Diversity and selection analyses identify transmission-blocking antigens as the optimal vaccine candidates in *Plasmodium falciparum*

Ilinca I. Ciubotariu[1], Bradley K. Broyles[1], Shaojun Xie[2], Jyothi Thimmapuram[2], Mulenga C. Mwenda[3], Brenda Mambwe[3], Conceptor Mulube[3], Japhet Matoba[4], Jessica L. Schue[5], William J. Moss[6,7], Daniel J. Bridges[8], Qixin He[1]*, and Giovanna Carpi[1,9]*

[1] Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA

[2] Bioinformatics Core, Purdue University, West Lafayette, Indiana, USA

[3] PATH-Malaria Control and Elimination Partnership in Africa (MACEPA), National Malaria Elimination Centre, Lusaka, Zambia

[4] Macha Research Trust, Choma, Zambia

[5] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[6] The W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[7] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[8] PATH, Lusaka, Zambia

[9] Purdue Institute of Inflammation, Immunology and Infectious Disease, West Lafayette, Indiana, USA

* Joint senior authors

Corresponding authors:

Correspondence to Giovanna Carpi (gcarpi@purdue.edu) and Qixin He (heqixin@purdue.edu).

Summary

Background

A highly effective vaccine for malaria remains an elusive target, at least in part due to the under-appreciated natural parasite variation. This study aimed to investigate genetic and structural variation, and immune selection of leading malaria vaccine candidates across the *Plasmodium falciparum*'s life cycle.

Methods

We analyzed 325 *P. falciparum* whole genome sequences from Zambia, in addition to 791 genomes from five other African countries available in the MalariaGEN Pf3k Rdatabase. Ten vaccine antigens spanning three life-history stages were examined for genetic and structural variations, using population genetics measures, haplotype network analysis, and 3D structure selection analysis.

Findings

Among the ten antigens analyzed, only three in the transmission-blocking vaccine category display *P. falciparum* 3D7 as the dominant haplotype. The antigens *AMA1, CSP, MSP1*$_{19}$ and *CelTOS,* are much more diverse than the other antigens, and their epitope regions are under moderate to strong balancing selection. In contrast, *Rh5*, a blood stage antigen, displays low diversity yet slightly stronger immune selection in the merozoite-blocking epitope region. Except for *CelTOS*, the transmission-blocking antigens *Pfs25*, *Pfs48/45*, *Pfs230*, *Pfs47*, and *Pfs28* exhibit minimal diversity and no immune selection in epitopes that induce strain-transcending antibodies, suggesting potential effectiveness of 3D7-based vaccines in blocking transmission.

Interpretations

These findings offer valuable insights into the selection of optimal vaccine candidates against *P. falciparum*. Based on our results, we recommend prioritizing conserved merozoite antigens and transmission-blocking antigens. Combining these antigens in multi-stage approaches may be particularly promising for malaria vaccine development initiatives.

Funding

Purdue Department of Biological Sciences; Puskas Memorial Fellowship; National Institute of Allergy and Infectious Diseases (U19AI089680).

Research in context

Evidence before this study

Decades of research on the most virulent malaria parasite, *Plasmodium falciparum*, have yielded multiple antigen candidates of pre-erythrocytic, blood-stage, and transmission-blocking vaccines in varying stages of development from preclinical development to more advanced clinical trials. The malaria vaccine, RTS,S/AS01, which was constructed using the C-terminal and NANP repeat region of the Circumsporozoite Protein (*CSP*) from the African reference strain 3D7, was approved and recommended for use in 2021. However, the vaccine's lower efficacy is likely a result of the genetic polymorphism of the target antigen shown by studies on natural variation in *CSP*. Similarly, another more recent pre-erythrocytic vaccine, R21/Matrix-M, showed great promise in clinical trials and was recommended in late 2023 by the WHO for use for prevention of malaria in children, but is also multi-dose and *CSP*-based. To maximize vaccine efficacy, it would be more strategic to first understand diversity and variation of antigens across the three types of vaccine classes, targeting various stages of the *P. falciparum* life cycle. Previous studies have reported analyses of vaccine candidate antigens but were mostly limited to pre-erythrocytic and blood-stage antigens, with less focus on transmission-blocking antigens. These studies revealed that most of the pre-erythrocytic and blood-stage antigens are of high diversity due to balancing selection, posing challenges for vaccine design to encompass the antigenic variation.

