (IV), PF3D7_0703900 (V), PF3D7_0711200 (VI), PF3D7_0713600 (VII), PF3D7_0808100 (VIII), PF3D7_0826000 (IX), PF3D7_1324300 (X), PF3D7_1335900 (XI), PF3D7_1361800 (XII). Variable sites in E140 (in chromosome 1), LSAP2 (in chromosome 2) and CSP (in chromosome 3) are shown (purple). In each of these three loci, one non-synonymous target site is present in one study but not the other.

Several factors can lead to potential false positive signals in wgSA, i.e., polymorphic sites identified as “target sites” which are not protective,[5] including differential genotype assortment into study arms by chance, a common challenge for studies with relatively small sample size, such as the two studies here, as well as genomic sites in linkage disequilibrium with protective sites.[17] In addition to contributing to false positives, these factors can also explain the presence of significantly underrepresented vaccine alleles in placebo recipients (**Table 1**, blue columns; **Figure 1**, blue dots). To determine if these wgSAs were likely to have revealed functionally

protective proteins, we sought to determine whether the target sites and associated target loci displayed patterns expected from antigens. Several lines of evidence suggest this to be the case, as follows.

Analyses of the 262 significantly differentiated non-synonymous sites in the Burkinabe wgSA revealed that the vaccine allele was depleted significantly more among samples from vaccine (n=202; 77·1%) *versus* placebo recipients (n=60; 22·9%) (**Table 1**; **Appendix Tab6**; $p<0{\cdot}001$), as would be expected from the genotype-specific sieve effect of a vaccine-primed immune system. In the Malian study, the vaccine allele was again significantly more depleted among samples from vaccine (61·5%) *versus* placebo recipients (38.5%) (**Table 1**; **Appendix Tab6**; $p<0{\cdot}001$). Additionally, a GO term enrichment analysis among 165 target loci identified with the Burkinabe clinical trial samples revealed a significant enrichment of membrane-associated proteins and proteins involved in host-parasite interactions, traits common to antigens (**Appendix F3**). A similar analysis of the 93 target loci identified in the Malian study, in which VE was not significant, did not reveal GO term enrichment, and neither did enrichment analyses on the loci with significantly differentiated non-synonymous SNPs in which the vaccine allele was depleted in controls, in either study (**Appendix F3**).

Eighteen loci were present among target loci in both studies (**Table 2**), a set larger than expected by chance ($E = 5{\cdot}0$; p-value ~ $3{\cdot}7\text{x}10^{-7}$), given the number of loci (n = 3326; (**Appendix Tab3**) with non-synonymous variable SNPs across studies (**Appendix Tab3)**. Among these 18 target loci, twelve were single copy genes and six were members of multigene families. Since wgSA false positives could arise from read mapping artifacts among members of multigene families, we focused further on single copy genes. Given 138 and 80 single copy target loci in the Burkinabe and Malian studies, respectively, and a total of 2848 single copy genes with non-synonymous variable sites, the probability of finding twelve single copy target loci in common between studies is also significantly higher than expected ($E = 3{\cdot}9$; p-value ~ $3{\cdot}6\text{x}10^{-4}$). Finally, as expected for targets of pre-erythrocytic vaccines, transcripts for all twelve target loci are present in sporozoites, and again during the liver stage, although one is only detected by day 6 (**Table 2**).[18,19]

The current vaccine development pipeline for malaria includes twelve pre-erythrocytic antigens, including CSP (PF3D7_0304600) and PfSSP2/TRAP (PF3D7_1335900).[20] We found 138 and 135 variable sites among these twelve antigens in Burkinabe and Malian sample sets, respectively, but only a small subset of those are target sites (**Appendix Tab4** and **Tab5**). Among Burkinabe samples, three non-synonymous target sites mapped to three of those antigens, one each in LSAP2 (PF3D7_0202100), CSP, and TRAP/PfSSP2 (**Figure 1**), and the target sites in CSP and PfSSP2/TRAP fall within predicted CD8+ T cell epitopes recognized by the 22 most prevalent HLA alleles in Burkina Faso and Mali. Among Malian samples, two target sites were identified among these twelve vaccine antigens, one of which maps to a predicted CD8+ T cell epitopes in PfSSP2/TRAP (**Figure 2**).

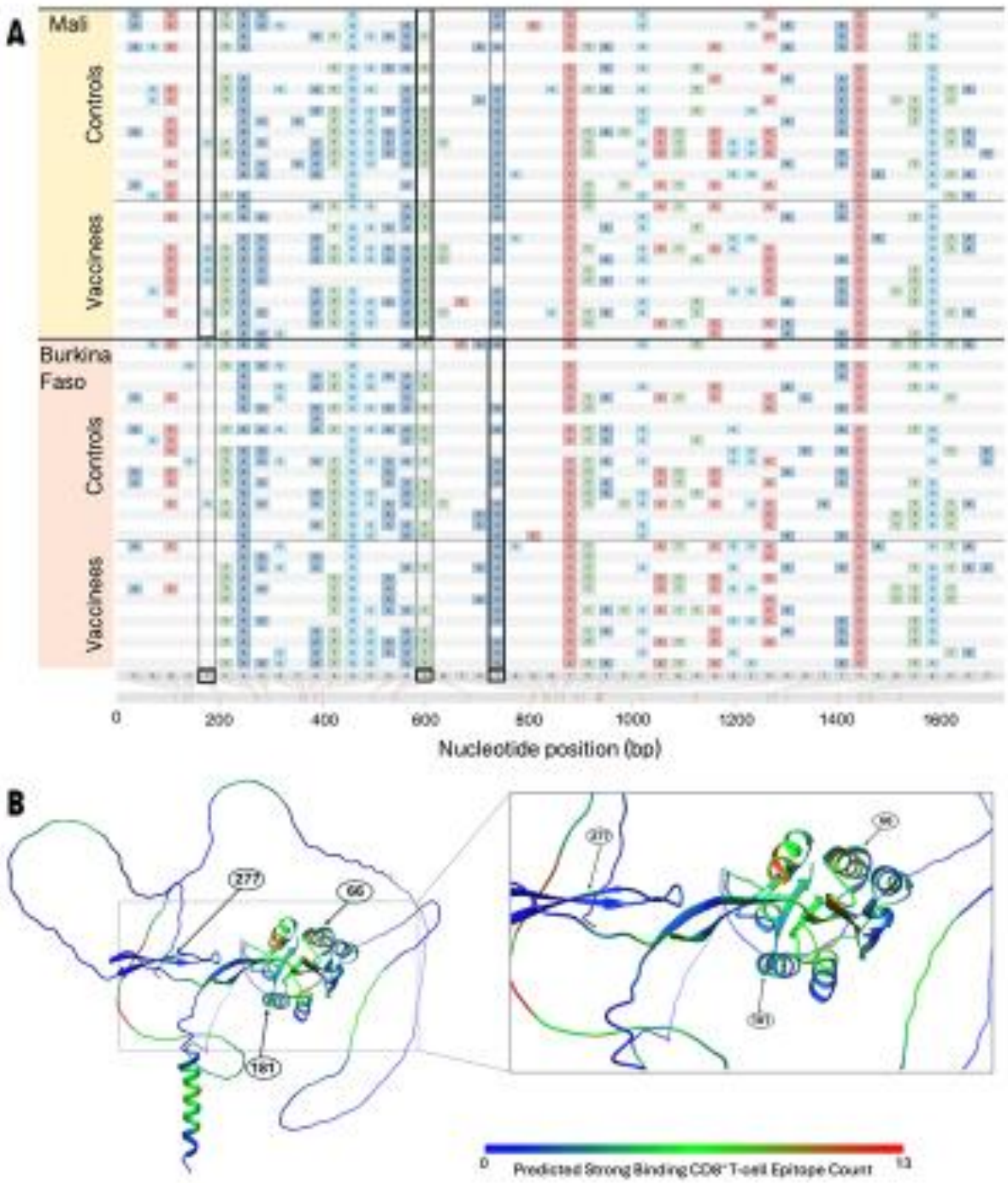


Figure 2. **Genetic variation and distribution of predicted CD8+ T-cell epitopes at PF3D7_1335900, the locus encoding PfSSP2/TRAP.** A) Variable genomic sites in the coding sequence of PF3D7_1335900 among clinical isolates (rows, alternating in white, and pink) from vaccinees and controls, from each clinical trial. 3D7 was used as reference sequence and its allelic state at each variable is shown below the alignment (grey). In sequences for each

clinical isolate (rows), nucleotides (shown in different colors) denote differences from the 3D7 allelic state; if identical to reference allele, no nucleotide is shown. Target sites identified by wgSA in at least one study are shown (solid box; dotted box in study where site is not a target site). Target sites in the Malian study are not significantly differentiated between vaccinee and control infections in the Burkina Faso trial samples, and vice-versa. Alignment visualization generated with SNIPIT (https://github.com/aineniamh/snipit). B) Protein data bank files for the 3D structure of TRAP, obtained from AlphaFold (https://alphafold.ebi.ac.uk/). Amino acid residues containing target sites are noted, with residue coordinate (L66I; R181L; L277I). Color spectrum shows to the count, from zero (blue) to 13 (red), of predicted strong binding CD8+ T-cell epitopes overlapping each residue. Display generated with UCSF Chimera v1.15.

Discussion

In 2001, Gilbert and colleagues[21] proposed an approach they termed “sieve analysis” to determine if VE is allele-specific, and to quantify the relative risk of infection, post-vaccination, associated with strains encoding different variants of a vaccine immunogen. Our group conducted the first sieve analysis in malaria, to identify residues associated with protection post-immunization with Pf apical membrane antigen 1 (AMA1).[22] We previously proposed that extension of the sieve analysis concept to the entire genome of a highly variable, recombining pathogen species such as Pf, using samples from a placebo-controlled vaccine trial, could identifying genomic sites in which the vaccine allele is significantly under-represented in vaccinees relative to placebo recipients, thereby identifying targets of vaccine-induced protection.[5] Here, we conducted two wgSA on Pf isolates collected from VE trials of PfSPZ vaccines conducted in malaria-endemic areas, with the explicit purpose of identifying potential targets of PfSPZ vaccine-induced protection. The wgSA conducted on the Burkinabe study isolates revealed 165 target loci (138 of which are single-copy genes) while the Malian wgSA revealed 93 target loci (80 single-copy genes). A total of 18 target loci were identified in both analyses, including twelve single-copy genes.

Several lines of evidence are consistent with the possibility that these twelve target loci are targets of PfSPZ vaccine induced protection. Most notably, PF3D7_1335900 encodes one of the most well-established pre-erythrocytic sporozoite stage antigens, PfSSP2/TRAP.[23] Also, PF3D7_0113800 encodes a DBL domain known to activate Rho GTPases involved in hepatocyte invasion during Pf infection,[24,25] and PF3D7_1361800 encodes the glideosome-associated connector (GAC), a crucial component of the gliding machinery that enables host tissue transversal, cell invasion and egress.[26] GAC connects actin fibers and surface adhesins (such as TRAP), playing a pivotal role in the forward movement of the parasite. Interestingly, GAC becomes exposed and is shed as parasites migrate through host tissue,[26] providing an opportunity for immune system priming as well as subsequent parasite detection. PF3D7_0703900 and PF3D7_1324300 encode conserved, membrane-associated proteins of unknown function, but may be accessible to either cell-mediated or humoral immunity if present on the sporozoite surface. The relationship between some genes identified by the wgSA and vaccine-induced protection is less clear. PF3D7_0518700 encodes the mRNA-binding protein PUF1, likely regulating stability and translation efficiency of specific mRNAs important for gametocyte development.[27] PF3D7_0808100 encodes AP-3 complex subunit delta, with a role in endosomal protein trafficking and vesicle formation within the parasite.[28,29] PF3D7_0713600 encodes a putative mitochondrial ribosomal protein S5. The wgSA revealed four additional parasite genes encoding conserved proteins of uncharacterized function (**Table 2**). One of

these, PF3D7_0421700, has been identified as a likely target of allele-specific naturally-acquired immunity to malaria,[30] suggesting an overlap between protective targets of sporozoite-based vaccination and natural infection. Finally, within these 12 target loci, most non-synonymous target sites are located within predicted CD8+ T cell epitopes, consistent with cell-mediated immunity (**Table 2; Appendix Tab7**).

The two wgSAs yielded different numbers of target loci, which partially overlapped between studies. Differences can result from several factors, including pre-treatment (or lack thereof), the vaccine tested and VE outcome. The Burkinabe clinical trial tested three doses of $2.7x10^6$ PfSPZ of PfSPZ Vaccine, which consists of radiation attenuated, non-replicating PfSPZ that arrest early in the liver stage. The Malian trial assessed three doses of $2x10^5$ PfSPZ of PfSPZ-CVac (CQ), which consists of non-attenuated, fully replicating PfSPZ that are eliminated by chloroquine after rupture of liver schizonts when the parasites reach the blood.[4] Therefore, different antigen sets are presented by the two vaccines, and for different periods of time, likely inducing responses to different, and only partly overlapping, sets of targets. Patients were treated to clear blood stage parasitemia pre-immunization in the Burkinabe study, but not in Mali. It is unclear if an ongoing infection with a genotype different from PfNF54 could have impacted protection outcome in a genotype-specific manner, hence contributing to an increase in noise to signal ratio, with a larger proportion of the differentially distributed genotypes being due to chance in the Mali study. In addition, the two studies differed in protection outcome, with the Burkinabe trial (but not the Malian trial) resulting in significant VE. Lack of significant VE likely also increases noise to signal ratio. This is consistent with the lower signal observed in enrichment analyses in Malian relative to Burkinabe samples, including a smaller proportion of differentially distributed non-synonymous sites in which the vaccine allele was underrepresented among vaccinee samples relative to controls, and a lack of GO term enrichment among target loci. Therefore, the studies almost certainly differ in power to detect an association between genetic variants and protection. Finally, differences between host populations, including in HLA allele frequencies, previous malaria exposure, as well as previous exposure to different Pf genotypes could all have led to differences in the target sites detected.

Overall, these results are consistent with wgSA's utility as an agnostic approach to identify novel potential targets of whole-organism-based vaccines. Analysis of additional PfSPZ-based VE trials in malaria-endemic areas can be used to further explore this approach, potentially strengthening identification and prioritization of candidate loci. Importantly, it is essential to validate these results for any targets intended for vaccine formulations, by demonstrating that they are both immunogenic and protective. The evidence presented suggests that multiple parasite proteins contribute to the protective immune response induced by vaccination with PfSPZ vaccines, and that these protective immune responses are often allele-specific, as is the case, by definition, of the target loci identified here. Therefore, to achieve strain-transcendent, broad protection, PfSPZ vaccines will need to induce optimal protective immune responses to multiple epitopes from multiple proteins, including subdominant epitopes, and could potentially be improved by including mixtures of PfSPZ, genetic crosses of parasites from different regions, and/or expression by PfSPZ of multiple variants of known protective epitopes from multiple strains.

References

1. WHO. Malaria vaccines: preferred product characteristics and clinical development considerations. Geneva, 2022.
2. Neafsey DE, Juraska M, Bedford T, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med* 2015; **373**(21): 2025-37.
3. World Health Organization = Organisation mondiale de la S. WHO Guidelines for malaria, 25 November 2022. Geneva: World Health Organization; 2022.
4. Richie TL, Church LWP, Murshedkar T, et al. Sporozoite immunization: innovative translational science to support the fight against malaria. *Expert Rev Vaccines* 2023; **22**(1): 964-1007.
5. Scalsky R, Dwivedi A, Stabler T, et al. Whole-genome sieve analysis, wgSA: identification of protective antigens by leveraging allele-specific vaccine efficacy. *Vaccine* 2024; **(Submitted)**.
6. Sirima SB, Ouedraogo A, Tiono AB, et al. A randomized controlled trial showing safety and efficacy of a whole sporozoite vaccine against endemic malaria. *Sci Transl Med* 2022; **14**(674): eabj3776.
7. Coulibaly D, Kone AK, Traore K, et al. PfSPZ-CVac malaria vaccine demonstrates safety among malaria-experienced adults: A randomized, controlled phase 1 trial. *EClinicalMedicine* 2022; **52**: 101579.
8. Walliker D, Quakyi IA, Wellems TE, et al. Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science* 1987; **236**(4809): 1661-6.
9. Moser KA, Drabek EF, Dwivedi A, et al. Strains used in whole organism Plasmodium falciparum vaccine trials differ in genome structure, sequence, and immunogenic potential. *Genome Med* 2020; **12**(1): 6.
10. Lee SA, Harrison AG, Tessier N, et al. Assessing clonality in malaria parasites using massively parallel sequencing. 2016. http://bahlolab.github.io/moimix/.
11. Zhu SJ, Hendry JA, Almagro-Garcia J, et al. The origins and relatedness structure of mixed infections vary with local prevalence of P. falciparum malaria. *Elife* 2019; **8**.
12. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 1984; **38**(6): 1358-70.
13. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020; **48**(W1): W449-W54.
14. Gonzalez-Galarza FF, McCabe A, Santos E, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 2020; **48**(D1): D783-D8.
15. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 2020; **36**(8): 2628-9.
16. Epstein JE, Paolino KM, Richie TL, et al. Protection against Plasmodium falciparum malaria by PfSPZ Vaccine. *JCI Insight* 2017; **2**(1): e89154.
17. Edlefsen PT, Rolland M, Hertz T, et al. Comprehensive sieve analysis of breakthrough HIV-1 sequences in the RV144 vaccine efficacy trial. *PLoS Comput Biol* 2015; **11**(2): e1003973.
18. Zanghi G, Patel H, Camargo N, et al. Global gene expression of human malaria parasite liver stages throughout intrahepatocytic development. *bioRxiv* 2023.
19. Real E, Howick VM, Dahalan FA, et al. A single-cell atlas of Plasmodium falciparum transmission through the mosquito. *Nat Commun* 2021; **12**(1): 3196.

20. Bettencourt P. Current Challenges in the Identification of Pre-Erythrocytic Malaria Vaccine Candidate Antigens. *Front Immunol* 2020; **11**: 190.
21. Gilbert P, Self S, Rao M, Naficy A, Clemens J. Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. *J Clin Epidemiol* 2001; **54**(1): 68-85.
22. Ouattara A, Takala-Harrison S, Thera MA, et al. Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications. *J Infect Dis* 2013; **207**(3): 511-9.
23. Ogwang C, Kimani D, Edwards NJ, et al. Prime-boost vaccination with chimpanzee adenovirus and modified vaccinia Ankara encoding TRAP provides partial protection against Plasmodium falciparum infection in Kenyan adults. *Sci Transl Med* 2015; **7**(286): 286re5.
24. Aghazadeh B, Zhu K, Kubiseski TJ, et al. Structure and mutagenesis of the Dbl homology domain. *Nat Struct Biol* 1998; **5**(12): 1098-107.
25. Schepis A, Kumar S, Kappe SHI. Malaria parasites harness Rho GTPase signaling and host cell membrane ruffling for productive invasion of hepatocytes. *Cell Rep* 2023; **42**(1): 111927.
26. Jacot D, Tosetti N, Pires I, et al. An Apicomplexan Actin-Binding Protein Serves as a Connector and Lipid Sensor to Coordinate Motility and Invasion. *Cell Host Microbe* 2016; **20**(6): 731-43.
27. Shrestha S, Li X, Ning G, Miao J, Cui L. The RNA-binding protein Puf1 functions in the maintenance of gametocytes in Plasmodium falciparum. *J Cell Sci* 2016; **129**(16): 3144-52.
28. Tomavo S, Slomianny C, Meissner M, Carruthers VB. Protein trafficking through the endosomal system prepares intracellular parasites for a home invasion. *PLoS Pathog* 2013; **9**(10): e1003629.
29. Nishimura N, Plutner H, Hahn K, Balch WE. The delta subunit of AP-3 is required for efficient transport of VSV-G from the trans-Golgi network to the cell surface. *Proc Natl Acad Sci U S A* 2002; **99**(10): 6755-60.
30. Shah Z, Naung MT, Moser KA, et al. Whole-genome analysis of Malawian Plasmodium falciparum isolates identifies possible targets of allele-specific immunity to clinical malaria. *PLoS Genet* 2021; **17**(5): e1009576.

ICMJE Authorship Statement

All authors meet ICMJE authorship criteria. T.L. Richie, B.K.L.Sim, and S.L. Hoffman are salaried employees of Sanaria Inc., the developer and owner of PfSPZ Vaccine and PfSPZ CVac. In addition, S.L. Hoffman and B.K.L. Sim have a financial interest in Sanaria Inc. No other authors have a competing interest.

Author Contributions

Study Design: JCS, AD, STH

Data Collection: BS, SJ, AO, ABT, DC, MAT, SBS

Data Analysis: AD, RJS, TCS, DGH, CG, JBM, OOI

Writing: JCS, MBL, AD, RJS

Editing: All

Author Signatures

Tables

Table 1. Classification of significantly differentiated sites between samples from vaccinees and placebo-recipients. The country and vaccine used in each case are mentioned in the header, for each study. Significantly differentiated sites in which the vaccine allele was underrepresented among vaccinee samples are termed "target sites" (dark blue columns); the loci to which non-synonymous (NSYN) target sites (tan) map are termed "target loci" (brown).

	Burkina Faso, PfSPZ Vaccine trial		Mali, PfSPZ CVac trial	
	PfNF54 allele depleted in vaccinee samples (target sites and target loci)	PfNF54 allele depleted in placebo recipient samples	PfNF54 allele depleted in vaccinee samples (target sites and target loci)	PfNF54 allele depleted in placebo recipient samples
Significantly differentiated sites	508	176	314	196
Synonymous (SYN)	94	34	47	47
Non-Synonymous (NSYN)	202	60	110	69
Other (intergenic, intronic, etc.)	212	82	157	80
Protein-coding loci with differentiated NSYN sites	165	48	93	58

Table 2. Target loci identified in both studies. The 18 target loci are shown, each containing at least one non-synonymous (NSYN) target site (i.e., significantly differentiated NSYN SNPs between samples from vaccinees and controls, with the vaccine allele depleted in the former) in both the studies and associated gene expression. Single-copy genes (not shaded) and members of multigene families (shaded rows) are shown.

Gene ID	Gene product	Protein length (AA)	No. NSYN target sites (BF / M)[a]	No. NSYN target sites in CD8+ T cell epitopes (BF / M)[b]	SPZ expression[c]	Liver expression[d]
PF3D7_0113800	DBL-containing protein	2900	2 / 1	2 / 1	+	- - +
PF3D7_0421700	Conserved Plasmodium protein	1936	1 / 1	1 / 1	+	+ + +
PF3D7_0518700	mRNA-binding protein PUF1	1894	1 / 1	1 / 1	+	+ + +
PF3D7_0609600	Conserved Plasmodium protein	1257	1 / 1	1 / 1	+	- + +

PF3D7_0703900	Conserved Plasmodium membrane protein	4358	2 / 1	2 / 1	+	+ + +
PF3D7_0711200	Conserved Plasmodium protein	1000	1 / 1	0 / 1	+	- + +
PF3D7_0713600	Ribosomal protein S5, mitochondrial, putative	917	1 / 1	0 / 1	+	+ + +
PF3D7_0808100	AP-3 Complex subunit Delta	1609	1 / 1	1 / 1	+	- + +
PF3D7_0826000	Conserved Plasmodium protein	1172	1 / 1	1 / 1	+	- + +
PF3D7_1324300	Conserved Plasmodium membrane protein	5415	1 / 1	1 / 0	+	+ + +
PF3D7_1335900	Thrombospondin-related anonymous protein, TRAP[e]	574	1 / 2	1 / 1	+	+ + +
PF3D7_1361800	Glideosome-associated connector, GAC	2605	1 / 1	1 / 0	+	- + +
PF3D7_0402200	surface-associated interspersed protein 4.1 (SURFIN 4.1), pseudogene	2217	4 / 1	4 / 1	+	- - -
PF3D7_0424400	surface-associated interspersed protein 4.2 (SURFIN 4.2)	2380	2 / 3	1 / 2	+	- - +
PF3D7_0713000	rifin	346	6 / 4	5 / 4	+	NA
PF3D7_0808800	rifin	388	1 / 1	1 / 1	+	- - -
PF3D7_0830800	surface-associated interspersed protein 8.2 (SURFIN 8.2)	2049	6 / 2	4 / 1	-	- + +
PF3D7_1200600	erythrocyte membrane protein 1, PfEMP1	3056	3 / 2	2 / 2	+	- - +

[a] Number of non-synonymous (NSYN) target sites in a locus, in each of the two studies; BF: Burkina Faso; M: Mali. [b] NSYN target sites in both studies, overlapping with predicted CD8+ T-cell epitopes corresponding to 22 most prevalent HLA alleles in Burkina Faso (BF) and in Mali (M). [c] Presence of transcripts in the sporozoite stage was classified dichotomously (presence/absence). [d] Liver stage expression is classified into five groups based on CPM values[18] (ubiquitously expressed: +++; present by day 4: -++; present by day 6 --+; not expressed in liver: --- ; and information not available: NA). [e] TRAP (but none of the other gene products) has a signal peptide.

Figure legends

Figure 1. Whole-genome sieve analysis of *P. falciparum* isolates from field trials of two PfSPZ vaccines, in Burkina Faso and in Mali. Site-wise F_{ST} values between placebo and vaccinee parasites are shown for Burkina Faso and Mali vaccine trials, separately. Significantly differentiated genomic sites (permutation method; k=5000; *p-value*<0·05) are shown for each trial (non-grey), including sites where the vaccine allele was underrepresented in isolates in controls (light blue) and those where the vaccine allele was underrepresented among vaccinee isolates (target sites; dark blue, and non-blue). Non-synonymous target sites (tan) are further highlighted they were observed in the same target locus in both studies, and the locus is a member of a multigene family (dark green diamond); and when they are observed in same single-copy target locus in both studies (brown triangle). The latter (n=12; shown in roman numerals) are shown across the genome (red, vertical line): PF3D7_0113800 (I), PF3D7_0421700 (II), PF3D7_0518700 (III), PF3D7_0609600 (IV), PF3D7_0703900 (V), PF3D7_0711200 (VI), PF3D7_0713600 (VII), PF3D7_0808100 (VIII), PF3D7_0826000 (IX),

PF3D7_1324300 (X), PF3D7_1335900 (XI), PF3D7_1361800 (XII). Variable sites in E140 (in chromosome 1), LSAP2 (in chromosome 2) and CSP (in chromosome 3) are shown (purple). In each of these three loci, one non-synonymous target site is present in one study but not the other.

Figure 2. Genetic variation and distribution of predicted CD8⁺ T-cell epitopes at PF3D7_1335900, the locus encoding PfSSP2/TRAP. A) Variable genomic sites in the coding sequence of PF3D7_1335900 among clinical isolates (rows, alternating in white, pink) from vaccinees and controls, from each clinical trial. 3D7 was used as reference sequence and its allelic state at each variable is shown below the alignment (grey). In sequences for each clinical isolate (rows), nucleotides (shown in different colors) denote differences from the 3D7 allelic state; if identical to reference allele, no nucleotide is shown. Target sites identified by wgSA in at least one study are shown (solid box; dotted box in study where site is not a target site). Target sites in the Malian study are not significantly differentiated between vaccinee and control infections in the Burkina Faso trial samples, and vice-versa. Alignment visualization generated with SNIPIT (https://github.com/aineniamh/snipit). **B)** Protein data bank files for the 3D structure of TRAP, obtained from AlphaFold (https://alphafold.ebi.ac.uk/). Amino acid residues containing target sites are noted, with residue coordinate (L66I; R181L; L277I). Color spectrum shows to the count, from zero (blue) to 13 (red), of predicted strong binding CD8⁺ T-cell epitopes overlapping each residue. Display generated with UCSF Chimera v1.15.

5 DISCUSSION

5.1 Placing molecular monitoring and next-generation whole genome sequencing into malaria control activities in the field

Historically, malaria has caused significant loss of human life, to the point of impacting the evolution of our collective DNA, selecting for certain genomic traits (Kariuki and Williams, 2020). Haemoglobinopathies, or mutations leading to abnormal hemoglobin structure or function, are believed to have increased in frequency through natural selection in malaria-affected human populations, as these variants provide a degree of protection from malaria death (Taylor et al., 2012, Weatherall, 2008). In recent decades, advancements in malaria control and treatment have reduced malaria-related morbidity and mortality. In many malaria-endemic countries the lowest socioeconomic groups now bear the highest burden (Degarege et al., 2019). Malaria remains a significant contributor to death for children under 5-years in low-middle income countries (LMIC), particularly in sub-Saharan Africa, despite significant investment in global funding for malaria control and drive towards eventual elimination (Cohen et al., 2022, WHO, 2023). As NMPs make progress, it is becoming increasingly challenging to achieve further reduction in malaria prevalence. Parasites reservoirs become harder to identify, and continuous *P. falciparum* and its vectors adapt rapidly to interventions, and competition with other global health priorities limits funding for malaria control (Cohen et al., 2022). With the progression of global warming and high population connectedness through national and international travel, malaria expansion or re-introduction into previously malaria-free regions becomes a serious possibility (Li and Managi, 2022, Fischer et al., 2020). Most recently, in 2023, autochthonous *P. vivax* cases were detected in Texas and Florida, and an autochthonous *P. falciparum* case in Maryland was confirmed by RDT, microscopy and qPCR (Blackburn, 2023, Duwell et al., 2023). These US cases were acquired locally, in each state, among individuals with no reported malaria risk, such as recent travel, or contact with a recent traveler, from a malaria-endemic region. For the malaria-free, or low-transmission regions, any interruption of malaria control, or surveillance activities could lead to severe spikes in malaria mortality, if unchecked, as population-level immunity wanes over time, leading to an increase in susceptibility to malaria disease and death (Ghani et al., 2009). One of the challenges is generating detailed, accurate data and acting quickly enough to make evidence-based decisions in malaria control policy. Decentralization of NMPs to a provincial-level approach to control implementation (Gosling et al., 2020) or decentralizing data generation to laboratories in malaria-endemic regions have been suggested as faster, more economic paths to prevent malaria expansion (Tessema et al., 2019).

With the presented doctoral work, I aimed to advance knowledge of *P. falciparum* as it relates to malaria diagnostics, transmission dynamics and vaccine development. Underlying the presented work is the additional effort to improve accessibility of the appropriate molecular techniques, when possible – and in this context, within collaborating laboratories in Equatorial Guinea. As malaria endemic countries experience higher malaria burden (WHO, 2023), the faster NMPs can monitor *Plasmodium* transmission with advanced surveillance techniques, the more effective control and prevention interventions can be deployed in response.

In this PhD thesis, a significant proportion of the molecular data generated originated from Bioko Island, Equatorial Guinea. All results assessing ultrasensitive RDTs were generated on Bioko Island (Mpina et al., 2022). The first ever application of the Diaxxo PlasmoPod device in Equatorial Guinea was conducted at the Baney District Reference Laboratory on Bioko Island by the local staff (Stabler et al., Manuscript in preparation-b). All the presented epidemiological data and DBS samples used were collected and processed by local teams. Ideally, the entire process of NGS data generation and downstream analysis will be completed locally. Hindering this goal in LMICs is the lack of infrastructure for computational science, NGS and bioinformatics, as there are few opportunities to learn and apply these concepts. The development of the GC_3 approach was meant to facilitate the generation of already existing NGS coverage files and subsequent interpretation of results by non-NGS specialists, with limited bioinformatics experience (Stabler et al., 2022). Some knowledge of UNIX command lines or Python/R is required to operate GC_3; however, publically available, detailed documentation provide instructions for the required input files and to facilitate running the program locally, on a personal laptop. GC_3's function semi-automates the processing of coverage files and automatically outputs informative visuals and databases with commonly used software (e.g., excel). Generally, as molecular methods and NGS technologies become more accessible, their integration into NMP's core approach to malaria control strategy, as recommended by the WHO (Carter et al., 2022), becomes more feasible.

Plasmodium spp. NGS can be challenging, in part, due to the need to account for low parasite quantities in relation to human/host DNA, which limits downstream analyses (Oyola et al., 2013). Exacerbating NGS issues is AT-rich composition of some *Plasmodium* genomes, which leads to biased representation of genomic segments during sequence library construction (Lan et al., 2015). The application of hybrid selection, digestion or amplification techniques can reduce the effect of these biases (Melnikov et al., 2011, Oyola et al., 2016, Oyola et al., 2013). As part of my work, we have successfully applied an in-house DNA extraction method, developed by the Malaria Research Program at the University of Maryland School of Medicine, to field and clinical samples, and generated high-quality data - mostly through the application of sWGA techniques (Oyola et al., 2013, Shah et al., 2020, Stabler et al., Manuscript in preparation-a). Further, we also applied this extraction method in Equatorial Guinea (Stabler

et al., Manuscript in preparation-b). Comparatively, whole blood samples have a much higher DNA yield than DBS (Holzschuh and Koepfli, 2022). However, in the context of active surveillance of *Plasmodium* and transmission dynamics, DBS (i.e. blood drops on filter paper) are an ideal field sample as they remain stable for a significant amount of time and do not require cold chain logistics (Färnert et al., 1999, Bereczky et al., 2005). An active surveillance approach, based on field samples, is best to obtain reliable parasite prevalence estimates, as passive surveillance can underestimate malaria prevalence, especially among malaria immune adults (Zhou et al., 2015, Tiono et al., 2014, malERA, 2017). Ultimately, DNA yields from DBS are lower than whole blood extractions; however, we and others have successfully demonstrated that high-quality WGS can be generated from DBS and this approach was used to describe for the first time the *P. falciparum* population on Bioko Island (Stabler et al., Manuscript in preparation-a). These WGS results provide a holistic representation of circulating malaria strains through accurate analysis of molecular data.

Standard library preparation protocols are biased against AT-rich genomic regions, and care needs to be had with genomes such as that of *P. falciparum* (Lan et al., 2015); advancements in library preparation and Illumina sequencing platforms have improved sequencing results (Le Roch et al., 2012, Van Vliet, 2010). Within the presented work, both long- and short-read sequencing data was used. Ideally, both long- and short-read data are generated for each sample to generate the highest quality coverage and read data (Rhie et al., 2021). However, this typically requires a much greater investment, and is not always feasible given limited resources. Bioinformatic methods can minimize sequence errors by mapping raw reads to a reference genome, like *P. falciparum* 3D7 (Pf3D7), using GATK HaplotypeCaller and bowtie2, which uses the reference genomes as a template to assign descriptive quality values to each called SNP (Van der Auwera et al., 2013, Langmead and Salzberg, 2012). In this way, low-quality *Plasmodium* data can be excluded, such as a sample with low target coverage, reads mapping to multiple loci, or SNPs with low quality scores. Polyclonal infections add another complication for molecular monitoring of *Plasmodium* as these infections can mask the genotypes of lower-frequency strains, particularly important when assessing *hrp2* deletions (Watson et al., 2019, Hosch et al., 2022). Bias from polyclonal infections can be mitigated by either calling major alleles or by deconvolution. When calling major alleles, a distinct, predefined threshold is assigned (e.g. 70%) and at each genomic site only the most frequent allele is assigned, disregarding low frequency alleles. Whereas deconvolution includes differentiating polyclonal isolates by the allele frequency, and assigning algorithmic scores to then reconstruct each strain separately (Zhu et al., 2017). Both of these methods have been used to handle polyclonal infections in downstream analyses, to reduce noise and provide accurate genetic variation estimates (Ruybal-Pesántez et al., 2024, Neafsey et al., 2021). This extends to my doctoral work where all sequence data was processed accordingly to account for polyclonal infections (Stabler et al., 2022, Stabler et al., Manuscript in preparation-a).

During the progress of this doctoral work, short-read WGS samples were generated for *P. falciparum* from Bioko Island for the first time. WGS is becoming the gold standard for pathogen epidemiology, allowing for simultaneous analysis of multiple genes, which other sequencing methods lack (Auburn and Barry, 2017, Akoniyon et al., 2022). Targeted sequencing (amplicon sequencing), does not have the versatility of WGS, but is a more affordable, informative method to sequence genomic segments of interest (Satam et al., 2023). WGS has been used to measure *Plasmodium* population structure (Shen et al., 2018, Hupalo et al., 2016), identify structural variants (Obaldia et al., 2014, Osborne et al., 2021), monitor for drug resistance (Cravo et al., 2015) and advance malaria vaccine development (Conway, 2015). Further, WGS provides the best method for resolving linkage disequilibrium (LD) (Pengelly et al., 2015). LD refers to a non-random association between SNPs or segments of DNA that may be conserved solely due to their genomic proximity to a selected SNP or DNA segment (Lewontin, 1988). Within the *Plasmodium* genome, the length of DNA segments in LD can vary between 1.5 kbp (~0.1 cM) and 17 kbp (~1 cM) (Volkman et al., 2007), which requires “pruning” to accurately conduct gene association studies (Martin and Chung, 2021). Especially in the context of my thesis work, the first description of a, so far, uncharacterized population, WGS provided appropriate data to best understand diversity, structure and transmission dynamics (Stabler et al., Manuscript in preparation-a). Although there is a higher up-front cost (in the order of USD $75-$150 per sample), the return on investment from WGS may be higher than other NGS methods due to the versatility, public availability in open access databases in combination with the volume of the data generated.

Despite decreasing costs, accessibility to WGS technology remains difficult for LMIC as significant investment in electrical infrastructure, powerful computational systems and laboratories is required (Adebamowo et al., 2018, Neafsey et al., 2021). Opportunities and experience for local staff is also needed to appropriately apply previously described bioinformatics methods and NGS data processing at the local level. Although the ONT sequencing technology was not used in the presented doctoral work, there are available units at the Baney reference laboratory in Equatorial Guinea. Due to its increased accessibility and ability to generate WGS for *P. falciparum* (Meulenaere et al., 2024), ONT may be an improved sequencing technology for WGS generation from field samples, but further validation is required. In summary, NGS and WGS can offer high-throughput sequence data that provide high-quality, detailed data applicable within a variety of various malaria research areas due to its unbiased analysis and coverage of the entire genome.

5.2 Adapting RDT diagnostics to identify *P. falciparum* parasites that escape routine surveillance

Our reliance on malaria RDTs to accurately diagnose *Plasmodium* spp. infections is apparent, in part due to their wide-spread use, low cost and user-friendly design. However, an issue with RDTs is their relatively high limit of parasite detection (Berzosa et al., 2018, Laban et al., 2015), leading to difficulty identifying low-parasite density infections (Bousema et al., 2014). From a controlled human malaria infection (CHMI), we analyzed whole blood samples from 48 individuals using various diagnostic methods, including an ultrasensitive, HRP2-based RDT (uRDT). With a lower limit of detection than conventional RDTs, uRDTs had a similar sensitivity to microscopy (Mpina et al., 2022); however, the higher cost of uRDTs, compared to conventional RDT's very low price point, may prevent them from being more readily adopted by NMPs at present (Wittenauer et al., 2022). The uRDT's performance with field samples has also shown its usability as a diagnostic tool in Malaria Indicator Surveys (MIS) and field diagnostic studies (Yimam et al., 2022, Danwang et al., 2021), and may prove useful for surveys conducted in low-transmissions areas. An apparent drawback to uRDTs is the continued use of HRP2 as their target antigen, which is not robust to *hrp2* and *hrp3* (*hrp*2/3) deletion-carrying strains, particularly in multi-clonal infections.

Understanding diagnostic escape via *hrp2* and/or *hrp3* deletions poses a complex question that remains partially unaddressed in the field. Analysis of WGS data with $GC_{3,}$ and other studies of *hrp2* and *hrp3*, all have documented that partial gene deletions can be identified (Stabler et al., 2022, Gamboa et al., 2010, Dharia et al., 2010b, Molina - de la Fuente et al., 2022). Many studies only consider complete deletions or lack of a HRP2 antigen (Poti et al., 2020). This neglects the effect of a partial deletion which may also affect the HRP2 and HRP3 protein structure and function to a point where RDT diagnosis is compromised. Interestingly, a decrease in HRP2 or HRP3 function seem to not significantly impact *P. falciparum* viability and transmissibility in the field (Sepúlveda et al., 2018, Poti et al., 2020). Therefore, the inclusion of monitoring for partial *hrp2* and *hrp3* gene deletion strains is warranted in diagnostic evasion monitoring as much as complete deletions. However, the proportion of a partial deletion affecting function, the location of a partial deletion, and the role they may play in diagnostic evasion of malaria RDTs remains to be determined.