A search conducted on PubMed on April 1, 2024, for relevant published research which used the terms "malaria vaccine", "*Plasmodium falciparum*" [not "*vivax*"], "selection" and "diversity" yielded 48 studies between 1996 and the present day, with only 14 published studies in the past 3 years. This emphasizes the need for more studies assessing genetic diversity and selection of potential *P. falciparum* vaccine candidates to aid in more effective vaccine development efforts. A similar search with the terms "transmission-blocking vaccine", "malaria", "*Plasmodium falciparum*", not "*vivax*", "selection" and

"diversity" without any date or language restrictions revealed three relevant studies. This warrants future studies to explore transmission-blocking vaccines in this context.

Added value of this study

By comparing the genetic and structural analyses of transmission-blocking antigens with pre-erythrocytic and blood-stage antigens, we identify promising *P. falciparum* vaccine antigens characterized by their conservation with low balancing selection and the presence of infection/transmission-blocking epitopes, which are essential for informing the development of new malaria vaccines. This comprehensive workflow can be adopted for studying the genetic and structural variation of other *P. falciparum* vaccine targets before developing the next generation of malaria vaccines for effectiveness against natural parasite populations.

Implications of this study

Our suggested strategies for designing malaria vaccines include two possible approaches. We emphasize the development of a multi-stage vaccine that combines critical components such as anti-merozoite (*Rh5*) and transmission-blocking antigens (*Pfs25*, *Pfs28*, *Pfs48/45*, *Pfs230*). Alternatively, we suggest the creation of transmission-blocking vaccines specifically targeting *Pfs25*, *Pfs28* and *Pfs48/45*. These innovative approaches show great potential in advancing the development of more potent and effective malaria vaccines for the future.

Introduction

For preventing and controlling many infectious diseases, vaccines are vital tools that increase immunity, leading to reduced disease severity, incidence, and transmission. For malaria, a vector-borne disease caused by *Plasmodium* species parasites, vaccine strategies can be characterized by the three main parasite life-cycle stages that they target (Fig. 1). Firstly, pre-erythrocytic vaccines target sporozoites inoculated by *Anopheles* mosquitoes into the human skin before they reach the liver or infected hepatocytes, with the aim of preventing initial infection[1,2]. Some pre-erythrocytic subunit vaccine candidates[3,4] for *P. falciparum* include the first licensed malaria vaccine RTS,S/AS01[5], low-dose

R21/Matrix-M vaccine (currently approved for use in some countries[6]), and the full-length *CSP* [7]. Blood-stage vaccines predominantly target merozoites or infected erythrocytes to prevent replication and clinical illness, and some candidates include *AMA1*, *MSP1*, and *Rh5* [3,8]. Transmission-blocking vaccines target antigens expressed in the gametes and ookinete stages in the mosquito midgut; this vaccine type induces antibodies in humans that impair the parasite's development inside the mosquito and thus prevent subsequent transmission. Some examples of transmission-blocking vaccine candidates are *Pfs25*, *Pfs230*, and *Pfs48/45* [3,9,10].

To date, among the numerous preclinical and clinical trials of these classes of vaccines, RTS,S and R21 are the only two malaria vaccine licensed and recommended by the WHO. The former is approved for widespread use among children in sub-Saharan Africa and other regions with moderate to high *P. falciparum* malaria transmission[5,11-14] and the latter by regulators in Ghana, Nigeria, and Burkina Faso[6,15]. While this is a tremendous breakthrough for the field considering the complexity of the parasite life cycle and the correspondingly large number of protein-coding genes (>5,000) in the *Plasmodium* genome[16], there is a wealth of vaccine targets[17-20]. Despite the large number, about 60% of the genes code for putative proteins[21] and have unknown function[16,22], which poses an enormous challenge, and many of these 5300+ protein coding genes do not make attractive vaccine targets due to expression patterns, localization, high genetic variability, etc., and candidates must be experimentally validated for vaccine development[23]. Yet, despite decades of effort, noticeably little progress past early phase clinical trials for many of these other targets[24,25]. Furthermore, the RTS,S vaccine shows only modest efficacy ranging from 30% to 40% that wanes with time and requires four doses to reduce severe malaria by 30%[12,26-28]. This low efficacy has been attributed to the high genetic diversity and many polymorphic residues of the target antigen, *CSP* [12,24,29]. Notably, RTS,S was developed using a standard laboratory strain (*P. falciparum* 3D7), which originates from Africa but is not highly representative[30] and has low frequency within many parasite populations[29,31-33]. This underlines the need to consider *CSP*'s functional importance as well as the natural variation in the population, especially in epitope regions[25,34], when designing a vaccine. In recent years, vaccines based on regions of conserved or semi-conserved epitopes are emerging in other important yet antigenically diverse pathogens such as influenza virus and meningococcus[35-37].