Monitoring for *hrp2* and *hrp3* deletions is complicated by the presence of polyclonal infections. Within a polyclonal infection, strains with *hrp2* deletion may be masked by the presence of another strain with an intact *hrp2* gene (Hosch et al., 2022). When analyzing 19 samples with known and unknown *hrp2* and *hrp3* deletion status, a qPCR assay was used that can also account for polyclonal infections based on the difference of Cq values between a control gene (*Pfrnr2e2*) and *hrp2* and *hrp3* (Schindler et al., 2019, Stabler et al., 2022). Using this method,

deletions can be detected among polyclonal infections, however partial deletions may be more difficult to identify. A complement of this *hrp*2/3 detecting assay is applying deconvolution, where NGS data and GC_3 can be used to also assess for *hrp2* and *hrp3* deletions among reconstructed strains in polyclonal infections. Of the 90 WGS samples generated in Bioko, none were identified to have *hrp2* or *hrp3* deletions by GC_3. Although no deletions were found, strains with *hrp2/3* deletions have been observed in Equatorial Guinea, including Bioko Island (Hosch et al., 2022, Berzosa et al., 2020, Molina-de la Fuente et al., 2024). Further, PlasmoPod assessment of the 1,479 samples from the 2023 MIS show a proportion of samples as false negative by RDT despite high parasite density, which may suggest the presence of *hrp2* deletions. However, the ability to identify polyclonal infections using 18S RT-qPCR with PlasmoPod remains to be developed in the context of this equipment.

Whereas GC_3 can be used to facilitate analysis of sequence coverage data, qPCR techniques are the most sensitive and appropriate method for monitoring *hrp2/3* deletions. Clearly, polyclonal infections complicate monitoring efforts, and as such, molecular techniques need to account for such infections when monitoring for the frequency of deletions, or risk underestimating the presence of RDT diagnostic evading strains. Neglecting to monitor for *hrp2/3* deletions may result in the selection and expansion of these strains as observed in Eritrea and Ethiopia (Mihreteab et al., 2021, Golassa et al., 2020). Especially in Eritrea, HRP2-based RDTs are significantly compromised and the NMP requires non-HRP2 RDTs to be used (WHO, 2023). Further hindering the Eritrean NMP control strategy is the observation of strains with both ART-resistance markers and *hrp2*/*hrp3*-deletions deletions (Mihreteab et al., 2023). The importance of proper surveillance of *Plasmodium* populations and appropriate responses to evolutionary adaptations cannot be understated.

As evident by the *hrp2*-specific qPCR assay (Schindler et al., 2019), molecular assays can be successfully designed to detect various targets. This versatility in molecular monitoring assays further support their incorporation into NMP malaria control strategies. The design of the PlasmoPod device aims to simplify and increase accessibility to such molecular assays while maintaining sensitivity and specificity comparable to more traditional qPCR techniques (Bechtold et al., 2023). On a larger scale, 1,500 blood samples collected from 1,500 volunteers were used to generated molecular results during the month of February 2024, with activities ranging from extracting DNA from DBS to running RT-qPCR, all while conducting the associated training of laboratory staff. In the context of this thesis, PlasmoPod was used to identify low-density infections using an 18S RT-qPCR assay, and can detect low-density or suspected *hrp2*-deletion infections. Although PlasmoPod is yet to substitute the ease and use of RDTs in the field, PlasmoPod offers a unique alternative that can generate results within 30 minutes of running qPCR. Its application as an easily implementable molecular method for Plasmodium diagnosis or surveillance is apparent.

The work presented here illustrates how NGS and novel technology tools (both software and hardware) can be used to improve monitoring of *Plasmodium*. Ideally in the near future, the sample collection and data generation will be followed entirely conducted at the local level in malaria endemic regions, cutting down on costs and shortening the turnaround needed for processing of samples (Golumbeanu et al., 2023). Advancements in malaria diagnostic methods and their affordability may allow for NMPs to better understand transmission dynamics and case tracking. Ultrasensitive RDTs may be a more appropriate option for field studies in low-transmission areas as they provide a similar level of sensitivity to that of microscopy without investment in laboratory specialization. However, their more widespread use is hampered by the cost and the lack of a different diagnostic antigen apart from HRP2. Using GC_3 can facilitate analysis of publicly available NGS data, as access to the nucleic material of samples is not always feasible. This can in turn be used as another method of estimating diagnostic escape if NGS data is available. Further, alternative diagnostic tools, like the 18S RT-qPCR assay with PlasmoPod, can offer relief in areas where the reliability of RDTs has become significantly compromised due to RDT escaping parasite populations.

5.3 Measuring connectedness and transmission dynamics to improve malaria interventions

Understanding transmission dynamics of *Plasmodium* spp. depends on various factors including mosquito biting behavior (Nzioki et al., 2023, Epopa et al., 2019), selective pressure and parasite fitness (Rosenthal, 2013, Lefevre et al., 2018, Schneider, 2021), and population genetics (Neafsey et al., 2021, Arnott et al., 2012, de Oliveira et al., 2020). In the context of Bioko Island, annual MIS data provides a granular epidemiological view of malaria transmission on the island beginning on 2004 (Cook et al., 2018), repeatedly noting higher odds of malaria infection among travelers to mainland Equatorial Guinea (Guerra et al., 2019, Bradley et al., 2015, García et al., 2023, Citron et al., 2021). As part of my doctoral work, the *P. falciparum* population structure was described using WGS data whole and the impact of importation on Bioko island was analyzed and linked to available epidemiological data. A significant proportion of infections on Bioko were polyclonal infections, a proxy measurement of transmission intensity (Galinsky et al., 2015), and evidence of moderate levels of transmission (Nchama et al., 2021, Gatton and Cheng, 2010, Lopez and Koepfli, 2021). The relatedness between Bioko, Cameroon and Gabon illustrated the connectedness between malaria on Bioko Island and its mainland neighbors. Further evidence of how importation may be preventing sustainable elimination efforts. Interestingly, the Bioko parasite population appeared slightly more closely related to the Gabonese population than Cameroonian, possibly hinting at stronger connectedness between mainland Equatorial Guinea and Gabon that extends to Bioko (Stabler et al., Manuscript in preparation-a). The predominant tribal group

in Equatorial Guinea and Gabon are the Fang, which represented 85% of Equatoguineans and 25% of Gabonese (Stokes, 2009, CIA, 2024b). Fang populations exist in Cameroon, but do not make up as significant proportion of the inhabitants (CIA, 2024a). Although speculative, the shared language and cultural links through Fang tribes present in both Equatorial Guinea and Gabon may promote increased border crossing and trade. In similar geographically isolated contexts, such as Zanzibar, São Tomé and Príncipe and Senegal, human migration (and the concomitant malaria importation) appears to be a significant obstacle to malaria elimination as island parasites have a significant proportion of their genome in IBD with parasites from the mainland (Fakih et al., 2023, Morgan et al., 2020, Wang et al., 2022, Daniels et al., 2020). Applying genetic data in conjunction with epidemiological data demonstrates the capability of detailing local and imported malaria transmission pathways, including those characterized in Equatorial Guinea. This first time population genetics analysis of malaria on Bioko Island suggests screening and treatment of travelers is likely needed to make continued progress towards malaria elimination.

Sustained surveillance of *Plasmodium* transmission is necessary as progress towards elimination continues. Malaria burden can change rapidly in response to effective malaria control (Cook et al., 2018, Kigozi et al., 2020). An active monitoring approach is considered advantageous, especially in regions nearing elimination (Nasir et al., 2020). Active surveillance on Bioko Island has demonstrated malaria prevalence is typically higher in rural communities than urban communities (García et al., 2023, Cook et al., 2018). However, based on population genetics analyses of Bioko *P. falciparum* isolates, parasites in Malabo, the capital city, appear to have more segments in IBD than rural parasites on average, suggestive of a more panmitic population (Salla et al., 2020, Ajogbasile et al., 2021). This is further supported by the higher frequency of polyclonal infections among individuals from Malabo, where more polyclonal infections are observed in regions with higher transmission intensity (Gatton and Cheng, 2010, Lopez and Koepfli, 2021). The presence of polyclonal infections is not unexpected as prevalence measurements have not decreased to a level where monoclonal infections would predominate (Lopez and Koepfli, 2021). Bioko Island has not yet reached a pre-elimination phase of malaria; however, once it does, evidence of low transmission in rural communities prior to urban communities may be expected, due to lower connectivity from rural communities to the mainland via Malabo (Stabler et al., Manuscript in preparation-a). It is clear that malaria spread is not constrained by geopolitical boundaries, either between countries or communities, and targeted control efforts at border crossing with high human traffic may be needed to prevent the import or export of malaria.

Merging epidemiological data and molecular data in real time offers valuable detail toward identifying reservoirs of low parasite-density infections. Low parasite-density reservoirs can contribute to local transmission while being hard to detect in time (Okell et al., 2012, Slater et

al., 2019, Whittaker et al., 2021). Low-density, asymptomatic infections appear often in adults (Heinemann et al., 2020, Topazian et al., 2020), as adults have more robust naturally acquired immunity against malaria and suppress infections better than younger age groups (Buchwald et al., 2018). Omitting malaria surveillance of adults potentially misses a key source of transmission. Using PlasmoPod, we showed that older age groups had higher association to malaria infection measured by RT-qPCR (Stabler et al., Manuscript in preparation-b). In the context of Bioko Island, the transmission dynamics between urban and rural communities differ (García et al., 2023). If RDTs are used as the diagnostic method during MIS, an underestimation of malaria may occur in urban environments, which are often associated with low density infections on Bioko (Stabler et al., Manuscript in preparation-b). In urban environments, there is higher population density, higher frequency of human migration and greater opportunities for mosquitos to transmit malaria (Olusegun-Joseph et al., 2019, Santos-Vega et al., 2016, Tatem et al., 2008). However, in Malabo, where most urban communities in Bioko Island are located, the population also has much greater access to health clinics and antimalarial treatment free of charge (Rehman et al., 2013). In areas with more robust health care systems, symptomatic infections are more likely to be diagnosed, treated and eliminated (Björkman and Morris, 2020), thus lower density infections may have greater chances of survival in urban areas like Malabo. Ultimately, as shown by our work on Bioko Island, utilizing both epidemiological and molecular data illuminates those under-detected malaria subpopulations present in the human population and which groups may be hidden transmission reservoirs only detectable at a molecular level.

5.4 Developing the next generation of malaria vaccines with improved protective efficacy

With the recent WHO approvals of the RTS,S/AS01 and R21/Matrix-M malaria vaccines, their widespread use is expected to alleviate malaria-related deaths, especially among children under 5 years old (Hammershaimb and Berry, 2024). Despite these advances, the need for other vaccine iterations with improved efficacy is pressing (Beeson et al., 2019). Efficacy measurements of approved malaria vaccines use clinical malaria as an endpoint (symptomatic malaria and diagnosis confirmation by microscopy), however their protection does not significantly prevent infection and subsequent transmission (Laurenson and Laurens, 2024, Lambach et al., 2024). Further, much like natural immunity, vaccine protection wanes over time, requiring regular boosting doses (Datoo et al., 2024, Chandramohan et al., 2021), which can complicate adherence to vaccine regimens especially in countries with weaker health systems (Deen and Clemens, 2006). Considering that RTS,S and R21 target a single gene product, PfCSP, and RTS,S elicited protection has been shown to be allele-specific (Neafsey et al., 2015), variants that evade vaccine-induced protection may minimize the positive impact

of those vaccines. It remains prudent to continue developing other malaria vaccine candidates, particularly focusing on other potential targets, applying novel techniques for discovery to create a vaccine capable of preventing malaria infection.

Plasmodium's complex life cycle and relatively high antigenic diversity necessitate a broader vaccine development approach as many gene products can fulfill redundant biological functions, especially those associated with erythrocyte attachment and invasion (Alves-Rosa et al., 2024). The merozoite surface protein 1 (*msp1*) of *P. falciparum* is one such gene whose product is thought of as a potential vaccine target due to its function as a facilitator of erythrocyte invasion of merozoites and high immunogenicity (Ferreira et al., 2003). The presented doctoral work reviewed the potential of *msp1* as a whole gene target (Thomson-Luque et al., 2024). Previous vaccine candidates explored MSP1 subunit vaccines, but did not elicit desired results (Tolle et al., 1993, Sheehy et al., 2012, Früh et al., 1991, Chauhan et al., 2010). The *msp1* gene is made up of several conserved, semi-conserved and highly polymorphic regions (Takala et al., 2002, Tanabe et al., 2007, Kiwanuka, 2009). Among these gene segments, the block 2 segment of *msp1* appears to be a primary immunological target among malaria-exposed individuals, and a potential vaccine target (Cavanagh et al., 2004, Cowan et al., 2011). However this segment is also highly polymorphic with tri-peptide repeats and accounts for most of the observed genetic variance between parasite populations (Thomson-Luque et al., 2024). The hypothesis for pursuing recombinant, full-length MSP1 as a vaccine candidate, despite previous suboptimal results for subunit vaccines, is that by using the full-length MSP1 product, a vaccine will induce broad immunity with various T-cell and B-cell epitopes, improving elicited, strain transcending protection (Thomson-Luque et al., 2024). This review compiles the previous clinical study outcomes and understand the genetic diversity and structure, and potential protective efficacy of full-length *msp1* for planned clinical trials in East and West Africa.

Of the approximately 5,500 genes in the *P. falciparum* genome, an estimated 40% are still annotated as putative proteins or proteins of unknown function (Gardner et al., 2002, Aurrecoechea et al., 2008). Although naturally acquired immunity is achievable through multiple episodes of sustained malaria exposure (Barry and Hansen, 2016), a direct correlate, or correlates, of protection is yet to be discovered. Available data are conflicting and the hypotheses of protective immune correlates differ on whether protection is mainly B-cell or T-cell mediated (Scalsky et al., Under review). Sieve analysis, the comparison of allele frequencies between infected vaccinated and infected unvaccinated individuals, has been used to describe allele specific immunity of PfCSP-based malaria vaccines (Neafsey et al., 2015). We hypothesized that protective antigens can be identified by way of differential frequency of gene variants in parasites collected from blood samples in controls (i.e. placebo recipients) and vaccinated individuals (Scalsky et al., Under review). We applied for the first

time in malaria a whole genome sieve analysis (wgSA) of *P. falciparum* in the context of an irradiation attenuated whole-sporozoite vaccine trial (i.e. PfSPZ Vaccine) and a non-attenuated form (PfSPZ-CVac). WGS data was generated from samples collected during two field trials in high seasonal transmission regions of Mali and Burkina Faso (Dwivedi et al., Manuscript in preparation). Mali and Burkina Faso sieve analyses results shared 18 target loci, 12 of which were single-copy genes, including the well-described Thrombospondin-related anonymous protein (TRAP) (Dwivedi et al., Manuscript in preparation). Previously developed TRAP-based vaccines reported suboptimal results (Tiono et al., 2018, Bejon et al., 2006), but it is clear this protein product plays an important role in acquired immunity, either naturally acquired or vaccine-induced. Another identified target includes PF3D7_0113800, a DBL domain that is associated to Rho GTPase activation involved in hepatocyte invasion (Aghazadeh et al., 1998, Schepis et al., 2023). PF3D7_1361800 encodes a glideosome-associated connector (GAC), a component involved in the movement of parasites through tissue, and cellular invasion and egress (Jacot et al., 2016). Many of the identified targets were conserved, membrane-bound or membrane-associated, which if present on the sporozoite surface may reflect why cellular-mediated or humoral immunity was elicited (Dwivedi et al., Manuscript in preparation). Ultimately, this is the first instance of applying wgSA as an agnostic approach for target discovery of next-generation vaccines. Due to its novelty, this approach requires further exploration and subsequent immunological examination in humans and mouse models to assess its validity to identify new relevant immune targets.

With the advent and large scale deployment of the single-target and single allele (PfCSP from the 3D7 reference strain) malaria vaccines, RTS,S and R21, and preliminary observations of induction of allele-specific immunity, we might expect to see vaccine escaping strains appearing and eventually expanding. Vaccine protection using PfCSP remains incomplete and short-lived, and a combination with non-PfCSP (e.g. *msp1*) targets are necessary to improve protection. Sustained surveillance of infecting strains is essential to optimizing vaccine effects and prevent the expansion of resistance strains. It is clear other iterations of malaria vaccines are needed. Leveraging WGS for wgSA of clinical samples among infected vaccinated and infected unvaccinated individuals can identify those variants involved in acquired immunity. This type of sieve analysis can inform malaria vaccine developers of antigen targets and their allelic variants that may elicit broader and more sustained vaccine-mediated protection in different populations globally.

6 CONCLUSION AND FUTURE DIRECTIONS

Through this doctoral work, I have successfully employed molecular methods and sequencing data to investigate key aspects of malaria control. A core objective was to enhance the

accessibility and application of these methods in malaria-endemic regions, offering more accurate surveillance and faster results. This work has contributed to capacity-building efforts in Equatorial Guinea, particularly at the national reference laboratory in Baney on Bioko Island. I hope to further scientific development and the understanding of malaria, to enable its progress towards eradication, in Central and West Africa by demonstrating the relevance and feasibility of integrating advanced molecular techniques in local, malaria-endemic settings.

Whole genome sequencing (WGS) provides the most comprehensive NGS data, applicable in various ways, as evidenced in this thesis. It supports the monitoring of *P. falciparum* for diagnostic escape, molecular epidemiology and vaccine development. Despite higher upfront costs and infrastructure demands, advancements in WGS technology are reducing these barriers, potentially broadening its application in support of evidence-based decision-making in NMPs.

The detailed and informative data derived from molecular methods make them ideal for incorporation into NMP control and prevention strategies. Further development of these tools should be prioritized to reduce costs and promote investment in infrastructure. In particular, enhancing bioinformatics and computational analyses of molecular data is crucial. Although local utilization of these data may be hindered by limited availability of human resources, the potential benefits are substantial. By integrating these tools into existing public health frameworks, we can improve the detection and monitoring of malaria, leading to more effective control strategies. The portable nature of the PlasmoPod and the scalability of WGS make these tools suitable for resource-limited settings, enabling decentralized and rapid responses to any change in malaria transmission dynamics. Moving forward, we plan to extend this technology within Equatorial Guinea, fostering local expertise and preparing for future outbreaks of emerging or re-emerging pathogens, even beyond malaria. This proactive approach will ensure that public health authorities can make informed decisions quickly, enhancing the effectiveness of intervention strategies and contributing to global health security.

Future projects are also in preparation to build upon the achievements of this work. Further development of GC_3 aims to identify gene duplications associated with drug resistance using NGS data. Additionally, the next GC_3 iteration aims to simplify the input data required by users, facilitating its use to an even broader audience of investigators. Other projects include the generation of WGS from the 2020 MIS conducted on Bioko Island, when a travel moratorium was in place in Equatorial Guinea during the SARS-CoV-2 pandemic, potentially sharply reducing the importation of parasites from the mainland (Rio Muni). This will ideally provide the most accurate representation of the local *P. falciparum* population, and improve the foundational knowledge to predict imported cases and rollout targeted malaria control interventions. Finally, similar samples to those used in the wgSA were collected on Bioko

Island and with their sequencing, these can be used to explore the versatility of wgSA and facilitate the identification of targets offering broader protection among varying African populations.

Many aspects of this doctoral work represent pioneering efforts in unexplored areas, including the first whole genome description of *P. falciparum* in Equatorial Guinea, the large-scale implementation of a novel qPCR device using MIS samples, and the first use of WGS to conduct a genome-wide selection analysis (wgSA). While the full impact of these contributions remains to be seen due to their novelty, they may provide valuable insights for molecular surveillance and vaccine development.

Ultimately, significant obstacles remain in the way for malaria eradication efforts. However, under the risk of increased mortality if efforts are stopped, the only way for the global community is constant adjustment of approaches and pushing forward.

7 REFERENCES

ABUKARI, Z., OKONU, R., NYARKO, S. B., LO, A. C., DIENG, C. C., SALIFU, S. P., GYAN, B. A., LO, E. & AMOAH, L. E. 2019. The Diversity, Multiplicity of Infection and Population Structure of P. falciparum Parasites Circulating in Asymptomatic Carriers Living in High and Low Malaria Transmission Settings of Ghana. *Genes (Basel),* 10.

ADEBAMOWO, S. N., FRANCIS, V., TAMBO, E., DIALLO, S. H., LANDOURÉ, G., NEMBAWARE, V., DARENG, E., MUHAMED, B., ODUTOLA, M., AKEREDOLU, T., NERIMA, B., OZUMBA, P. J., MBHELE, S., GHANASH, A., WACHINOU, A. P. & NGOMI, N. 2018. Implementation of genomics research in Africa: challenges and recommendations. *Global Health Action,* 11**,** 1419033.

AGHAZADEH, B., ZHU, K., KUBISESKI, T. J., LIU, G. A., PAWSON, T., ZHENG, Y. & ROSEN, M. K. 1998. Structure and mutagenesis of the Dbl homology domain. *Nat Struct Biol,* 5**,** 1098-107.

AJOGBASILE, F. V., KAYODE, A. T., OLUNIYI, P. E., AKANO, K. O., UWANIBE, J. N., ADEGBOYEGA, B. B., PHILIP, C., JOHN, O. G., EROMON, P. J., EMECHEBE, G., FINIMO, F., OGBULAFOR, N., JIYA, N., OKAFOR, U., AMBE, J., WAMMANDA, R. D., OGUCHE, S., MOKUOLU, O. A., SOWUNMI, A., FOLARIN, O. A. & HAPPI, C. T. 2021. Genetic diversity and population structure of Plasmodium falciparum in Nigeria: insights from microsatellite loci analysis. *Malaria Journal,* 20**,** 236.

AKONIYON, O. P., ADEWUMI, T. S., MAHARAJ, L., OYEGOKE, O. O., ROUX, A., ADELEKE, M. A., MAHARAJ, R. & OKPEKU, M. 2022. Whole Genome Sequencing Contributions and Challenges in Disease Reduction Focused on Malaria. *Biology,* 11**,** 587.

ALVES-ROSA, M. F., TAYLER, N. M., DORTA, D., CORONADO, L. M. & SPADAFORA, C. 2024. P. falciparum Invasion and Erythrocyte Aging. *Cells,* 13**,** 334.

AMAMBUA-NGWA, A., AMENGA-ETEGO, L., KAMAU, E., AMATO, R., GHANSAH, A., GOLASSA, L., RANDRIANARIVELOJOSIA, M., ISHENGOMA, D., APINJOH, T., MAÏGA-ASCOFARÉ, O., ANDAGALU, B., YAVO, W., BOUYOU-AKOTET, M., KOLAPO, O., MANE, K., WORWUI, A., JEFFRIES, D., SIMPSON, V., D'ALESSANDRO, U., KWIATKOWSKI, D. & DJIMDE, A. A. 2019. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. 365**,** 813-816.

ARIEY, F., WITKOWSKI, B., AMARATUNGA, C., BEGHAIN, J., LANGLOIS, A.-C., KHIM, N., KIM, S., DURU, V., BOUCHIER, C., MA, L., LIM, P., LEANG, R., DUONG, S., SRENG, S., SUON, S., CHUOR, C. M., BOUT, D. M., MÉNARD, S., ROGERS, W. O., GENTON, B., FANDEUR, T., MIOTTO, O., RINGWALD, P., LE BRAS, J., BERRY, A., BARALE, J.-C., FAIRHURST, R. M., BENOIT-VICAL, F., MERCEREAU-PUIJALON, O. & MÉNARD, D. 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature,* 505**,** 50-55.

ARNOTT, A., BARRY, A. E. & REEDER, J. C. 2012. Understanding the population genetics of Plasmodium vivax is essential for malaria control and elimination. *Malaria Journal,* 11**,** 14.

ASHLEY, E. A., PYAE PHYO, A. & WOODROW, C. J. 2018. Malaria. *Lancet,* 391**,** 1608-1621.

AUBURN, S. & BARRY, A. E. 2017. Dissecting malaria biology and epidemiology using population genetics and genomics. *International Journal for Parasitology,* 47**,** 77-85.

AUDIBERT, C., ASPINALL, A., TCHOUATIEU, A.-M. & HUGO, P. 2024. Evaluation of Segmentation, Rotation, and Geographic Delivery Approaches for Deployment of Multiple First-Line Treatment (MFT) to Respond to Antimalarial Drug Resistance in Africa: A Qualitative Study in Seven Sub-Sahara Countries. *Tropical Medicine and Infectious Disease,* 9**,** 93.

AURRECOECHEA, C., BRESTELLI, J., BRUNK, B. P., DOMMER, J., FISCHER, S., GAJRIA, B., GAO, X., GINGLE, A., GRANT, G., HARB, O. S., HEIGES, M., INNAMORATO, F., IODICE, J., KISSINGER, J. C., KRAEMER, E., LI, W., MILLER, J. A., NAYAK, V., PENNINGTON, C., PINNEY, D. F., ROOS, D. S., ROSS, C., STOECKERT, C. J., JR., TREATMAN, C. & WANG, H. 2008. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research,* 37**,** D539-D543.

BALIKAGALA, B., FUKUDA, N., IKEDA, M., KATURO, O. T., TACHIBANA, S.-I., YAMAUCHI, M., OPIO, W., EMOTO, S., ANYWAR, D. A. & KIMURA, E. 2021. Evidence of artemisinin-resistant malaria in Africa. *New England Journal of Medicine,* 385**,** 1163-1171.

BARRY, A. & HANSEN, D. 2016. Naturally acquired immunity to malaria. *Parasitology,* 143**,** 125-128.

BECHTOLD, P., WAGNER, P., HOSCH, S., GREGORINI, M., STARK, W. J., GODY, J. C., KODIA-LENGUETAMA, E. R., PAGONENDJI, M. S., DONFACK, O. T., PHIRI, W. P., GARCÍA, G. A., NSANZANBANA, C., DAUBENBERGER, C. A., SCHINDLER, T. & VICKOS, U. 2023. Development and evaluation of PlasmoPod: A cartridge-based nucleic acid amplification test for rapid malaria diagnosis and surveillance. *PLOS Global Public Health,* 3**,** e0001516.

BECHTOLD, P., WAGNER, P., HOSCH, S., SIEGRIST, D., RUIZ-SERRANO, A., GREGORINI, M., MPINA, M., ONDÓ, F. A., OBAMA, J., AYEKABA, M. O. O., ENGLER, O., STARK, W. J., DAUBENBERGER, C. A. & SCHINDLER, T. 2021. Rapid Identification of SARS-CoV-2 Variants of Concern Using a Portable peakPCR Platform. *Analytical Chemistry,* 93**,** 16350-16359.

BEESON, J. G., KURTOVIC, L., DOBAÑO, C., OPI, D. H., CHAN, J.-A., FENG, G., GOOD, M. F., REILING, L. & BOYLE, M. J. 2019. Challenges and strategies for developing efficacious and long-lasting malaria vaccines. *Science Translational Medicine,* 11**,** eaau1458.

BEHJATI, S. & TARPEY, P. S. 2013. What is next generation sequencing? *Archives of disease in childhood. Education and practice edition,* 98**,** 236-238.

BEJON, P., MWACHARO, J., KAI, O., MWANGI, T., MILLIGAN, P., TODRYK, S., KEATING, S., LANG, T., LOWE, B., GIKONYO, C., MOLYNEUX, C., FEGAN, G., GILBERT, S. C., PESHU, N., MARSH, K. & HILL, A. V. S. 2006. A Phase 2b Randomised Trial of the Candidate Malaria Vaccines FP9 ME-TRAP and MVA ME-TRAP among Children in Kenya. *PLOS Clinical Trials,* 1**,** e29.

BELETE, T. M. 2020. Recent Progress in the Development of New Antimalarial Drugs with Novel Targets. *Drug Design, Development and Therapy,* 14**,** 3875-3889.

BENNINK, S., KIESOW, M. J. & PRADEL, G. 2016. The development of malaria parasites in the mosquito midgut. *Cellular Microbiology,* 18**,** 905-918.

BERECZKY, S., MARTENSSON, A., GIL, J. P. & FARNERT, A. 2005. Rapid DNA extraction from archive blood spots on filter paper for genotyping of Plasmodium falciparum. *American Journal of Tropical Medicine and Hygiene,* 72**,** 249-251.

BERZOSA, P., DE LUCIO, A., ROMAY-BARJA, M., HERRADOR, Z., GONZÁLEZ, V., GARCÍA, L., FERNÁNDEZ-MARTÍNEZ, A., SANTANA-MORALES, M., NCOGO, P., VALLADARES, B., RILOHA, M. & BENITO, A. 2018. Comparison of three diagnostic methods (microscopy, RDT, and PCR) for the detection of malaria parasites in representative samples from Equatorial Guinea. *Malaria Journal,* 17**,** 333.

BERZOSA, P., GONZÁLEZ, V., TARAVILLO, L., MAYOR, A., ROMAY-BARJA, M., GARCÍA, L., NCOGO, P., RILOHA, M. & BENITO, A. 2020. First evidence of the deletion in the pfhrp2 and pfhrp3 genes in Plasmodium falciparum from Equatorial Guinea. *Malaria Journal,* 19**,** 99.

BHATT, S., WEISS, D. J., CAMERON, E., BISANZIO, D., MAPPIN, B., DALRYMPLE, U., BATTLE, K. E., MOYES, C. L., HENRY, A., ECKHOFF, P. A., WENGER, E. A., BRIËT, O., PENNY, M. A., SMITH, T. A., BENNETT, A., YUKICH, J., EISELE, T. P., GRIFFIN, J. T., FERGUS, C. A., LYNCH, M., LINDGREN, F., COHEN, J. M., MURRAY, C. L. J., SMITH, D. L., HAY, S. I., CIBULSKIS, R. E. & GETHING, P. W. 2015. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature,* 526**,** 207-211.

BIRNBAUM, J., SCHARF, S., SCHMIDT, S., JONSCHER, E., HOEIJMAKERS, W. A. M., FLEMMING, S., TOENHAKE, C. G., SCHMITT, M., SABITZKI, R., BERGMANN, B., FRÖHLKE, U., MESÉN-RAMÍREZ, P., BLANCKE SOARES, A., HERRMANN, H., BÁRTFAI, R. & SPIELMANN, T. 2020. A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. *Science,* 367**,** 51-59.

BJÖRKMAN, A. & MORRIS, U. 2020. Why Asymptomatic *Plasmodium falciparum* Infections Are Common in Low-Transmission Settings. *Trends in Parasitology,* 36**,** 898-905.
BLACKBURN, D. 2023. Outbreak of locally acquired mosquito-transmitted (autochthonous) malaria—Florida and Texas, May–July 2023. *MMWR. Morbidity and Mortality Weekly Report,* 72.
BOUSEMA, T., OKELL, L., FELGER, I. & DRAKELEY, C. 2014. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nature Reviews Microbiology,* 12**,** 833-840.
BRADLEY, J., MONTI, F., REHMAN, A. M., SCHWABE, C., VARGAS, D., GARCIA, G., HERGOTT, D., RILOHA, M. & KLEINSCHMIDT, I. 2015. Infection importation: a key challenge to malaria elimination on Bioko Island, Equatorial Guinea. *Malaria Journal,* 14**,** 46.
BRANCUCCI, N. M., GERDT, J. P., WANG, C., DE NIZ, M., PHILIP, N., ADAPA, S. R., ZHANG, M., HITZ, E., NIEDERWIESER, I. & BOLTRYK, S. D. 2017. Lysophosphatidylcholine regulates sexual stage differentiation in the human malaria parasite Plasmodium falciparum. *Cell,* 171**,** 1532-1544. e15.
BRAY, P. G., MARTIN, R. E., TILLEY, L., WARD, S. A., KIRK, K. & FIDOCK, D. A. 2005. Defining the role of PfCRT in Plasmodium falciparum chloroquine resistance. *Molecular Microbiology,* 56**,** 323-333.
BUCHWALD, A. G., SORKIN, J. D., SIXPENCE, A., CHIMENYA, M., DAMSON, M., WILSON, M. L., SEYDEL, K., HOCHMAN, S., MATHANGA, D., TAYLOR, T. E. & LAUFER, M. K. 2018. Association Between Age and Plasmodium falciparum Infection Dynamics. *American Journal of Epidemiology,* 188**,** 169-176.
CARTER, L. L., YU, M. A., SACKS, J. A., BARNADAS, C., PEREYASLOV, D., COGNAT, S., BRIAND, S., RYAN, M. J. & SAMAAN, G. 2022. Global genomic surveillance strategy for pathogens with pandemic and epidemic potential 2022-2032. *Bulletin of the World Health Organization,* 100**,** 239-239A.
CAVANAGH, D. R., DODOO, D., HVIID, L., KURTZHALS, J. A. L., THEANDER, T. G., AKANMORI, B. D., POLLEY, S., CONWAY, D. J., KORAM, K. & MCBRIDE, J. S. 2004. Antibodies to the N-Terminal Block 2 of *Plasmodium falciparum* Merozoite Surface Protein 1 Are Associated with Protection against Clinical Malaria. *Infection and Immunity,* 72**,** 6492-6502.
CHANDRAMOHAN, D., ZONGO, I., SAGARA, I., CAIRNS, M., YERBANGA, R.-S., DIARRA, M., NIKIÈMA, F., TAPILY, A., SOMPOUGDOU, F., ISSIAKA, D., ZOUNGRANA, C., SANOGO, K., HARO, A., KAYA, M., SIENOU, A.-A., TRAORE, S., MAHAMAR, A., THERA, I., DIARRA, K., DOLO, A., KUEPFER, I., SNELL, P., MILLIGAN, P., OCKENHOUSE, C., OFORI-ANYINAM, O., TINTO, H., DJIMDE, A., OUÉDRAOGO, J.-B., DICKO, A. & GREENWOOD, B. 2021. Seasonal Malaria Vaccination with or without Seasonal Malaria Chemoprevention. *New England Journal of Medicine,* 385**,** 1005-1017.
CHAUHAN, V. S., YAZDANI, S. S. & GAUR, D. 2010. Malaria vaccine development based on merozoite surface proteins of Plasmodium falciparum. *Human Vaccines,* 6**,** 757-762.
CHEAVEAU, J., MOGOLLON, D. C., MOHON, M. A. N., GOLASSA, L., YEWHALAW, D. & PILLAI, D. R. 2019. Asymptomatic malaria in the clinical and public health context. *Expert Review of Anti-infective Therapy,* 17**,** 997-1010.
CHEN, I. & HSIANG, M. S. 2022. Triple artemisinin-based combination therapies for malaria: a timely solution to counter antimalarial drug resistance. *The Lancet Infectious Diseases,* 22**,** 751-753.
CHIN, W., CONTACOS, P. G., COLLINS, W. E., JETER, M. H. & ALPERT, E. 1968. Experimental mosquito-transmission of Plasmodium knowlesi to man and monkey.
CHOI, L., MAJAMBERE, S. & WILSON, A. L. 2019. Larviciding to prevent malaria transmission. *Cochrane Database of Systematic Reviews*.
CIA. 2024a. *Cameroon: People and Society* [Online]. Available: https://www.cia.gov/the-world-factbook/countries/cameroon/ [Accessed 01.08.2024].
CIA. 2024b. *Equatorial Guinea: People and Society* [Online]. Available: https://www.cia.gov/the-world-factbook/countries/equatorial-guinea/ [Accessed 01.08.2024].

CITRON, D. T., GUERRA, C. A., GARCÍA, G. A., WU, S. L., BATTLE, K. E., GIBSON, H. S. & SMITH, D. L. 2021. Quantifying malaria acquired during travel and its role in malaria elimination on Bioko Island. *Malaria Journal,* 20**,** 359.

CLYDE, D. F. 1975. Immunization of man against falciparum and vivax malaria by use of attenuated sporozoites. *The American journal of tropical medicine and hygiene,* 24**,** 397-401.

COCKBURN, I. A. & SEDER, R. A. 2018. Malaria prevention: from immunological concepts to effective vaccines and protective antibodies. *Nat Immunol,* 19**,** 1199-1211.

COHEN, J. M., OKUMU, F. & MOONEN, B. 2022. The fight against malaria: Diminishing gains and growing challenges. *Science Translational Medicine,* 14**,** eabn3256.

CONRAD, M. D. & ROSENTHAL, P. J. 2019. Antimalarial drug resistance in Africa: the calm before the storm? *The Lancet Infectious Diseases,* 19**,** e338-e351.

CONWAY, D. J. 2015. Paths to a malaria vaccine illuminated by parasite genomics. *Trends in Genetics,* 31**,** 97-107.

COOK, J., HERGOTT, D., PHIRI, W., RIVAS, M. R., BRADLEY, J., SEGURA, L., GARCIA, G., SCHWABE, C. & KLEINSCHMIDT, I. 2018. Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. *Malaria Journal,* 17**,** 62.

COWAN, G. J. M., CREASEY, A. M., DHANASARNSOMBUT, K., THOMAS, A. W., REMARQUE, E. J. & CAVANAGH, D. R. 2011. A Malaria Vaccine Based on the Polymorphic Block 2 Region of MSP-1 that Elicits a Broad Serotype-Spanning Immune Response. *PLOS ONE,* 6**,** e26616.

COWELL, A. N. & WINZELER, E. A. 2019. The genomic architecture of antimalarial drug resistance. *Brief Funct Genomics*.

CRAVO, P., NAPOLITANO, H. & CULLETON, R. 2015. How genomics is contributing to the fight against artemisinin-resistant malaria parasites. *Acta Tropica,* 148**,** 1-7.

DALRYMPLE, U., ARAMBEPOLA, R., GETHING, P. W. & CAMERON, E. 2018. How long do rapid diagnostic tests remain positive after anti-malarial treatment? *Malaria Journal,* 17**,** 228.

DAMBACH, P., BÄRNIGHAUSEN, T., YADOULETON, A., DAMBACH, M., TRAORÉ, I., KORIR, P., OUEDRAOGO, S., NIKIEMA, M., SAUERBORN, R., BECKER, N. & LOUIS, V. R. 2021. Is biological larviciding against malaria a starting point for integrated multi-disease control? Observations from a cluster randomized trial in rural Burkina Faso. *PLoS One,* 16**,** e0253597.

DANIELS, R. F., SCHAFFNER, S. F., DIEYE, Y., DIENG, G., HAINSWORTH, M., FALL, F. B., DIOUF, C. N., NDIOP, M., CISSE, M., GUEYE, A. B., SARR, O., GUINOT, P., DEME, A. B., BEI, A. K., SY, M., THWING, J., MACINNIS, B., EARLE, D., GUINOVART, C., SENE, D., HARTL, D. L., NDIAYE, D., STEKETEE, R. W., WIRTH, D. F. & VOLKMAN, S. K. 2020. Genetic evidence for imported malaria and local transmission in Richard Toll, Senegal. *Malaria Journal,* 19**,** 276.

DANWANG, C., KIRAKOYA-SAMADOULOUGOU, F. & SAMADOULOUGOU, S. 2021. Assessing field performance of ultrasensitive rapid diagnostic tests for malaria: a systematic review and meta-analysis. *Malaria Journal,* 20**,** 245.

DATOO, M. S., DICKO, A., TINTO, H., OUÉDRAOGO, J.-B., HAMALUBA, M., OLOTU, A., BEAUMONT, E., RAMOS LOPEZ, F., NATAMA, H. M., WESTON, S., CHEMBA, M., COMPAORE, Y. D., ISSIAKA, D., SALOU, D., SOME, A. M., OMENDA, S., LAWRIE, A., BEJON, P., RAO, H., CHANDRAMOHAN, D., ROBERTS, R., BHARATI, S., STOCKDALE, L., GAIROLA, S., GREENWOOD, B. M., EWER, K. J., BRADLEY, J., KULKARNI, P. S., SHALIGRAM, U., HILL, A. V. S., MAHAMAR, A., SANOGO, K., SIDIBE, Y., DIARRA, K., SAMASSEKOU, M., ATTAHER, O., TAPILY, A., DIALLO, M., DICKO, O. M., KAYA, M., MAGUIRAGA, S. O., SANKARE, Y., YALCOUYE, H., DIARRA, S., NIAMBELE, S. M., THERA, I., SAGARA, I., SYLLA, M., DOLO, A., MISIDAI, N., SIMANDO, S., MSAMI, H., JUMA, O., GUTAPAKA, N., PAUL, R., MSWATA, S., SASAMALO, I., JOHANESS, K., SULTAN, M., ALEXANDER, A., KIMARO, I., LWANGA, K., MTUNGWE, M., KHAMIS, K., RUGARABAM, L., KALINGA, W., MOHAMMED, M., KAMANGE, J., MSANGI, J., MWAIJANDE, B., MTAKA, I., MHAPA, M., MLAGANILE, T., MBAGA, T., YERBANGA, R. S., SAMTOUMA, W.,

SIENOU, A. A., KABRE, Z., OUEDRAOGO, W. J. M., YARBANGA, G. A. B., ZONGO, I., SAVADOGO, H., SANON, J., COMPAORE, J., KERE, I., YONI, F. L., SANRE, T. M., OUATTARA, S. B., PROVSTGAARD-MORYS, S., WOODS, D., SNOW, R. W., AMEK, N., NGETSA, C. J., OCHOLA-OYIER, L. I., MUSYOKI, J., MUNENE, M., MUMBA, N., ADETIFA, U. J., MUIRURI, C. M., MWAWAKA, J. S., et al. 2024. Safety and efficacy of malaria vaccine candidate R21/Matrix-M in African children: a multicentre, double-blind, randomised, phase 3 trial. *The Lancet,* 403**,** 533-544.