With the release of thousands of publicly available *P. falciparum* whole genome sequences (WGS), it is possible to study the natural genetic and functional variation of targets for malaria vaccine candidate antigens, yet, to date few studies have done so[34,38-40]. Two recent studies reported comprehensive analyses of vaccine candidate antigens[25,41], but were focused primarily on pre-erythrocytic and blood-stage antigens, with limited inclusion of transmission-blocking antigens. Here, we aimed to characterize the natural genetic and structural variation of ten *P. falciparum* vaccine antigens spanning the three different vaccine classes, including transmission-blocking antigens. We focused on Africa which carries the heaviest burden of malaria, and augmented parasite genomic data from Zambia, where malaria continues to be endemic and a major public health with over 8 million cases and 1300 deaths reported in 2022[42].

Methods

Ethics Statement

The parents or legal guardians provided parental permission for study participants and this study was conducted with the approval of the Biomedical Research Ethics Committee from the University of Zambia (Ref 011-02-18) and from the Zambian National Health Research Authority.

Sample Selection and Processing

This work is a secondary analysis of a larger parent study focused on understanding malaria transmission across Zambia and establishing a baseline for parasite genetic metrics for this country that are informative about transmission intensity, genetic relatedness between parasites, and selection[43]. Briefly, dried blood spot (DBS) samples were collected as part of the 2018 Zambia Malaria Indicator Survey (MIS), which used a nationally representative two-stage stratified clustering sampling strategy to obtain samples across Zambia for the estimation of malaria prevalence[44]. The 2018 MIS included DBS samples from children under the age of 5 from 179 standard enumeration areas from across ten provinces, with high transmission provinces oversampled[44]. Due to undersampling in the low transmission provinces (Central, Lusaka and Southern), we augmented the DBS samples from the Zambia 2018 MIS with DBS

157 samples from Southern Province collected in the community surveillance program from the Southern and
158 Central Africa International Centers of Excellence for Malaria Research (ICEMR)[45]. These additional
159 samples temporally matched the MIS (from 2018-2019) and were collected as part of a longitudinal cohort
160 study conducted to understand sources of focal transmission from an area with an expected relatively
161 high number of incident malaria cases.
162 In total we included 459 PET-PCR positive *P. falciparum* samples from seven provinces[43] and additional
163 28 PCR positive *P. falciparum* samples from Southern Province for sequencing. The samples were
164 processed for genomic DNA extraction as previously described[43,46] and then genomic library preparation,
165 parasite enrichment and sequencing were carried out.
166
167 **Whole-Genome Sequencing**
168 To selectively enrich and sequence whole *P. falciparum* genomes from DBS we used a 4-plexed hybrid
169 capture method previously described[43]. Genomic library preparation, hybridization capture, and
170 sequencing was conducted at the Yale Center for Genomic Analysis (YCGA). Samples were sequenced
171 using 101 bp paired-end read sequencing on an Illumina NovaSeq 6000 with a target of 30 million reads
172 per sample.
173
174 **Acquisition of Additional Genomic Data**
175 To contextualize Zambian *P. falciparum* genome diversity within Africa, we included and analyzed 781
176 publicly available *P. falciparum* WGS data from 5 countries across West, Central, and East Africa
177 (Democratic Republic of the Congo, Ghana, Guinea, Malawi, and Tanzania) from the MalariaGEN Pf3k
178 project release 5.1 (https://www.malariagen.net/data/pf3k-5) and the MalariaGEN *P. falciparum*
179 Community Project (http://www.malariagen.net/resource/16). Raw Fastq files were downloaded from SRA
180 using pysradb (https://github.com/saketkc/pysradb)[47] and processed in the same way as the newly
181 sequenced WGS.
182
183 **Read Mapping and Variant Identification**