DE OLIVEIRA, T. C., CORDER, R. M., EARLY, A., RODRIGUES, P. T., LADEIA-ANDRADE, S., ALVES, J. M. P., NEAFSEY, D. E. & FERREIRA, M. U. 2020. Population genomics reveals the expansion of highly inbred Plasmodium vivax lineages in the main malaria hotspot of Brazil. *PLOS Neglected Tropical Diseases,* 14**,** e0008808.

DEEN, J. L. & CLEMENS, J. D. 2006. Issues in the design and implementation of vaccine trials in less developed countries. *Nature Reviews Drug Discovery,* 5**,** 932-940.

DEGAREGE, A., FENNIE, K., DEGAREGE, D., CHENNUPATI, S. & MADHIVANAN, P. 2019. Improving socioeconomic status may reduce the burden of malaria in sub Saharan Africa: A systematic review and meta-analysis. *PLOS ONE,* 14**,** e0211205.

DHARIA, N. V., BRIGHT, A. T., WESTENBERGER, S. J., BARNES, S. W., BATALOV, S., KUHEN, K., BORBOA, R., FEDERE, G. C., MCCLEAN, C. M., VINETZ, J. M., NEYRA, V., LLANOS-CUENTAS, A., BARNWELL, J. W., WALKER, J. R. & WINZELER, E. A. 2010a. Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. *Proc Natl Acad Sci U S A,* 107**,** 20045-50.

DHARIA, N. V., PLOUFFE, D., BOPP, S. E. R., GONZÁLEZ-PÁEZ, G. E., LUCAS, C., SALAS, C., SOBERON, V., BURSULAYA, B., KOCHEL, T. J., BACON, D. J. & WINZELER, E. A. 2010b. Genome scanning of Amazonian Plasmodium falciparum shows subtelomeric instability and clindamycin-resistant parasites. *Genome Research,* 20**,** 1534-1544.

DONDORP, A. M., NOSTEN, F., YI, P., DAS, D., PHYO, A. P., TARNING, J., LWIN, K. M., ARIEY, F., HANPITHAKPONG, W. & LEE, S. J. 2009. Artemisinin resistance in Plasmodium falciparum malaria. *New England journal of medicine,* 361**,** 455-467.

DUWELL, M., DEVITA, T., CHEN, J., MYERS, R. A., MACE, K., RIDPATH, A. D., ODONGO, W., RAPHAEL, B. H., LENHART, A. & TONGREN, J. E. 2023. Locally Acquired Mosquito-Transmitted (Autochthonous) Plasmodium falciparum Malaria--National Capital Region, Maryland, August 2023. *MMWR: Morbidity & Mortality Weekly Report,* 72.

DVORIN, J. D. 2017. Getting Your Head around Cerebral Malaria. *Cell Host Microbe,* 22**,** 586-588.

DWIVEDI, A., SCALSKY, R. J., STABLER, T. C., SHRESTH, B., JOSHI, S., HARRIS, D. G., GANDHI, C., MUNRO, J. B., IFEONU, O. O., OUEDRAOGO, A., TIONO, A. B., COULIBALY, D., OUATTARA, A., RICHIE, T. L., SIM, B. K. L., PLOWE, C. V., LYKE, K. E., TAKALA-HARRISON, S., HOFFMAN, S. L., THERA, M. A., SIRIMA, S. B., LAURENS, M. B. & SILVA, J. C. Manuscript in preparation. Protective targets of PfSPZ vaccines identified from whole-genome sieve analyses of isolates from malaria vaccine efficacy trials.

EASTMAN, R. T. & FIDOCK, D. A. 2009. Artemisinin-based combination therapies: a vital tool in efforts to eliminate malaria. *Nature Reviews Microbiology,* 7**,** 864-874.

EPOPA, P. S., COLLINS, C. M., NORTH, A., MILLOGO, A. A., BENEDICT, M. Q., TRIPET, F. & DIABATE, A. 2019. Seasonal malaria vector and transmission dynamics in western Burkina Faso. *Malaria Journal,* 18**,** 113.

FAKIH, B. S., HOLZSCHUH, A., ROSS, A., STUCK, L., ABDUL, R., AL-MAFAZY, A.-W. H., IREMA, I., MBENA, A., THAWER, S. G., SHIJA, S. J., ALIY, S. M., ALI, A., FINK, G., YUKICH, J. & HETZEL, M. W. 2023. Risk of imported malaria infections in Zanzibar: a cross-sectional study. *Infectious Diseases of Poverty,* 12**,** 80.

FÄRNERT, A., AREZ, A. P., CORREIA, A. T., BJÖRKMAN, A., SNOUNOU, G. & DO ROSÁRIO, V. 1999. Sampling and storage of blood and the detection of malaria parasites by polymerase chain reaction. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 93**,** 50-53.

FERREIRA, M. U., RIBEIRO, W. L., TONON, A. P., KAWAMOTO, F. & RICH, S. M. 2003. Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of Plasmodium falciparum. *Gene,* 304**,** 65-75.

FISCHER, L., GÜLTEKIN, N., KAELIN, M. B., FEHR, J. & SCHLAGENHAUF, P. 2020. Rising temperature and its impact on receptivity to malaria transmission in Europe: A systematic review. *Travel Medicine and Infectious Disease,* 36**,** 101815.

FOX, E. J., REID-BAYLISS, K. S., EMOND, M. J. & LOEB, L. A. 2014. Accuracy of Next Generation Sequencing Platforms. *Next Gener Seq Appl,* 1.

FRÜH, K., DOUMBO, O., MÜLLER, H. M., KOITA, O., MCBRIDE, J., CRISANTI, A., TOURÉ, Y. & BUJARD, H. 1991. Human antibody response to the major merozoite surface antigen of Plasmodium falciparum is strain specific and short-lived. *Infection and Immunity,* 59**,** 1319-1324.

FUEHRER, H.-P., CAMPINO, S. & SUTHERLAND, C. J. 2022. The primate malaria parasites Plasmodium malariae, Plasmodium brasilianum and Plasmodium ovale spp.: genomic insights into distribution, dispersal and host transitions. *Malaria Journal,* 2**1,** 138.

FUENTES-RAMÍREZ, A., JIMÉNEZ-SOTO, M., CASTRO, R., ROMERO-ZUÑIGA, J. J. & DOLZ, G. 2017. Molecular Detection of Plasmodium malariae/Plasmodium brasilianum in Non-Human Primates in Captivity in Costa Rica. *PLOS ONE,* 12**,** e0170704.

GALINSKY, K., VALIM, C., SALMIER, A., DE THOISY, B., MUSSET, L., LEGRAND, E., FAUST, A., BANIECKI, M. L., NDIAYE, D., DANIELS, R. F., HARTL, D. L., SABETI, P. C., WIRTH, D. F., VOLKMAN, S. K. & NEAFSEY, D. E. 2015. COIL: a methodology for evaluating malarial complexity of infection using likelihood from single nucleotide polymorphism data. *Malaria Journal,* 14**,** 4.

GAMBOA, D., HO, M.-F., BENDEZU, J., TORRES, K., CHIODINI, P. L., BARNWELL, J. W., INCARDONA, S., PERKINS, M., BELL, D., MCCARTHY, J. & CHENG, Q. 2010. A Large Proportion of P. falciparum Isolates in the Amazon Region of Peru Lack pfhrp2 and pfhrp3: Implications for Malaria Rapid Diagnostic Tests. *PLOS ONE,* 5**,** e8091.

GARCÍA, G. A., JANKO, M., HERGOTT, D. E. B., DONFACK, O. T., SMITH, J. M., MBA EYONO, J. N., DEBOER, K. R., NGUEMA AVUE, R. M., PHIRI, W. P., ALDRICH, E. M., SCHWABE, C., STABLER, T. C., RIVAS, M. R., CAMERON, E., GUERRA, C. A., COOK, J., KLEINSCHMIDT, I. & BRADLEY, J. 2023. Identifying individual, household and environmental risk factors for malaria infection on Bioko Island to inform interventions. *Malaria Journal,* 22**,** 72.

GARDNER, M. J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R. W., CARLTON, J. M., PAIN, A., NELSON, K. E., BOWMAN, S., PAULSEN, I. T., JAMES, K., EISEN, J. A., RUTHERFORD, K., SALZBERG, S. L., CRAIG, A., KYES, S., CHAN, M. S., NENE, V., SHALLOM, S. J., SUH, B., PETERSON, J., ANGIUOLI, S., PERTEA, M., ALLEN, J., SELENGUT, J., HAFT, D., MATHER, M. W., VAIDYA, A. B., MARTIN, D. M., FAIRLAMB, A. H., FRAUNHOLZ, M. J., ROOS, D. S., RALPH, S. A., MCFADDEN, G. I., CUMMINGS, L. M., SUBRAMANIAN, G. M., MUNGALL, C., VENTER, J. C., CARUCCI, D. J., HOFFMAN, S. L., NEWBOLD, C., DAVIS, R. W., FRASER, C. M. & BARRELL, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature,* 419**,** 498-511.

GATTON, M. L. & CHENG, Q. 2010. Interrupting Malaria Transmission: Quantifying the Impact of Interventions in Regions of Low to Moderate Transmission. *PLOS ONE,* 5**,** e15149.

GATTON, M. L., MARTIN, L. B. & CHENG, Q. 2004. Evolution of resistance to sulfadoxine-pyrimethamine in Plasmodium falciparum. *Antimicrobial agents and chemotherapy,* 48**,** 2116-2123.

GENDROT, M., FAWAZ, R., DORMOI, J., MADAMET, M. & PRADINES, B. 2019. Genetic diversity and deletion of Plasmodium falciparum histidine-rich protein 2 and 3: a threat to diagnosis of P. falciparum malaria. *Clinical Microbiology and Infection,* 25**,** 580-585.

GHANI, A. C., SUTHERLAND, C. J., RILEY, E. M., DRAKELEY, C. J., GRIFFIN, J. T., GOSLING, R. D. & FILIPE, J. A. N. 2009. Loss of Population Levels of Immunity to Malaria as a Result of Exposure-Reducing Interventions: Consequences for Interpretation of Disease Trends. *PLOS ONE,* 4**,** e4383.

GIMNIG, J. E., STEINHARDT, L. C., AWOLOLA, T. S., IMPOINVIL, D., ZOHDY, S. & LINDBLADE, K. A. 2024. Reducing Malaria Transmission through Reactive Indoor

Residual Spraying: A Systematic Review. *The American Journal of Tropical Medicine and Hygiene,* 110**,** 94-100.
GITTA, B. & KILIAN, N. 2020. Diagnosis of Malaria Parasites Plasmodium spp. in Endemic Areas: Current Strategies for an Ancient Disease. *BioEssays,* 42**,** 1900138.
GOLASSA, L., MESSELE, A., AMAMBUA-NGWA, A. & SWEDBERG, G. 2020. High prevalence and extended deletions in Plasmodium falciparum hrp2/3 genomic loci in Ethiopia. *PLOS ONE,* 15**,** e0241807.
GOLUMBEANU, M., EDI, C. A., HETZEL, M. W., KOEPFLI, C. & NSANZABANA, C. 2023. Bridging the Gap from Molecular Surveillance to Programmatic Decisions for Malaria Control and Elimination. *The American Journal of Tropical Medicine and Hygiene,* 1.
GOSLING, R., CHIMUMBWA, J., UUSIKU, P., ROSSI, S., NTUKU, H., HARVARD, K., WHITE, C., TATARSKY, A., CHANDRAMOHAN, D. & CHEN, I. 2020. District-level approach for tailoring and targeting interventions: a new path for malaria control and elimination. *Malaria Journal,* 19**,** 125.
GUERRA, C. A., KANG, S. Y., CITRON, D. T., HERGOTT, D. E. B., PERRY, M., SMITH, J., PHIRI, W. P., OSÁ NFUMU, J. O., MBA EYONO, J. N., BATTLE, K. E., GIBSON, H. S., GARCÍA, G. A. & SMITH, D. L. 2019. Human mobility patterns and malaria importation on Bioko Island. *Nature Communications,* 10**,** 2332.
GUO, B., BORDA, V., LABOULAYE, R., SPRING, M. D., WOJNARSKI, M., VESELY, B. A., SILVA, J. C., WATERS, N. C., O'CONNOR, T. D. & TAKALA-HARRISON, S. 2024. Strong positive selection biases identity-by-descent-based inferences of recent demography and population structure in Plasmodium falciparum. *Nature Communications,* 15**,** 2499.
HACKL, S. T., HARBIG, T. A. & NIESELT, K. 2022. Technical report on best practices for hybrid and long read de novo assembly of bacterial genomes utilizing Illumina and Oxford Nanopore Technologies reads. *bioRxiv***,** 2022.10.25.513682.
HAMMERSHAIMB, E. A. & BERRY, A. A. 2024. Pre-erythrocytic malaria vaccines: RTS,S, R21, and beyond. *Expert Review of Vaccines,* 23**,** 49-52.
HASSETT, M. R. & ROEPE, P. D. 2019. Origin and Spread of Evolving Artemisinin-Resistant Plasmodium falciparum Malarial Parasites in Southeast Asia. *Am J Trop Med Hyg,* 101**,** 1204-1211.
HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.
HEINEMANN, M., PHILLIPS, R. O., VINNEMEIER, C. D., ROLLING, C. C., TANNICH, E. & ROLLING, T. 2020. High prevalence of asymptomatic malaria infections in adults, Ashanti Region, Ghana, 2018. *Malaria Journal,* 19**,** 366.
HENDRIKSEN, I. C. E., MTOVE, G., PEDRO, A. J., GOMES, E., SILAMUT, K., LEE, S. J., MWAMBULI, A., GESASE, S., REYBURN, H., DAY, N. P. J., WHITE, N. J., VON SEIDLEIN, L. & DONDORP, A. M. 2011. Evaluation of a PfHRP2 and a pLDH-based Rapid Diagnostic Test for the Diagnosis of Severe Malaria in 2 Populations of African Children. *Clinical Infectious Diseases,* 52**,** 1100-1107.
HOFFMAN, S. L., VEKEMANS, J., RICHIE, T. L. & DUFFY, P. E. 2015. The March Toward Malaria Vaccines. *Am J Prev Med,* 49**,** S319-33.
HOFMANN, N., MWINGIRA, F., SHEKALAGHE, S., ROBINSON, L. J., MUELLER, I. & FELGER, I. 2015. Ultra-Sensitive Detection of Plasmodium falciparum by Amplification of Multi-Copy Subtelomeric Targets. *PLOS Medicine,* 12**,** e1001788.
HOLZSCHUH, A. & KOEPFLI, C. 2022. Tenfold difference in DNA recovery rate: systematic comparison of whole blood vs. dried blood spot sample collection for malaria molecular surveillance. *Malaria Journal,* 21**,** 88.
HOSCH, S., YOBOUE, C. A., DONFACK, O. T., GUIROU, E. A., DANGY, J.-P., MPINA, M., NYAKURUNGU, E., BLÖCHLIGER, K., GUERRA, C. A., PHIRI, W. P., AYEKABA, M. O. O., GARCÍA, G. A., TANNER, M., DAUBENBERGER, C. & SCHINDLER, T. 2022. Analysis of nucleic acids extracted from rapid diagnostic tests reveals a significant proportion of false positive test results associated with recent malaria treatment. *Malaria Journal,* 21**,** 23.
HUPALO, D. N., LUO, Z., MELNIKOV, A., SUTTON, P. L., ROGOV, P., ESCALANTE, A., VALLEJO, A. F., HERRERA, S., ARÉVALO-HERRERA, M., FAN, Q., WANG, Y., CUI,

L., LUCAS, C. M., DURAND, S., SANCHEZ, J. F., BALDEVIANO, G. C., LESCANO, A. G., LAMAN, M., BARNADAS, C., BARRY, A., MUELLER, I., KAZURA, J. W., EAPEN, A., KANAGARAJ, D., VALECHA, N., FERREIRA, M. U., ROOBSOONG, W., NGUITRAGOOL, W., SATTABONKOT, J., GAMBOA, D., KOSEK, M., VINETZ, J. M., GONZÁLEZ-CERÓN, L., BIRREN, B. W., NEAFSEY, D. E. & CARLTON, J. M. 2016. Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nature Genetics,* 48**,** 953-958.

JACOT, D., TOSETTI, N., PIRES, I., STOCK, J., GRAINDORGE, A., HUNG, Y.-F., HAN, H., TEWARI, R., KURSULA, I. & SOLDATI-FAVRE, D. 2016. An Apicomplexan Actin-Binding Protein Serves as a Connector and Lipid Sensor to Coordinate Motility and Invasion. *Cell Host & Microbe,* 20**,** 731-743.

JAIN, M., KOREN, S., MIGA, K. H., QUICK, J., RAND, A. C., SASANI, T. A., TYSON, J. R., BEGGS, A. D., DILTHEY, A. T., FIDDES, I. T., MALLA, S., MARRIOTT, H., NIETO, T., O'GRADY, J., OLSEN, H. E., PEDERSEN, B. S., RHIE, A., RICHARDSON, H., QUINLAN, A. R., SNUTCH, T. P., TEE, L., PATEN, B., PHILLIPPY, A. M., SIMPSON, J. T., LOMAN, N. J. & LOOSE, M. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology,* 36**,** 338-345.

JONG, E. C. & STEVENS, D. L. 2021. *Netter's Infectious Diseases-E-Book: Netter's Infectious Diseases-E-Book*, Elsevier Health Sciences.

JONGO, S. A., CHURCH, L. W. P., NCHAMA, V. U. N. N., HAMAD, A., CHUQUIYAURI, R., KASSIM, K. R., ATHUMAN, T., DEAL, A., NATASHA, K. C., MTORO, A., MPINA, M., NYAKARUNGU, E., BIDJIMI, G. O., OWONO, M. A., MAYE, E. R. M., MANGUE, M. E. O., OKOMO, G. N. N., PASIALO, B. E. N., MANDUMBI, D. M. O., MIKUE, M.-S. A. L., MOCHOMUEMUE, F. L., OBONO, M. O., BESAHA, J. C. M., BIJERI, J. R., ABEGUE, G. M., VERI, Y. R., BELA, I. T., CHOCHI, F. C., LIMA SÁNCHEZ, J. E., PENCELLI, V., GAYOZO, G., NLANG, J. A. E. M., SCHINDLER, T., JAMES, E. R., ABEBE, Y., LEMIALE, L., STABLER, T. C., MURSHEDKAR, T., CHEN, M.-C., SCHWABE, C., RATSIRARSON, J., RIVAS, M. R., AYEKABA, M. O. O., MILANG, D. V. N., FALLA, C. C., PHIRI, W. P., GARCÍA, G. A., MAAS, C. D., NLAVO, B. M., TANNER, M., BILLINGSLEY, P. F., KIM LEE SIM, B., DAUBENBERGER, C., HOFFMAN, S. L., ABDULLA, S. & RICHIE, T. L. 2022. Multi-Dose Priming Regimens of PfSPZ Vaccine: Safety and Efficacy against Controlled Human Malaria Infection in Equatoguinean Adults. *The American journal of tropical medicine and hygiene,* 106**,** 1215-1226.

JOSLING, G. A. & LLINÁS, M. 2015. Sexual development in Plasmodium parasites: knowing when it's time to commit. *Nature Reviews Microbiology,* 13**,** 573-587.

KAMAU, E., ALEMAYEHU, S., FEGHALI, K. C., SAUNDERS, D. & OCKENHOUSE, C. F. 2013. Multiplex qPCR for Detection and Absolute Quantification of Malaria. *PLOS ONE,* 8**,** e71539.

KAMAU, E., TOLBERT, L. S., KORTEPETER, L., PRATT, M., NYAKOE, N., MURINGO, L., OGUTU, B., WAITUMBI, J. N. & OCKENHOUSE, C. F. 2011. Development of a highly sensitive genus-specific quantitative reverse transcriptase real-time PCR assay for detection and quantitation of plasmodium by amplifying RNA and DNA of the 18S rRNA genes. *J Clin Microbiol,* 49**,** 2946-53.