We identified *P. falciparum* genomic variation from whole genome sequence data using a bioinformatics pipeline previously described[48]. Briefly, the Illumina paired-end reads were trimmed for low-quality bases (Phred-scaled base quality <30) and mapped to the *P. falciparum* 3D7 reference genome (ftp://ftp.sanger.ac.uk/pub/project/pathogens/gff3/2015-08/Pfalciparum.genome.fasta.gz) using BWA-MEM v0.7.17[49]. Aligned reads for the 1248 samples (487 from Zambia and 781 from Pf3k) were marked for duplicates and sorted using Picard Tools 2.20.8. For variant calling only samples with >50% of the regions spanning the selected antigens (Table 1) with >5X read coverage were included, resulting in a total of 1092 *P. falciparum* WGS samples. Variants were called using GATK v4.1.4.1 following best practices (https://software.broadinstitute.org/gatk/best-practices)[50,51]. We applied GATK Base Quality Score Recalibration using default parameters and using variants from the *P. falciparum* crosses 1.0 release as a set of known sites (ftp://www.malariagen.net/data/pf-crosses-1.0)[52]. GATK HaplotypeCaller in GVCF mode was used to call single-sample variants (*ploidy* 2 and *standard-min-confidence-threshold for calling* = 30), followed by GenotypeGVCFs to genotype the cohort. We restricted our analyses to variants from the 14 nuclear chromosomes only, excluding variants from telomeric and hypervariable regions (ftp://ngs.sanger.ac.uk/production/malaria/pf-crosses/1.0/regions-20130225.onebased.txt). We further removed variants with VQSLOD <0 and SNPs with missingness >0.2. After variant filtering, we scored 426,757 genome-wide SNPs across the 1092 samples.

Variants were functionally annotated with SnpEff (version 4.3t) for genomic variant annotations and functional effect prediction[53].

Multiplicity of Infection

To ensure high quality haplotype reconstruction, we first determined the multiplicity of infection (MOI) for each sample using within-sample F statistic (F_{WS})[54] implemented in the R moimix package (available at https://github.com/bahlolab/moimix). F_{WS} scores infections on a continuous scale of multiplicity, with infections having an F_{WS} > 0.95 considered clonal (MOI=1), those with F_{WS} < 0.95 and > 0.8 containing two strains (MOI=2), and those with F_{WS} < 0.8 harboring more than two strains (MOI>2). F_{WS} metric derived from the 426,757 genome-wide SNPs were used to stratify samples into single infections, two infections or more than two infections.

Selection of Vaccine Candidate Antigens

Antigens were selected based on their inclusion[14] in either two licensed and recommended malaria vaccines (e.g. *CSP*)[28], clinical development in Phase I (e.g. *AMA1*, *Rh5*, *MSP1*, *Pfs25*, *Pfs48/45*, and *Pfs230*)[3], clinical development in Phase II (e.g. *Rh5* and *Pfs230*)[55-57], or their recognition as a promising vaccine target (e.g. *Pfs47*, *Pfs28*, and *CelTOS*)[58,59]. The ten selected antigens encompassed all parasite life cycle stages (Table 1 and Fig. 1), and genomic information and coordinates are provided in Table S1.

Haplotype Reconstruction

Vaccine target regions (Table S1) were extracted from the filtered VCF file (VQSLOD <0) in BED format using vcftools (version 0.1.16). Haplotypes were reconstructed using a custom code available at https://github.com/giocarpi/Pf_capture_div_code, which converted the VCF file to FASTA format, incorporating all variants (SNPs and INDELs) within coding sequences (CDS) relative to the *P. falciparum* 3D7 reference gene sequences. For a given variant position in a sample, a minimum read depth was required to be 5 for samples of MOI=1 and a minimum read depth of 10 for samples of MOI=2. A minimum minor allele proportion for a heterozygous call to be made was set to 0.2. If the mean frequency of minor alleles was below 0.4, these minor alleles constituted a minor haplotype, which would be included in the following analyses; otherwise, the minor haplotype was not retained (according to paper[25] and code). Samples failing read depth criteria were excluded from downstream analysis, as were samples containing one or more 'N' nucleotides.