KARIUKI, S. N. & WILLIAMS, T. N. 2020. Human genetics and malaria resistance. *Human Genetics,* 139**,** 801-811.

KEITANY, G. J., KIM, K. S., KRISHNAMURTY, A. T., HONDOWICZ, B. D., HAHN, W. O., DAMBRAUSKAS, N., SATHER, D. N., VAUGHAN, A. M., KAPPE, S. H. I. & PEPPER, M. 2016. Blood Stage Malaria Disrupts Humoral Immunity to the Pre-erythrocytic Stage Circumsporozoite Protein. *Cell Reports,* 17**,** 3193-3205.

KIGOZI, S. P., KIGOZI, R. N., EPSTEIN, A., MPIMBAZA, A., SSERWANGA, A., YEKA, A., NANKABIRWA, J. I., HALLIDAY, K., PULLAN, R. L., RUTAZAANA, D., SEBUGUZI, C. M., OPIGO, J., KAMYA, M. R., STAEDKE, S. G., DORSEY, G., GREENHOUSE, B. & RODRIGUEZ-BARRAQUER, I. 2020. Rapid shifts in the age-specific burden of malaria following successful control interventions in four regions of Uganda. *Malaria Journal,* 19**,** 128.

KISAND, V. & LETTIERI, T. 2013. Genome sequencing of bacteria: sequencing, de novoassembly and rapid analysis using open source tools. *BMC Genomics,* 14**,** 211.
KIWANUKA, G. N. 2009. Genetic diversity in Plasmodium falciparum merozoite surface protein 1 and 2 coding genes and its implications in malaria epidemiology: a review of published studies from 1997–2007.
KIYUKA, P. K., MERI, S. & KHATTAB, A. 2020. Complement in malaria: immune evasion strategies and role in protective immunity. *FEBS Letters,* 594**,** 2502-2517.
KOGAN, F. 2020. Malaria Burden. *Remote Sensing for Malaria: Monitoring and Predicting Malaria from Operational Satellites.* Cham: Springer International Publishing.
KOLLURI, N., KLAPPERICH, C. & CABODI, M. 2018. Towards lab-on-a-chip diagnostics for malaria elimination. *Lab on a Chip,* 18**,** 75-94.
KONG, A., WILSON, S. A., AH, Y., NACE, D., ROGIER, E. & AIDOO, M. 2021. HRP2 and HRP3 cross-reactivity and implications for HRP2-based RDT use in regions with Plasmodium falciparum hrp2 gene deletions. *Malaria Journal,* 20**,** 207.
KONO, N. & ARAKAWA, K. 2019. Nanopore sequencing: Review of potential applications in functional genomics. 61**,** 316-326.
KUBLIN, J. G., CORTESE, J. F., NJUNJU, E. M., G. MUKADAM, R. A., WIRIMA, J. J., KAZEMBE, P. N., DJIMDÉ, A. A., KOURIBA, B., TAYLOR, T. E. & PLOWE, C. V. 2003. Reemergence of Chloroquine-Sensitive Plasmodium falciparum Malaria after Cessation of Chloroquine Use in Malawi. *The Journal of Infectious Diseases,* 187**,** 1870-1875.
LABAN, N. M., KOBAYASHI, T., HAMAPUMBU, H., SULLIVAN, D., MHARAKURWA, S., THUMA, P. E., SHIFF, C. J., MOSS, W. J. & FOR SOUTHERN AFRICA INTERNATIONAL CENTERS OF EXCELLENCE FOR MALARIA, R. 2015. Comparison of a PfHRP2-based rapid diagnostic test and PCR for malaria in a low prevalence setting in rural southern Zambia: implications for elimination. *Malaria Journal,* 14**,** 25.
LAMBACH, P., SILAL, S., SBARRA, A. N., KOH, M., AGGARWAL, R., FAROOQUI, H. H., FLASCHE, S., HOGAN, A. B., KIM, S.-Y., LEUNG, K., MOSS, W. J., MUNYWOKI, P. K., PORTNOY, A., SHEEL, M. & WANG, X.-Y. 2024. Report from the World Health Organization's immunization and vaccines-related implementation research advisory committee (IVIR-AC) meeting, virtual gathering, 26 February–1 March 2024. *Vaccine,* 42**,** 3379-3383.
LAN, J. H., YIN, Y., REED, E. F., MOUA, K., THOMAS, K. & ZHANG, Q. 2015. Impact of three Illumina library construction methods on GC bias and HLA genotype calling. *Hum Immunol,* 76**,** 166-75.
LANGHORNE, J., NDUNGU, F. M., SPONAAS, A.-M. & MARSH, K. 2008. Immunity to malaria: more questions than answers. *Nature Immunology,* 9**,** 725-732.
LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9**,** 357-359.
LAUFER, M. K., TAKALA-HARRISON, S., DZINJALAMALA, F. K., STINE, O. C., TAYLOR, T. E. & PLOWE, C. V. 2010. Return of Chloroquine-Susceptible Falciparum Malaria in Malawi Was a Reexpansion of Diverse Susceptible Parasites. *The Journal of Infectious Diseases,* 202**,** 801-808.
LAUFER, M. K., THESING, P. C., EDDINGTON, N. D., MASONGA, R., DZINJALAMALA, F. K., TAKALA, S. L., TAYLOR, T. E. & PLOWE, C. V. 2006. Return of Chloroquine Antimalarial Efficacy in Malawi. *New England Journal of Medicine,* 355**,** 1959-1966.
LAURENSON, A. J. & LAURENS, M. B. 2024. A new landscape for malaria vaccine development. *PLOS Pathogens,* 20**,** e1012309.
LE ROCH, K. G., CHUNG, D.-W. D. & PONTS, N. 2012. Genomics and integrated systems biology in Plasmodium falciparum: a path to malaria control and eradication. *Parasite Immunology,* 34**,** 50-60.
LEE, K.-S., DIVIS, P. C. S., ZAKARIA, S. K., MATUSOP, A., JULIN, R. A., CONWAY, D. J., COX-SINGH, J. & SINGH, B. 2011. Plasmodium knowlesi: Reservoir Hosts and Tracking the Emergence in Humans and Macaques. *PLOS Pathogens,* 7**,** e1002015.
LEFEVRE, T., OHM, J., DABIRÉ, K. R., COHUET, A., CHOISY, M., THOMAS, M. B. & CATOR, L. 2018. Transmission traits of malaria parasites within the mosquito: Genetic

variation, phenotypic plasticity, and consequences for control. *Evolutionary Applications,* 11**,** 456-469.
LEWONTIN, R. C. 1988. On measures of gametic disequilibrium. *Genetics,* 120**,** 849-852.
LI, C. & MANAGI, S. 2022. Global malaria infection risk from climate change. *Environmental Research,* 214**,** 114028.
LIAO, X., LI, M., ZOU, Y., WU, F.-X., YI-PAN & WANG, J. 2019. Current challenges and solutions of de novo assembly. *Quantitative Biology,* 7**,** 90-109.
LIU, Q., JING, W., KANG, L., LIU, J. & LIU, M. 2021. Trends of the global, regional and national incidence of malaria in 204 countries from 1990 to 2019 and implications for malaria prevention. *Journal of Travel Medicine,* 28.
LOPEZ, L. & KOEPFLI, C. 2021. Systematic review of Plasmodium falciparum and Plasmodium vivax polyclonal infections: Impact of prevalence, study population characteristics, and laboratory procedures. *PLOS ONE,* 16**,** e0249382.
LU, H., GIORDANO, F. & NING, Z. 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics,* 14**,** 265-279.
LUKE, T. C. & HOFFMAN, S. L. 2003. Rationale and plans for developing a non-replicating, metabolically active, radiation-attenuated Plasmodium falciparum sporozoite vaccine. *Journal of Experimental Biology,* 206**,** 3803-3808.
MALERA 2017. malERA: An updated research agenda for characterising the reservoir and measuring transmission in malaria elimination and eradication. *PLoS Med,* 14**,** e1002452.
MARTIN, E. R. & CHUNG, R.-H. 2021. Linkage Disequilibrium and Association Analysis. *Genetic Analysis of Complex Diseases.*
MASON, C. E. & ELEMENTO, O. 2012. Faster sequencers, larger datasets, new challenges. *Genome Biology,* 13**,** 314.
MEEK, M. H. & LARSON, W. A. 2019. The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources,* 19**,** 795-803.
MELNIKOV, A., GALINSKY, K., ROGOV, P., FENNELL, T., VAN TYNE, D., RUSS, C., DANIELS, R., BARNES, K. G., BOCHICCHIO, J., NDIAYE, D., SENE, P. D., WIRTH, D. F., NUSBAUM, C., VOLKMAN, S. K., BIRREN, B. W., GNIRKE, A. & NEAFSEY, D. E. 2011. Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biology,* 12**,** R73.
MÉNARD, R., SULTAN, A. A., CORTES, C., ALTSZULER, R., VAN DIJK, M. R., JANSE, C. J., WATERS, A. P., NUSSENZWEIG, R. S. & NUSSENZWEIG, V. 1997. Circumsporozoite protein is required for development of malaria sporozoites in mosquitoes. *Nature,* 385**,** 336-340.
MERRIMAN, B., D TEAM, I. T. & ROTHBERG, J. M. 2012. Progress in Ion Torrent semiconductor chip based sequencing. *ELECTROPHORESIS,* 33**,** 3397-3417.
METZKER, M. L. 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics,* 11**,** 31-46.
MEULENAERE, K. D., CUYPERS, W. L., GAUGLITZ, J. M., GUETENS, P., ROSANAS-URGELL, A., LAUKENS, K. & CUYPERS, B. 2024. Selective whole-genome sequencing of *Plasmodium* parasites directly from blood samples by nanopore adaptive sampling. *mBio,* 15**,** e01967-23.
MIHRETEAB, S., ANDERSON, K., PASAY, C., SMITH, D., GATTON, M. L., CUNNINGHAM, J., BERHANE, A. & CHENG, Q. 2021. Epidemiology of mutant Plasmodium falciparum parasites lacking histidine-rich protein 2/3 genes in Eritrea 2 years after switching from HRP2-based RDTs. *Scientific Reports,* 11**,** 21082.
MIHRETEAB, S., PLATON, L., BERHANE, A., STOKES, B. H., WARSAME, M., CAMPAGNE, P., CRISCUOLO, A., MA, L., PETIOT, N., DODERER-LANG, C., LEGRAND, E., WARD, K. E., KASSAHUN, A. Z., RINGWALD, P., FIDOCK, D. A. & MÉNARD, D. 2023. Increasing Prevalence of Artemisinin-Resistant HRP2-Negative Malaria in Eritrea. *New England Journal of Medicine,* 389**,** 1191-1202.
MIOTTO, O., ALMAGRO-GARCIA, J., MANSKE, M., MACINNIS, B., CAMPINO, S., ROCKETT, K. A., AMARATUNGA, C., LIM, P., SUON, S., SRENG, S., ANDERSON, J. M., DUONG, S., NGUON, C., CHUOR, C. M., SAUNDERS, D., SE, Y., LON, C.,

FUKUDA, M. M., AMENGA-ETEGO, L., HODGSON, A. V., ASOALA, V., IMWONG, M., TAKALA-HARRISON, S., NOSTEN, F., SU, X. Z., RINGWALD, P., ARIEY, F., DOLECEK, C., HIEN, T. T., BONI, M. F., THAI, C. Q., AMAMBUA-NGWA, A., CONWAY, D. J., DJIMDE, A. A., DOUMBO, O. K., ZONGO, I., OUEDRAOGO, J. B., ALCOCK, D., DRURY, E., AUBURN, S., KOCH, O., SANDERS, M., HUBBART, C., MASLEN, G., RUANO-RUBIO, V., JYOTHI, D., MILES, A., O'BRIEN, J., GAMBLE, C., OYOLA, S. O., RAYNER, J. C., NEWBOLD, C. I., BERRIMAN, M., SPENCER, C. C., MCVEAN, G., DAY, N. P., WHITE, N. J., BETHELL, D., DONDORP, A. M., PLOWE, C. V., FAIRHURST, R. M. & KWIATKOWSKI, D. P. 2013. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat Genet,* 45**,** 648-55.

MOBEGI, V. A., DUFFY, C. W., AMAMBUA-NGWA, A., LOUA, K. M., LAMAN, E., NWAKANMA, D. C., MACINNIS, B., ASPELING-JONES, H., MURRAY, L., CLARK, T. G., KWIATKOWSKI, D. P. & CONWAY, D. J. 2014. Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity. *Mol Biol Evol,* 31**,** 1490-9.

MOLINA-DE LA FUENTE, I., PACHECO, M. A., GARCÍA, L., GONZÁLEZ, V., RILOHA, M., OKI, C., BENITO, A., ESCALANTE, A. A. & BERZOSA, P. 2024. Evolution of pfhrp2 and pfhrp3 deletions in Equatorial Guinea between the pre– and post–RDT introduction. *Malaria Journal,* 23**,** 215.

MOLINA - DE LA FUENTE, I., YIMAR, M., GARCÍA, L., GONZÁLEZ, V., AMOR, A., ANEGAGRIE, M., BENITO, A., MARTÍNEZ, J., MORENO, M. & BERZOSA, P. 2022. Deletion patterns, genetic variability and protein structure of pfhrp2 and pfhrp3: implications for malaria rapid diagnostic test in Amhara region, Ethiopia. *Malaria Journal,* 21**,** 287.

MOODY, A. 2002. Rapid Diagnostic Tests for Malaria Parasites. 15**,** 66-78.

MORGAN, A. P., BRAZEAU, N. F., NGASALA, B., MHAMILAWA, L. E., DENTON, M., MSELLEM, M., MORRIS, U., FILER, D. L., AYDEMIR, O., BAILEY, J. A., PARR, J. B., MÅRTENSSON, A., BJORKMAN, A. & JULIANO, J. J. 2020. Falciparum malaria from coastal Tanzania and Zanzibar remains highly connected despite effective control efforts on the archipelago. *Malaria Journal,* 19**,** 47.

MOSER, K. A., DRÁBEK, E. F., DWIVEDI, A., STUCKE, E. M., CRABTREE, J., DARA, A., SHAH, Z., ADAMS, M., LI, T., RODRIGUES, P. T., KOREN, S., PHILLIPPY, A. M., MUNRO, J. B., OUATTARA, A., SPARKLIN, B. C., DUNNING HOTOPP, J. C., LYKE, K. E., SADZEWICZ, L., TALLON, L. J., SPRING, M. D., JONGSAKUL, K., LON, C., SAUNDERS, D. L., FERREIRA, M. U., NYUNT, M. M., LAUFER, M. K., TRAVASSOS, M. A., SAUERWEIN, R. W., TAKALA-HARRISON, S., FRASER, C. M., SIM, B. K. L., HOFFMAN, S. L., PLOWE, C. V. & SILVA, J. C. 2020. Strains used in whole organism Plasmodium falciparum vaccine trials differ in genome structure, sequence, and immunogenic potential. *Genome Medicine,* 12**,** 6.

MOURIER, T., DE ALVARENGA, D. A. M., KAUSHIK, A., DE PINA-COSTA, A., DOUVROPOULOU, O., GUAN, Q., GUZMÁN-VEGA, F. J., FORRESTER, S., DE ABREU, F. V. S., JÚNIOR, C. B., DE SOUZA JUNIOR, J. C., MOREIRA, S. B., HIRANO, Z. M. B., PISSINATTI, A., FERREIRA-DA-CRUZ, M. D. F., DE OLIVEIRA, R. L., AROLD, S. T., JEFFARES, D. C., BRASIL, P., DE BRITO, C. F. A., CULLETON, R., DANIEL-RIBEIRO, C. T. & PAIN, A. 2021. The genome of the zoonotic malaria parasite Plasmodium simium reveals adaptations to host switching. *BMC Biology,* 19**,** 219.

MPINA, M., STABLER, T. C., SCHINDLER, T., RASO, J., DEAL, A., ACUCHE PUPU, L., NYAKARUNGU, E., DEL CARMEN OVONO DAVIS, M., URBANO, V., MTORO, A., HAMAD, A., LOPEZ, M. S. A., PASIALO, B., EYANG, M. A. O., RIVAS, M. R., FALLA, C. C., GARCÍA, G. A., MOMO, J. C., CHUQUIYAURI, R., SAVERINO, E., PRESTON CHURCH, L. W., KIM LEE SIM, B., MANGUIRE, B., TANNER, M., MAAS, C., ABDULLA, S., BILLINGSLEY, P. F., HOFFMAN, S. L., JONGO, S., RICHIE, T. L. & DAUBENBERGER, C. A. 2022. Diagnostic performance and comparison of ultrasensitive and conventional rapid diagnostic test, thick blood smear and quantitative PCR for detection of low-density Plasmodium falciparum infections during a controlled human malaria infection study in Equatorial Guinea. *Malaria Journal,* 21**,** 99.

NASIR, S. M. I., AMARASEKARA, S., WICKREMASINGHE, R., FERNANDO, D. & UDAGAMA, P. 2020. Prevention of re-establishment of malaria: historical perspective and future prospects. *Malaria Journal,* 19**,** 452.

NCHAMA, V. U. N. N., SAID, A. H., MTORO, A., BIDJIMI, G. O., OWONO, M. A., MAYE, E. R. M., MANGUE, M. E. O., OKOMO, G. N. N., PASIALO, B. E. N., ONDO, D. M., LOPEZ, M.-S. A., MOCHOMUEMUE, F. L., OBONO, M. O., BESAHA, J. C. M., CHUQUIYAURI, R., JONGO, S. A., KAMAKA, K., KIBONDO, U. A., ATHUMAN, T., FALLA, C. C., EYONO, J. N. M., SMITH, J. M., GARCÍA, G. A., RASO, J., NYAKARUNGU, E., MPINA, M., SCHINDLER, T., DAUBENBERGER, C., LEMIALE, L., BILLINGSLEY, P. F., SIM, B. K. L., RICHIE, T. L., CHURCH, L. W. P., OLOTU, A., TANNER, M., HOFFMAN, S. L. & ABDULLA, S. 2021. Incidence of Plasmodium falciparum malaria infection in 6-month to 45-year-olds on selected areas of Bioko Island, Equatorial Guinea. *Malaria Journal,* 20**,** 322.

NEAFSEY, D. E., JURASKA, M., BEDFORD, T., BENKESER, D., VALIM, C., GRIGGS, A., LIEVENS, M., ABDULLA, S., ADJEI, S., AGBENYEGA, T., AGNANDJI, S. T., AIDE, P., ANDERSON, S., ANSONG, D., APONTE, J. J., ASANTE, K. P., BEJON, P., BIRKETT, A. J., BRULS, M., CONNOLLY, K. M., D'ALESSANDRO, U., DOBANO, C., GESASE, S., GREENWOOD, B., GRIMSBY, J., TINTO, H., HAMEL, M. J., HOFFMAN, I., KAMTHUNZI, P., KARIUKI, S., KREMSNER, P. G., LEACH, A., LELL, B., LENNON, N. J., LUSINGU, J., MARSH, K., MARTINSON, F., MOLEL, J. T., MOSS, E. L., NJUGUNA, P., OCKENHOUSE, C. F., OGUTU, B. R., OTIENO, W., OTIENO, L., OTIENO, K., OWUSU-AGYEI, S., PARK, D. J., PELLE, K., ROBBINS, D., RUSS, C., RYAN, E. M., SACARLAL, J., SOGOLOFF, B., SORGHO, H., TANNER, M., THEANDER, T., VALEA, I., VOLKMAN, S. K., YU, Q., LAPIERRE, D., BIRREN, B. W., GILBERT, P. B. & WIRTH, D. F. 2015. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med,* 373**,** 2025-2037.

NEAFSEY, D. E., SCHAFFNER, S. F., VOLKMAN, S. K., PARK, D., MONTGOMERY, P., MILNER, D. A., LUKENS, A., ROSEN, D., DANIELS, R., HOUDE, N., CORTESE, J. F., TYNDALL, E., GATES, C., STANGE-THOMANN, N., SARR, O., NDIAYE, D., NDIR, O., MBOUP, S., FERREIRA, M. U., MORAES, S. D. L., DASH, A. P., CHITNIS, C. E., WIEGAND, R. C., HARTL, D. L., BIRREN, B. W., LANDER, E. S., SABETI, P. C. & WIRTH, D. F. 2008. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biology,* 9**,** R171.