Genetic Diversity Analysis

The software DNA Sequence Polymorphism (DnaSP) version 6.12.03 was used to calculate the number of unique haplotypes (h) of each respective vaccine candidate antigen in each country, the number of polymorphic sites (S), and the average number of pairwise nucleotide differences (k) between samples from within and between countries[60]. Furthermore, we calculated the haplotype diversity (Hd) for each antigen, with 1 representing a population in which every haplotype is unique and 0 equating a population where each haplotype is identical. Nei's nucleotide diversity statistic[61] (π) within each country and among

all samples was also calculated, which represents the average number of nucleotide differences between haplotypes drawn from the same population. *Tajima's D* statistic values[61-63] were calculated using Molecular Evolutionary Genetics Analysis (MEGA X) version 10.1.7 to measure selection (*Tajima's D* values that are approximately zero occur in neutral conditions, >1 indicates balancing selection, and <1 represents purifying selection). Gaps in the multiple sequence alignments were treated as complete deletion. To assess the genealogical relationships between the haplotypes found in Pf3k from African countries and those from Zambia for each vaccine candidate antigen of interest, we constructed haplotype networks based on the method by Templeton, Crandall, and Sing (TCS) using PopArt version 1.7[64,65].

Structural Analyses

Protein structures used for the downstream analyses were either selected from existing solved ones or generated from AlphaFold2[66]. PDB structures were retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) database on 5/23/2023. The final solved structures to be used for each protein were selected based on three criteria: 1) the structure was based on *P. falciparum* 3D7 reference gene sequences; 2) most complete in terms of gene coverage; 3) resolved as many residues with polymorphisms as possible. Often, the SNP-containing residues occurred in solvent-exposed loop regions that were not resolved in published Protein Data Bank (PDB) structures. In these cases (i.e., *AMA1*, *Rh5*, *Pfs230* (D1, D2)*, Pfs25, MSP1*$_{19}$) and the case of no solved structures for *P. falciparum* (i.e., *Pfs47, CelTOS*), we assessed the qualities of AlphaFold2's predictions of these structures, that is whether the RMSD is small or < 3 Å[67,68] (Fig. S3). *Pfs230* (D1, D2), *Pfs25, Pfs4845,* and *MSP1*$_{19}$ were successfully folded by AlphaFold2, agreeing with the native fold at or below 2 Å Root Mean Square Deviation (RMSD). *Rh5* prediction was unable to meet this threshold, therefore we proceeded with the most complete solved structure of *Rh5* available. *AMA1* structures were missing domain III, as well as missing resolution of many residues in a loop of domain II (residues 351–389). AlphaFold2 was able to confidently predict domain III and the overall structure (RMSD 0.36), but the missing loop residues of domain II were of low confidence, and the placement of the loop disagreed with the resolved residues of this loop in solved

structures. For this reason, we proceeded with an AlphaFold2 model of *AMA1* including domain III but excluding the low confidence loop region of domain II.

Pfs47 had no deposited structures. Hence, it was predicted entirely with AlphaFold2 with good confidence (Fig. S3). *CelTOS* had no deposited structures for *P. falciparum,* but the *P. vivax* homolog has been experimentally solved. We assessed a structure modeled with AlphaFold2 for *P. falciparum CelTOS* and verified its fit to *P. vivax* (RMSD 1.59), trimming off residues outside the solved *P. vivax* structure.

AlphaFold2 structures were predicted on 3D7 reference sequences, utilizing the ColabFold[69] pipeline (https://colab.research.google.com/github/sokrypton/ColabFold). Multiple sequence alignments were performed with MMseqs2 searching through UniRef+Environmental datasets. For each protein, three different structure models were run with six recycles and early stops at 100 PLDDT. Templates and AMBER relaxation were set to “off”. Models of rank one were chosen as final structures. For predicted structures with solved references that needed loop or domain modeling, an agreement between predicted and solved reference of 2 Å RMSD or less was deemed suitable for downstream analysis. For predicted structures with no solved reference, a convergence of three models to a structure within 2 Å RMSD between three models was assessed as a confident prediction model.

All the available solved antibody-antigen structures of seven of ten vaccine target proteins were obtained from RCSB (as of 7/12/23). Using a custom R script, we extracted interface residues between the antibody and target protein that occur within 5 Å of each other. *Pfs47* lacks antibody-antigen structures, so we have included a known linear epitope (see Table 1). These interface residues were sorted into bins for visualization purposes at 0, 1, 2, or 3+ antibody interfaces, to represent common epitopes vs. infrequent and non-targeted residues. We consolidated identical antibodies with multiple structures into a single record. This ensures that the count of epitopes involving each interface residue only includes unique antibody contacts.