NEAFSEY, D. E., TAYLOR, A. R. & MACINNIS, B. L. 2021. Advances and opportunities in malaria population genomics. *Nature Reviews Genetics,* 22**,** 502-517.

NG, P. C. & KIRKNESS, E. F. 2010. Whole Genome Sequencing. *In:* BARNES, M. R. & BREEN, G. (eds.) *Genetic Variation: Methods and Protocols.* Totowa, NJ: Humana Press.

NOAKES, M. T., BRINKERHOFF, H., LASZLO, A. H., DERRINGTON, I. M., LANGFORD, K. W., MOUNT, J. W., BOWMAN, J. L., BAKER, K. S., DOERING, K. M., TICKMAN, B. I. & GUNDLACH, J. H. 2019. Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nature Biotechnology,* 37**,** 651-656.

NUSSENZWEIG, R. S., VANDERBERG, J., MOST, H. & ORTON, C. 1967. Protective Immunity produced by the Injection of X-irradiated Sporozoites of Plasmodium berghei. *Nature,* 216**,** 160-162.

NWAKANMA, D. C., DUFFY, C. W., AMAMBUA-NGWA, A., ORIERO, E. C., BOJANG, K. A., PINDER, M., DRAKELEY, C. J., SUTHERLAND, C. J., MILLIGAN, P. J., MACINNIS, B., KWIATKOWSKI, D. P., CLARK, T. G., GREENWOOD, B. M. & CONWAY, D. J. 2014. Changes in malaria parasite drug resistance in an endemic population over a 25-year period with resulting genomic evidence of selection. *J Infect Dis,* 209**,** 1126-35.

NZIOKI, I., MACHANI, M. G., ONYANGO, S. A., KABUI, K. K., GITHEKO, A. K., OCHOMO, E., YAN, G. & AFRANE, Y. A. 2023. Differences in malaria vector biting behavior and changing vulnerability to malaria transmission in contrasting ecosystems of western Kenya. *Parasites & Vectors,* 16**,** 376.

OBALDIA, N., III, BARO, N. K., CALZADA, J. E., SANTAMARIA, A. M., DANIELS, R., WONG, W., CHANG, H.-H., HAMILTON, E. J., AREVALO-HERRERA, M., HERRERA, S., WIRTH, D. F., HARTL, D. L., MARTI, M. & VOLKMAN, S. K. 2014. Clonal Outbreak of

Plasmodium falciparum Infection in Eastern Panama. *The Journal of Infectious Diseases,* 211**,** 1087-1096.
OKELL, L. C., BOUSEMA, T., GRIFFIN, J. T., OUÉDRAOGO, A. L., GHANI, A. C. & DRAKELEY, C. J. 2012. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nature Communications,* 3**,** 1237.
OLSON, N. D., LUND, S. P., COLMAN, R. E., FOSTER, J. T., SAHL, J. W., SCHUPP, J. M., KEIM, P., MORROW, J. B., SALIT, M. L. & ZOOK, J. M. 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. 6.
OLUSEGUN-JOSEPH, T., OBOH, M., OVIOMA, G., FAGBOHUN, I., OKORAFOR, U. & AINA, D. 2019. Differential prevalence of malaria infection in rural and urban out-patient clinics in Lagos state, Nigeria. *Pan African J Life Sci,* 2**,** 79-84.
OSBORNE, A., MANKO, E., TAKEDA, M., KANEKO, A., KAGAYA, W., CHAN, C., NGARA, M., KONGERE, J., KITA, K., CAMPINO, S., KANEKO, O., GITAKA, J. & CLARK, T. G. 2021. Characterizing the genomic variation and population dynamics of Plasmodium falciparum malaria parasites in and around Lake Victoria, Kenya. *Scientific Reports,* 11**,** 19809.
OTTO, T. D., BÖHME, U., SANDERS, M., REID, A., BRUSKE, E. I., DUFFY, C. W., BULL, P. C., PEARSON, R. D., ABDI, A., DIMONTE, S., STEWART, L. B., CAMPINO, S., KEKRE, M., HAMILTON, W. L., CLAESSENS, A., VOLKMAN, S. K., NDIAYE, D., AMAMBUA-NGWA, A., DIAKITE, M., FAIRHURST, R. M., CONWAY, D. J., FRANCK, M., NEWBOLD, C. I. & BERRIMAN, M. 2018. Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Res,* 3**,** 52.
OUATTARA, A., TAKALA-HARRISON, S., THERA, M. A., COULIBALY, D., NIANGALY, A., SAYE, R., TOLO, Y., DUTTA, S., HEPPNER, D. G., SOISSON, L., DIGGS, C. L., VEKEMANS, J., COHEN, J., BLACKWELDER, W. C., DUBE, T., LAURENS, M. B., DOUMBO, O. K. & PLOWE, C. V. 2012. Molecular Basis of Allele-Specific Efficacy of a Blood-Stage Malaria Vaccine: Vaccine Development Implications. *The Journal of Infectious Diseases,* 207**,** 511-519.
OYOLA, S. O., ARIANI, C. V., HAMILTON, W. L., KEKRE, M., AMENGA-ETEGO, L. N., GHANSAH, A., RUTLEDGE, G. G., REDMOND, S., MANSKE, M., JYOTHI, D., JACOB, C. G., OTTO, T. D., ROCKETT, K., NEWBOLD, C. I., BERRIMAN, M. & KWIATKOWSKI, D. P. 2016. Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malaria Journal,* 15**,** 597.
OYOLA, S. O., GU, Y., MANSKE, M., OTTO, T. D., O'BRIEN, J., ALCOCK, D., MACINNIS, B., BERRIMAN, M., NEWBOLD, C. I. & KWIATKOWSKI, D. P. 2013. Efficient depletion of host DNA contamination in malaria clinical sequencing. *Journal of clinical microbiology,* 51**,** 745-751.
PAYNE, D. 1987. Spread of chloroquine resistance in Plasmodium falciparum. *Parasitol Today,* 3**,** 241-6.
PENGELLY, R. J., TAPPER, W., GIBSON, J., KNUT, M., TEARLE, R., COLLINS, A. & ENNIS, S. 2015. Whole genome sequences are required to fully resolve the linkage disequilibrium structure of human populations. *BMC Genomics,* 16**,** 666.
PHAM, N. M., KARLEN, W., BECK, H.-P. & DELAMARCHE, E. 2018. Malaria and the 'last' parasite: how can technology help? *Malaria Journal,* 17**,** 260.
PORETSKY, R., RODRIGUEZ-R, L. M., LUO, C., TSEMENTZI, D. & KONSTANTINIDIS, K. T. 2014. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLOS ONE,* 9**,** e93827.
POTI, K. E., SULLIVAN, D. J., DONDORP, A. M. & WOODROW, C. J. 2020. HRP2: Transforming Malaria Diagnosis, but with Caveats. *Trends in Parasitology,* 36**,** 112-126.
PRESTON, M. D., CAMPINO, S., ASSEFA, S. A., ECHEVERRY, D. F., OCHOLLA, H., AMAMBUA-NGWA, A., STEWART, L. B., CONWAY, D. J., BORRMANN, S., MICHON, P., ZONGO, I., OUEDRAOGO, J. B., DJIMDE, A. A., DOUMBO, O. K., NOSTEN, F., PAIN, A., BOUSEMA, T., DRAKELEY, C. J., FAIRHURST, R. M., SUTHERLAND, C. J., ROPER, C. & CLARK, T. G. 2014. A barcode of organellar genome polymorphisms

identifies the geographic origin of Plasmodium falciparum strains. *Nat Commun,* 5**,** 4052.
PRUDÊNCIO, M., RODRIGUEZ, A. & MOTA, M. M. 2006. The silent path to thousands of merozoites: the Plasmodium liver stage. *Nature Reviews Microbiology,* 4**,** 849-856.
PRYCE, J., RICHARDSON, M. & LENGELER, C. 2018. Insecticide-treated nets for preventing malaria. *Cochrane Database Syst Rev,* 11**,** Cd000363.
RAJNEESH, TIWARI, R., SINGH, V. K., KUMAR, A., GUPTA, R. P., SINGH, A. K., GAUTAM, V. & KUMAR, R. 2023. Advancements and Challenges in Developing Malaria Vaccines: Targeting Multiple Stages of the Parasite Life Cycle. *ACS Infectious Diseases,* 9**,** 1795-1814.
RANG, F. J., KLOOSTERMAN, W. P. & DE RIDDER, J. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology,* 19**,** 90.
RANJAN, R., RANI, A., METWALLY, A., MCGEE, H. S. & PERKINS, D. L. 2016. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications,* 469**,** 967-977.
RATHORE, D., SACCI, J. B., DE LA VEGA, P. & MCCUTCHAN, T. F. 2002. Binding and Invasion of Liver Cells by *Plasmodium falciparum* Sporozoites: ESSENTIAL INVOLVEMENT OF THE AMINO TERMINUS OF CIRCUMSPOROZOITE PROTEIN *. *Journal of Biological Chemistry,* 277**,** 7092-7098.
REHMAN, A. M., MANN, A. G., SCHWABE, C., REDDY, M. R., RONCON GOMES, I., SLOTMAN, M. A., YELLOTT, L., MATIAS, A., CACCONE, A., NSENG NCHAMA, G. & KLEINSCHMIDT, I. 2013. Five years of malaria control in the continental region, Equatorial Guinea. *Malaria Journal,* 12**,** 154.
RHIE, A., MCCARTHY, S. A., FEDRIGO, O., DAMAS, J., FORMENTI, G., KOREN, S., ULIANO-SILVA, M., CHOW, W., FUNGTAMMASAN, A., KIM, J., LEE, C., KO, B. J., CHAISSON, M., GEDMAN, G. L., CANTIN, L. J., THIBAUD-NISSEN, F., HAGGERTY, L., BISTA, I., SMITH, M., HAASE, B., MOUNTCASTLE, J., WINKLER, S., PAEZ, S., HOWARD, J., VERNES, S. C., LAMA, T. M., GRUTZNER, F., WARREN, W. C., BALAKRISHNAN, C. N., BURT, D., GEORGE, J. M., BIEGLER, M. T., IORNS, D., DIGBY, A., EASON, D., ROBERTSON, B., EDWARDS, T., WILKINSON, M., TURNER, G., MEYER, A., KAUTT, A. F., FRANCHINI, P., DETRICH, H. W., SVARDAL, H., WAGNER, M., NAYLOR, G. J. P., PIPPEL, M., MALINSKY, M., MOONEY, M., SIMBIRSKY, M., HANNIGAN, B. T., PESOUT, T., HOUCK, M., MISURACA, A., KINGAN, S. B., HALL, R., KRONENBERG, Z., SOVIĆ, I., DUNN, C., NING, Z., HASTIE, A., LEE, J., SELVARAJ, S., GREEN, R. E., PUTNAM, N. H., GUT, I., GHURYE, J., GARRISON, E., SIMS, Y., COLLINS, J., PELAN, S., TORRANCE, J., TRACEY, A., WOOD, J., DAGNEW, R. E., GUAN, D., LONDON, S. E., CLAYTON, D. F., MELLO, C. V., FRIEDRICH, S. R., LOVELL, P. V., OSIPOVA, E., AL-AJLI, F. O., SECOMANDI, S., KIM, H., THEOFANOPOULOU, C., HILLER, M., ZHOU, Y., HARRIS, R. S., MAKOVA, K. D., MEDVEDEV, P., HOFFMAN, J., MASTERSON, P., CLARK, K., MARTIN, F., HOWE, K., FLICEK, P., WALENZ, B. P., KWAK, W., CLAWSON, H., et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature,* 592**,** 737-746.
RHOADS, A. & AU, K. F. 2015. PacBio Sequencing and its Applications. *Genomics, Proteomics & Bioinformatics,* 13**,** 278-289.
RODRIGUE, S., MALMSTROM, R. R., BERLIN, A. M., BIRREN, B. W., HENN, M. R. & CHISHOLM, S. W. 2009. Whole Genome Amplification and De novo Assembly of Single Bacterial Cells. *PLOS ONE,* 4**,** e6864.
ROSENTHAL, P. J. 2013. The interplay between drug resistance and fitness in malaria parasites. *Molecular Microbiology,* 89**,** 1025-1038.
ROUGERON, V., DARON, J., FONTAINE, M. C. & PRUGNOLLE, F. 2022. Evolutionary history of Plasmodium vivax and Plasmodium simium in the Americas. *Malaria Journal,* 21**,** 141.
RTS, S CLINICAL TRIALS PARTNERSHIP, 2015. 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. *The Lancet,* 386**,** 31-45.

RUYBAL-PESÁNTEZ, S., MCCANN, K., VIBIN, J., SIEGEL, S., AUBURN, S. & BARRY, A. E. 2024. Molecular markers for malaria genetic epidemiology: progress and pitfalls. *Trends in Parasitology,* 40**,** 147-163.

RYAN, E. T., HILL, D. R., SOLOMON, T., ARONSON, N. & ENDY, T. P. 2019. *Hunter's tropical medicine and emerging infectious diseases*, Elsevier Health Sciences.

SABINA, J. & LEAMON, J. H. 2015. Bias in Whole Genome Amplification: Causes and Considerations. *In:* KRONEIS, T. (ed.) *Whole Genome Amplification: Methods and Protocols.* New York, NY: Springer New York.

SALLA, L. C., RODRIGUES, P. T., CORDER, R. M., JOHANSEN, I. C., LADEIA-ANDRADE, S. & FERREIRA, M. U. 2020. Molecular evidence of sustained urban malaria transmission in Amazonian Brazil, 2014–2015. *Epidemiology and Infection,* 148**,** e47.

SANTOS-VEGA, M., BOUMA, M. J., KOHLI, V. & PASCUAL, M. 2016. Population Density, Climate Variables and Poverty Synergistically Structure Spatial Risk in Urban Malaria in India. *PLOS Neglected Tropical Diseases,* 10**,** e0005155.

SATAM, H., JOSHI, K., MANGROLIA, U., WAGHOO, S., ZAIDI, G., RAWOOL, S., THAKARE, R. P., BANDAY, S., MISHRA, A. K., DAS, G. & MALONIA, S. K. 2023. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology,* 12**,** 997.

SATO, S. 2021. Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology. *Journal of Physiological Anthropology,* 40**,** 1.

SCALSKY, R., DWIVEDI, A., STABLER, T., MBAMBO, G., OUATTARA, A., LYKE, K. E., TAKALA-HARRISON, S. & SILVA, J. C. Under review. Whole-genome sieve analysis: identification of protective malaria antigens by leveraging allele-specific vaccine efficacy. *Vaccine*.

SCHEPIS, A., KUMAR, S. & KAPPE, S. H. 2023. Malaria parasites harness Rho GTPase signaling and host cell membrane ruffling for productive invasion of hepatocytes. *Cell Reports,* 42.

SCHINDLER, T., DEAL, A. C., FINK, M., GUIROU, E., MOSER, K. A., MWAKASUNGULA, S. M., MIHAYO, M. G., JONGO, S. A., CHAKI, P. P., ABDULLA, S., VALVERDE, P. C. M., TORRES, K., BIJERI, J. R., SILVA, J. C., HOFFMAN, S. L., GAMBOA, D., TANNER, M. & DAUBENBERGER, C. 2019. A multiplex qPCR approach for detection of pfhrp2 and pfhrp3 gene deletions in multiple strain infections of Plasmodium falciparum. *Scientific Reports,* 9**,** 13107.

SCHIRMER, M., IJAZ, U. Z., D'AMORE, R., HALL, N., SLOAN, W. T. & QUINCE, C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research,* 43**,** e37-e37.

SCHNEIDER, K. A. 2021. Charles Darwin Meets Ronald Ross: A Population-Genetic Framework for the Evolutionary Dynamics of Malaria. *In:* TEBOH-EWUNGKEM, M. I. & NGWA, G. A. (eds.) *Infectious Diseases and Our Planet.* Cham: Springer International Publishing.

SEDER, R. A., CHANG, L. J., ENAMA, M. E., ZEPHIR, K. L., SARWAR, U. N., GORDON, I. J., HOLMAN, L. A., JAMES, E. R., BILLINGSLEY, P. F., GUNASEKERA, A., RICHMAN, A., CHAKRAVARTY, S., MANOJ, A., VELMURUGAN, S., LI, M., RUBEN, A. J., LI, T., EAPPEN, A. G., STAFFORD, R. E., PLUMMER, S. H., HENDEL, C. S., NOVIK, L., COSTNER, P. J., MENDOZA, F. H., SAUNDERS, J. G., NASON, M. C., RICHARDSON, J. H., MURPHY, J., DAVIDSON, S. A., RICHIE, T. L., SEDEGAH, M., SUTAMIHARDJA, A., FAHLE, G. A., LYKE, K. E., LAURENS, M. B., ROEDERER, M., TEWARI, K., EPSTEIN, J. E., SIM, B. K., LEDGERWOOD, J. E., GRAHAM, B. S., HOFFMAN, S. L. & TEAM, V. R. C. S. 2013. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science,* 341**,** 1359-65.

SEPÚLVEDA, N., PHELAN, J., DIEZ-BENAVENTE, E., CAMPINO, S., CLARK, T. G., HOPKINS, H., SUTHERLAND, C., DRAKELEY, C. J. & BESHIR, K. B. 2018. Global analysis of Plasmodium falciparum histidine-rich protein-2 (pfhrp2) and pfhrp3 gene deletions using whole-genome sequencing data and meta-analysis. *Infection, Genetics and Evolution,* 62**,** 211-219.

SHAH, Z., ADAMS, M., MOSER, K. A., SHRESTHA, B., STUCKE, E. M., LAUFER, M. K., SERRE, D., SILVA, J. C. & TAKALA-HARRISON, S. J. M. J. 2020. Optimization of

parasite DNA enrichment approaches to generate whole genome sequencing data for Plasmodium falciparum from low parasitaemia samples. *Malaria Journal,* 19**,** 1-10.

SHEEHY, S. H., DUNCAN, C. J. A., ELIAS, S. C., CHOUDHARY, P., BISWAS, S., HALSTEAD, F. D., COLLINS, K. A., EDWARDS, N. J., DOUGLAS, A. D., ANAGNOSTOU, N. A., EWER, K. J., HAVELOCK, T., MAHUNGU, T., BLISS, C. M., MIURA, K., POULTON, I. D., LILLIE, P. J., ANTROBUS, R. D., BERRIE, E., MOYLE, S., GANTLETT, K., COLLOCA, S., CORTESE, R., LONG, C. A., SINDEN, R. E., GILBERT, S. C., LAWRIE, A. M., DOHERTY, T., FAUST, S. N., NICOSIA, A., HILL, A. V. S. & DRAPER, S. J. 2012. ChAd63-MVA–vectored Blood-stage Malaria Vaccines Targeting MSP1 and AMA1: Assessment of Efficacy Against Mosquito Bite Challenge in Humans. *Molecular Therapy,* 20**,** 2355-2368.

SHEN, H.-M., CHEN, S.-B., CUI, Y.-B., XU, B., KASSEGNE, K., ABE, E. M., WANG, Y. & CHEN, J.-H. 2018. Whole-genome sequencing and analysis of *Plasmodium falciparum* isolates from China-Myanmar border area. *Infectious Diseases of Poverty,* 07**,** 82-88.

SHENDURE, J., BALASUBRAMANIAN, S., CHURCH, G. M., GILBERT, W., ROGERS, J., SCHLOSS, J. A. & WATERSTON, R. H. 2017. DNA sequencing at 40: past, present and future. *Nature,* 550**,** 345-353.

SHRESTHA, B., SHAH, Z., MORGAN, A. P., SAINGAM, P., CHAISATIT, C., CHAORATTANAKAWEE, S., PRADITPOL, C., BOONYALAI, N., LERTSETHTAKARN, P., WOJNARSKI, M., DEUTSCH-FELDMAN, M., ADAMS, M., SEA, D., CHANN, S., TYNER, S. D., LANTERI, C. A., SPRING, M. D., SAUNDERS, D. L., SMITH, P. L., LON, C., GOSI, P., SOK, S., SATHARATH, P., REKOL, H., LEK, D., VESELY, B. A., LIN, J. T., WATERS, N. C. & TAKALA-HARRISON, S. 2021. Distribution and Temporal Dynamics of Plasmodium falciparum Chloroquine Resistance Transporter Mutations Associated With Piperaquine Resistance in Northern Cambodia. *The Journal of Infectious Diseases,* 224**,** 1077-1085.

SIBLEY, C. H., HYDE, J. E., SIMS, P. F. G., PLOWE, C. V., KUBLIN, J. G., MBERU, E. K., COWMAN, A. F., WINSTANLEY, P. A., WATKINS, W. M. & NZILA, A. M. 2001. Pyrimethamine–sulfadoxine resistance in *Plasmodium falciparum*: what next? *Trends in Parasitology,* 17**,** 582-588.

SINGH, A. P., BUSCAGLIA, C. A., WANG, Q., LEVAY, A., NUSSENZWEIG, D. R., WALKER, J. R., WINZELER, E. A., FUJII, H., FONTOURA, B. M. A. & NUSSENZWEIG, V. 2007. Plasmodium Circumsporozoite Protein Promotes the Development of the Liver Stages of the Parasite. *Cell,* 131**,** 492-504.

SLATER, H. C., ROSS, A., FELGER, I., HOFMANN, N. E., ROBINSON, L., COOK, J., GONÇALVES, B. P., BJÖRKMAN, A., OUEDRAOGO, A. L., MORRIS, U., MSELLEM, M., KOEPFLI, C., MUELLER, I., TADESSE, F., GADISA, E., DAS, S., DOMINGO, G., KAPULU, M., MIDEGA, J., OWUSU-AGYEI, S., NABET, C., PIARROUX, R., DOUMBO, O., DOUMBO, S. N., KORAM, K., LUCCHI, N., UDHAYAKUMAR, V., MOSHA, J., TIONO, A., CHANDRAMOHAN, D., GOSLING, R., MWINGIRA, F., SAUERWEIN, R., PAUL, R., RILEY, E. M., WHITE, N. J., NOSTEN, F., IMWONG, M., BOUSEMA, T., DRAKELEY, C. & OKELL, L. C. 2019. The temporal dynamics and infectiousness of subpatent Plasmodium falciparum infections in relation to parasite density. *Nature Communications,* 10**,** 1433.

STABLER, T. C., DWIVEDI, A., GUO, B., SHRESTHA, B., JOSHI, S., RIVAS, M. R., DONFACK, O. T., GUERRA, C. A., GARCÍA, G. A., DAUBENBERGER, C. & SILVA, J. C. Manuscript in preparation-a. Whole genome sequences of *Plasmodium falciparum* on Bioko Island, Equatorial Guinea illustrate connectedness with geographical neighbors and island-wide partial population structure.

STABLER, T. C., DWIVEDI, A., SHRESTHA, B., JOSHI, S., SCHINDLER, T., OUATTARA, A., GARCÍA, G. A., DAUBENBERGER, C. & SILVA, J. C. 2022. Gene Coverage Count and Classification (GC3): a locus sequence coverage assessment tool using short-read whole genome sequencing data, and its application to identify and classify histidine-rich protein 2 and 3 deletions in Plasmodium falciparum. *Malaria Journal,* 21**,** 357.

STABLER, T. C., HOSCH, S., NYAKARUNGU, E., GIGER, J. N., ELONGA, M. K., BIBANG, R. N., NDONG, V. M., MOMETOLO, I. E., BOLOPA, A. L., BEDOUNG, R. M. O., BIJERI, J. R., ROKA, E. W., NGUEMA, A. G., ICHINDA, V. P., BAKALE, S.,

ESOKOLETE, C. N., RIVAS, M. R., GALICK, D. S., PHIRI, W. P., GUERRA, C. A., GARCÍA, G. A., MONTEMARTINI, L., SCHINDLER, T., SILVA, J. C. & DAUBENBERGER, C. Manuscript in preparation-b. Integrating local malaria molecular monitoring into regular malaria indicator surveys on Bioko Island: high association between urban communities and low density infections.

STANISIC, D. I. & GOOD, M. F. 2023. Malaria Vaccines: Progress to Date. *BioDrugs,* 37**,** 737-756.

STANISIC, D. I., RICHARDS, J. S., MCCALLUM, F. J., MICHON, P., KING, C. L., SCHOEPFLIN, S., GILSON, P. R., MURPHY, V. J., ANDERS, R. F. & MUELLER, I. 2009. Immunoglobulin G subclass-specific responses against Plasmodium falciparum merozoite antigens are associated with control of parasitemia and protection from symptomatic illness. *Infection and immunity,* 77**,** 1165-1174.

STOKES, J. 2009. *Encyclopedia of the Peoples of Africa and the Middle East*, Infobase Publishing.

STRUBBIA, S., PHAN, M. V. T., SCHAEFFER, J., KOOPMANS, M., COTTEN, M. & LE GUYADER, F. S. 2019. Characterization of Norovirus and Other Human Enteric Viruses in Sewage and Stool Samples Through Next-Generation Sequencing. *Food and Environmental Virology,* 11**,** 400-409.

STUCKE, E. M., DARA, A., DWIVEDI, A., HODGES, T. K., OTT, S., COULIBALY, D., KONÉ, A. K., TRAORÉ, K., GUINDO, B., TANGARA, B. M., NIANGALY, A., DAOU, M., DIARRA, I., TOLO, Y., SISSOKO, M., TALLON, L. J., SADZEWICZ, L., ZHOU, A. E., LAURENS, M. B., OUATTARA, A., KOURIBA, B., DOUMBO, O. K., TAKALA-HARRISON, S., SERRE, D., PLOWE, C. V., THERA, M. A., TRAVASSOS, M. A. & SILVA, J. C. 2021. Successful Profiling of Plasmodium falciparum var Gene Expression in Clinical Samples via a Custom Capture Array. *mSystems,* 6**,** e0022621.

SUH, P. F., ELANGA-NDILLE, E., TCHOUAKUI, M., SANDEU, M. M., TAGNE, D., WONDJI, C. & NDO, C. 2023. Impact of insecticide resistance on malaria vector competence: a literature review. *Malaria Journal,* 22**,** 19.

SUTHERLAND, C. J., TANOMSING, N., NOLDER, D., OGUIKE, M., JENNISON, C., PUKRITTAYAKAMEE, S., DOLECEK, C., HIEN, T. T., DO ROSÁRIO, V. E., AREZ, A. P., PINTO, J., MICHON, P., ESCALANTE, A. A., NOSTEN, F., BURKE, M., LEE, R., BLAZE, M., OTTO, T. D., BARNWELL, J. W., PAIN, A., WILLIAMS, J., WHITE, N. J., DAY, N. P. J., SNOUNOU, G., LOCKHART, P. J., CHIODINI, P. L., IMWONG, M. & POLLEY, S. D. 2010. Two Nonrecombining Sympatric Forms of the Human Malaria Parasite Plasmodium ovale Occur Globally. *The Journal of Infectious Diseases,* 201**,** 1544-1550.

SYED, Y. Y. 2022. RTS,S/AS01 malaria vaccine (Mosquirix®): a profile of its use. *Drugs & Therapy Perspectives,* 38**,** 373-381.

TAKALA-HARRISON, S. & LAUFER, M. K. 2015. Antimalarial drug resistance in Africa: key lessons for the future. *Annals of the New York Academy of Sciences,* 1342**,** 62-67.

TAKALA, S., BRANCH, O., ESCALANTE, A. A., KARIUKI, S., WOOTTON, J. & LAL, A. A. 2002. Evidence for intragenic recombination in Plasmodium falciparum: identification of a novel allele family in block 2 of merozoite surface protein-1: Asembo Bay Area Cohort Project XIV. *Molecular and Biochemical Parasitology,* 125**,** 163-171.

TANABE, K., SAKIHAMA, N., WALLIKER, D., BABIKER, H., ABDEL-MUHSIN, A.-M. A., BAKOTE'E, B., OHMAE, H., ARISUE, N., HORII, T., ROOTH, I., FÄRNERT, A., BJÖRKMAN, A. & RANFORD-CARTWRIGHT, L. 2007. Allelic dimorphism-associated restriction of recombination in Plasmodium falciparum msp1. *Gene,* 397**,** 153-160.

TATEM, A. J., GUERRA, C. A., KABARIA, C. W., NOOR, A. M. & HAY, S. I. 2008. Human population, urban settlement patterns and their impact on Plasmodium falciparum malaria endemicity. *Malaria Journal,* 7**,** 218.

TAYLOR, S. M., PAROBEK, C. M. & FAIRHURST, R. M. 2012. Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *The Lancet Infectious Diseases,* 12**,** 457-468.

TAZI, L. & AYALA, F. J. 2011. Unresolved direction of host transfer of Plasmodium vivax v. P. simium and P. malariae v. P. brasilianum. *Infection, Genetics and Evolution,* 11**,** 209-221.

TESSEMA, S. K., RAMAN, J., DUFFY, C. W., ISHENGOMA, D. S., AMAMBUA-NGWA, A. & GREENHOUSE, B. 2019. Applying next-generation sequencing to track falciparum malaria in sub-Saharan Africa. *Malar J,* 18**,** 268.

THERA, M. A., DOUMBO, O. K., COULIBALY, D., LAURENS, M. B., OUATTARA, A., KONE, A. K., GUINDO, A. B., TRAORE, K., TRAORE, I., KOURIBA, B., DIALLO, D. A., DIARRA, I., DAOU, M., DOLO, A., TOLO, Y., SISSOKO, M. S., NIANGALY, A., SISSOKO, M., TAKALA-HARRISON, S., LYKE, K. E., WU, Y., BLACKWELDER, W. C., GODEAUX, O., VEKEMANS, J., DUBOIS, M.-C., BALLOU, W. R., COHEN, J., THOMPSON, D., DUBE, T., SOISSON, L., DIGGS, C. L., HOUSE, B., LANAR, D. E., DUTTA, S., HEPPNER, D. G. & PLOWE, C. V. 2011. A Field Trial to Assess a Blood-Stage Malaria Vaccine. *New England Journal of Medicine,* 365**,** 1004-1013.

THOMSON-LUQUE, R., STABLER, T. C., FÜRLE, K., SILVA, J. C. & DAUBENBERGER, C. 2024. Plasmodium falciparum merozoite surface protein 1 as asexual blood stage malaria vaccine candidate. *Expert Review of Vaccines,* 23**,** 160-173.

THOMSON, R., PARR, J. B., CHENG, Q., CHENET, S., PERKINS, M. & CUNNINGHAM, J. 2020. Prevalence of Plasmodium falciparum lacking histidine-rich proteins 2 and 3: a systematic review. *Bulletin of the World Health Organization,* 98**,** 558-568F.

TIONO, A. B., KANGOYE, D. T., REHMAN, A. M., KARGOUGOU, D. G., KABORÉ, Y., DIARRA, A., OUEDRAOGO, E., NÉBIÉ, I., OUÉDRAOGO, A., OKECH, B., MILLIGAN, P. & SIRIMA, S. B. 2014. Malaria Incidence in Children in South-West Burkina Faso: Comparison of Active and Passive Case Detection Methods. *PLOS ONE,* 9**,** e86936.

TIONO, A. B., NÉBIÉ, I., ANAGNOSTOU, N., COULIBALY, A. S., BOWYER, G., LAM, E., BOUGOUMA, E. C., OUEDRAOGO, A., YARO, J. B. B., BARRY, A., ROBERTS, R., RAMPLING, T., BLISS, C., HODGSON, S., LAWRIE, A., OUEDRAOGO, A., IMOUKHUEDE, E. B., EWER, K. J., VIEBIG, N. K., DIARRA, A., LEROY, O., BEJON, P., HILL, A. V. S. & SIRIMA, S. B. 2018. First field efficacy trial of the ChAd63 MVA ME-TRAP vectored malaria vaccine candidate in 5-17 months old infants and children. *PLOS ONE,* 13**,** e0208328.

TOLLE, R., FRÜH, K., DOUMBO, O., KOITA, O., N'DIAYE, M., FISCHER, A., DIETZ, K. & BUJARD, H. 1993. A prospective study of the association between the human humoral immune response to Plasmodium falciparum blood stage antigen gp190 and control of malarial infections. *Infection and Immunity,* 61**,** 40-47.

TOPAZIAN, H. M., GUMBO, A., PUERTO-MEREDITH, S., NJIKO, R., MWANZA, A., KAYANGE, M., MWALILINO, D., MVULA, B., TEGHA, G., MVALO, T., EDWARDS, J. K., EMCH, M., PETTIFOR, A., SMITH, J. S., HOFFMAN, I., MESHNICK, S. R. & JULIANO, J. J. 2020. Asymptomatic Plasmodium falciparum malaria prevalence among adolescents and adults in Malawi, 2015–2016. *Scientific Reports,* 10**,** 18740.

TRAN, T. M., LI, S., DOUMBO, S., DOUMTABE, D., HUANG, C.-Y., DIA, S., BATHILY, A., SANGALA, J., KONE, Y., TRAORE, A., NIANGALY, M., DARA, C., KAYENTAO, K., ONGOIBA, A., DOUMBO, O. K., TRAORE, B. & CROMPTON, P. D. 2013. An Intensive Longitudinal Cohort Study of Malian Children and Adults Reveals No Evidence of Acquired Immunity to Plasmodium falciparum Infection. *Clinical Infectious Diseases,* 57**,** 40-47.

TRAVASSOS, M. A., NIANGALY, A., BAILEY, J. A., OUATTARA, A., COULIBALY, D., LYKE, K. E., LAURENS, M. B., PABLO, J., JASINSKAS, A., NAKAJIMA, R., BERRY, A. A., ADAMS, M., JACOB, C. G., PIKE, A., TAKALA-HARRISON, S., LIANG, L., KOURIBA, B., KONE, A. K., ROWE, J. A., MOULDS, J., DIALLO, D. A., DOUMBO, O. K., THERA, M. A., FELGNER, P. L. & PLOWE, C. V. 2018. Children with cerebral malaria or severe malarial anaemia lack immunity to distinct variant surface antigen subsets. *Scientific Reports,* 8**,** 6281.

US FDA 2020. FAQs on testing for SARS-CoV-2.

VAN DER AUWERA, G. A., CARNEIRO, M. O., HARTL, C., POPLIN, R., DEL ANGEL, G., LEVY-MOONSHINE, A., JORDAN, T., SHAKIR, K., ROAZEN, D., THIBAULT, J., BANKS, E., GARIMELLA, K. V., ALTSHULER, D., GABRIEL, S. & DEPRISTO, M. A. 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. 43**,** 11.10.1-11.10.33.

VAN DIJK, E. L., NAQUIN, D., GORRICHON, K., JASZCZYSZYN, Y., OUAZAHROU, R., THERMES, C. & HERNANDEZ, C. 2023. Genomics in the long-read sequencing era. *Trends in Genetics,* 39**,** 649-671.

VAN VLIET, A. H. M. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiology Letters,* 302**,** 1-7.

VARETA, J., HORSTMAN, N. A., ADAMS, M., SEYDEL, K. B., MCCANN, R. S., COHEE, L. M., LAUFER, M. K. & TAKALA-HARRISON, S. 2024. Genotyping Plasmodium falciparum gametocytes using amplicon deep sequencing. *Malaria Journal,* 23**,** 96.

VARO, R., BALANZA, N., MAYOR, A. & BASSAT, Q. 2021. Diagnosis of clinical malaria in endemic settings. *Expert Review of Anti-infective Therapy,* 19**,** 79-92.

VENKATESAN, M., AMARATUNGA, C., CAMPINO, S., AUBURN, S., KOCH, O., LIM, P., UK, S., SOCHEAT, D., KWIATKOWSKI, D. P., FAIRHURST, R. M. & PLOWE, C. V. 2012. Using CF11 cellulose columns to inexpensively and effectively remove human DNA from Plasmodium falciparum-infected whole blood samples. *Malaria Journal,* 11**,** 41.

VINAYAK, S., ALAM, M. T., MIXSON-HAYDEN, T., MCCOLLUM, A. M., SEM, R., SHAH, N. K., LIM, P., MUTH, S., ROGERS, W. O., FANDEUR, T., BARNWELL, J. W., ESCALANTE, A. A., WONGSRICHANALAI, C., ARIEY, F., MESHNICK, S. R. & UDHAYAKUMAR, V. 2010. Origin and Evolution of Sulfadoxine Resistant Plasmodium falciparum. *PLOS Pathogens,* 6**,** e1000830.

VOLKMAN, S. K., SABETI, P. C., DECAPRIO, D., NEAFSEY, D. E., SCHAFFNER, S. F., MILNER, D. A., DAILY, J. P., SARR, O., NDIAYE, D., NDIR, O., MBOUP, S., DURAISINGH, M. T., LUKENS, A., DERR, A., STANGE-THOMANN, N., WAGGONER, S., ONOFRIO, R., ZIAUGRA, L., MAUCELI, E., GNERRE, S., JAFFE, D. B., ZAINOUN, J., WIEGAND, R. C., BIRREN, B. W., HARTL, D. L., GALAGAN, J. E., LANDER, E. S. & WIRTH, D. F. 2007. A genome-wide map of diversity in Plasmodium falciparum. *Nature Genetics,* 39**,** 113-119.

WALKER, I. S. & ROGERSON, S. J. 2023. Pathogenicity and virulence of malaria: Sticky problems and tricky solutions. *Virulence,* 14**,** 2150456.

WANG, Y., LI, M., GUO, W., DENG, C., ZOU, G. & SONG, J. 2022. Burden of Malaria in Sao Tome and Principe, 1990–2019: Findings from the Global Burden of Disease Study 2019. *International Journal of Environmental Research and Public Health,* 19**,** 14817.

WARRELL, D. A. 2017. Clinical features of malaria. *Essential Malariology, 4Ed.* CRC Press.

WATSON, O. J., VERITY, R., GHANI, A. C., GARSKE, T., CUNNINGHAM, J., TSHEFU, A., MWANDAGALIRWA, M. K., MESHNICK, S. R., PARR, J. B. & SLATER, H. C. 2019. Impact of seasonal variations in Plasmodium falciparum malaria transmission on the surveillance of pfhrp2 gene deletions. *eLife,* 8**,** e40339.

WEATHERALL, D. J. 2008. Genetic variation and susceptibility to infection: the red cell and malaria. *British Journal of Haematology,* 141**,** 276-286.

WELLEMS, T. E. & PLOWE, C. V. 2001. Chloroquine-Resistant Malaria. *The Journal of Infectious Diseases,* 184**,** 770-776.

WETTERSTRAND, K. A. 2020. *The Cost of Sequencing a Human Genome* [Online]. National Human Genome Research Institute. [Accessed].

WHITE, N. J. 2022. Severe malaria. *Malaria Journal,* 21**,** 284.

WHITE, N. J., PUKRITTAYAKAMEE, S., HIEN, T. T., FAIZ, M. A., MOKUOLU, O. A. & DONDORP, A. M. 2014. Malaria. *Lancet,* 383**,** 723-35.

WHITTAKER, C., SLATER, H., NASH, R., BOUSEMA, T., DRAKELEY, C., GHANI, A. C. & OKELL, L. C. 2021. Global patterns of submicroscopic *Plasmodium falciparum* malaria infection: insights from a systematic review and meta-analysis of population surveys. *The Lancet Microbe,* 2**,** e366-e374.

WHO 2015. *Guidelines for the treatment of malaria*, World Health Organization.

WHO 2023. *World malaria report 2023*, World Health Organization.

WITTENAUER, R., NOWAK, S. & LUTER, N. 2022. Price, quality, and market dynamics of malaria rapid diagnostic tests: analysis of Global Fund 2009–2018 data. *Malaria Journal,* 21**,** 12.

WU, J., WU, M., CHEN, T. & JIANG, R. 2016. Whole genome sequencing and its applications in medical genetics. *Quantitative Biology,* 4**,** 115-128.

YIMAM, Y., MOHEBALI, M. & ABBASZADEH AFSHAR, M. J. 2022. Comparison of diagnostic performance between conventional and ultrasensitive rapid diagnostic tests for diagnosis of malaria: A systematic review and meta-analysis. *PLOS ONE,* 17**,** e0263770.

ZHOU, G., AFRANE, Y. A., MALLA, S., GITHEKO, A. K. & YAN, G. 2015. Active case surveillance, passive case surveillance and asymptomatic malaria parasite screening illustrate different age distribution, spatial clustering and seasonality in western Kenya. *Malaria Journal,* 14**,** 41.

ZHU, S. J., ALMAGRO-GARCIA, J. & MCVEAN, G. 2017. Deconvolution of multiple infections in Plasmodium falciparum from high throughput sequencing data. *Bioinformatics,* 34**,** 9-15.