Calculation of 3D population genetics measures

To calculate spatially derived nucleotide diversity (π*) and Tajima’s D (D*) values, we utilized the BiostructMap Python[70,71] package, installed from https://github.com/andrewguy/biostructmap. The package provides a method for mapping DNA sequence data to residues on protein structures, defining a

spatial neighborhood of the residue, and calculating the statistics using a three-dimensional (3D) sliding window approach. In our study, we selected a 3D window with a radius of 15 Å, which represents the maximum surface of an antibody-antigen binding site[72]. We chose not to apply a relative solvent accessibility (RSA) filter and included all the residues within the radius of 15 Å, because polymorphisms at buried residues could have structural effects on an epitope, and thus should be reflected in calculated diversity and selection pressure.

Role of the Funding Sources

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to submit for publication.

Results

Enhancing *P. falciparum* genomic diversity from African countries

We augmented 241 *P. falciparum* whole genome sequences previously reported[43] with an additional 84 genomes from Zambia, a malaria endemic country in South-Central Africa currently lacking *P. falciparum* genomic representation in public databases. Combining these data with publicly available WGS sequences from the MalariaGEN *P. falciparum Community Project Consortium*[73] (http://www.malariagen.net/resource/16), WGS data of 1248 *P. falciparum* samples from 6 African countries were processed to obtain gene sequence haplotypes for ten antigens, with the final data set consisting of 1092 samples after quality filtering (Table 1, Table 2) for population genetic analyses. To ensure high-quality haplotype reconstruction, the multiplicity of infections (MOI) was determined using F_{WS}[54], and samples with a high multiplicity of infections (MOI>2) were removed. Overall, 44.2% (483) of 1092 African field samples were found to harbor more than two infections, with Malawi having the highest percentage (50.8%), followed by Zambia (48.9%) (Fig. S2). Moreover, 17.6% of samples (192) had MOI=2 and 38.2% (417) were characterized as MOI=1 (Fig. S2). Additionally, there was a mean of 457 (range:196-579) sequences per antigen when considering full-length genes for analysis, and a mean of

591 (range: 555-623) sequences per antigen when considering partial genes for those with specific regions of interest related to vaccine targets (Table 2).

Antigen haplotypes lack geographic structure

To assess the genealogical relationships amongst sequence haplotypes obtained from our newly derived Zambian sequences and those from other African countries, we used TCS network analysis for the ten antigens. Full length genes, as well as known vaccine target regions, were examined and compared (Fig. 2 and 3). In general, haplotypes did not exhibit clustering, irrespective of vaccine antigen type and geographic origin. This observation held true whether the analysis considered the entire gene or specific functionally relevant domains, suggesting a lack of geographic structure. (Fig. 2 and 3).

We identified the following percent of unique haplotypes across all countries for individual full-length genes: 41.2% (100/243) for *CSP*, 62.5% (362/579) for *AMA1*, 6.8% (28/413) for *Rh5*, 89.3% (175/196) for *MSP1*, 0.9% (5/552) for *Pfs25*, 4.5% (24/539) for *Pfs48/45*, 89.1% (383/430) for *Pfs230*, 12.3% (64/522) for *Pfs47*, 4.7% (25/531) for *Pfs28*, and 33.9% (190/560) for *CelTOS*. For certain genes, we examined specific regions of interest related to vaccine targets and identified the following percent of unique haplotypes across all countries: 17.2% (107/623) for *CSP* C-terminal, 52.8% (308/583) for *AMA1* (Domains I-III), 2.8% (17/612) for *MSP1* (C-terminal region: 19), 2.2% (12/562) for *Pfs48/45* (C-terminal 6-Cys domain), 24.5% (149/609) for *Pfs230* (N-terminal prodomain and Domains I-III), and 1.4% (8/555) for *Pfs47* (Del2). We further separated unique haplotypes by country for each gene (see Table 2 for full diversity statistics results breakdown). To contextualize this specifically for Zambia, when considering unique haplotypes solely from this country, we observed a modest number across all genes: 44.3% (31/70) for *CSP* full-length and 26.5% (18/68) for *CSP* partial, 20.6% (65/315) for *AMA1* full-length and 20.2% (50/248) for *AMA1* partial, 18.8% (3/16) for *Rh5* full-length, 29.5% (51/173) for *MSP1* full-length and 25% (1/4) for *MSP1* partial, 33.3% (1/3) for *Pfs25* full-length, 25% (4/16) for *Pfs48/45 full-length* and 28.6% (2/7) for *Pfs48/45* partial, 22.6% (86/381) for *Pfs230* full-length and 17.2% (20/116) for *Pfs230* partial, 15.8% (6/38) for *Pfs47* full-length and 25% (1/4) for *Pfs47* partial, 23.8% (5/21) for *Pfs28* full-length, and 24.4% (32/131) for *CelTOS* full-length. The 3D7 vaccine strain haplotype was dominant for antigens *Pfs48/45* (full-length and partial), and *Pfs25* and *Pfs28* (both full-length) (Fig. 3). In contrast, for

the more diverse antigens i.e., *Pfs230*, *AMA1*, *CSP* and *MSP1,* the 3D7 vaccine strain haplotype was found at very low frequencies (Fig. 2 and 3).

Differing signals of selection, haplotype, and nucleotide diversity across antigen types for vaccine development considerations

To understand haplotype diversity and investigate evidence of selection, we derived Nei's nucleotide diversity (π), haplotype diversity (Hd) and *Tajima's* D statistics for each antigen. When taking full-length genes into consideration, overall, five antigens, *Pfs230*, *AMA1*, *CSP*, *MSP1*, and *CelTOS*, representing genes across the three categories of vaccine targets, exhibited the highest values of haplotype diversity (Hd >0.9) across all parasite populations analyzed (Fig. 4a, Table 2). Two transmission-blocking antigens, *Pfs47* (0.85) and *Pfs48/45* (0.73) showed a moderate-high level of haplotype diversity, with similar levels as those demonstrated by the blood-stage antigen *Rh5* (0.72) (Fig 4a, Table 2). The transmission-blocking antigen *Pfs28* exhibited low haplotype diversity (0.29), and the lowest haplotype diversity was observed for another transmission-blocking antigen, *Pfs25* (0.03) (Fig. 4a, Table 2). When considering antigens with specific regions of interest for vaccine development, the same trend was observed: *AMA1*, *CSP*, and *Pfs230* showed extremely high levels of haplotype diversity (0.99, 0.96, and 0.94, respectively), followed by *MSP1* with high haplotype diversity (0.77), while *Pfs48/45* had moderate haplotype diversity (0.43), and finally followed by *Pfs47* with the lowest haplotype diversity (0.35) (Fig. 4a, Table 2). Furthermore, across the parasite populations, when examining all of the full-length genes, we detected the highest levels of nucleotide diversity for *CelTOS* (0.0144) and *AMA1* (0.0137) (Fig. 4b, Table 2). The next highest levels of nucleotide diversity were shown for *CSP* (0.0067), *MSP1* (0.0033), and this was followed by similar levels observed for *Pfs47* (0.0015), *Pfs230* (0.0012), and *Pfs48/45* (0.0011) (Fig. 4b, Table 2). These values were tenfold higher than the nucleotide diversity observed for genes *Rh5 and Pfs28* (0.0009 and 0.0005, respectively, and one hundred-fold lower than the nucleotide diversity observed for *Pfs25* of 0.00001 (Fig. 4b, Table 2). When considering partial regions of interest for vaccine development for the included genes, a similar pattern was observed; namely, *CSP* and *AMA1* had the highest nucleotide diversity (0.016 and 0.014, respectively), while *MSP1*, *Pfs47*, and *Pfs230* had tenfold lower nucleotide diversity (0.0057, 0.0024, and 0.0012), and *Pfs48/45* had the lowest recorded nucleotide

diversity of 0.0011 among the genes for which partial regions were studied (Fig. 4b, Table 2). When examining specimens from Zambia alone, haplotype and nucleotide diversity aligned with the observed ranges in the parasite populations across the three categories of vaccine targets.

To assess SNP neutrality across genes, we conducted *Tajima's D* statistical test, comparing average pairwise differences between samples and the total number of segregating sites (Fig. 4c, Table 2). *Tajima's D* values of 0 indicate neutrality in constant-size populations, negative values suggest purifying selection, and positive values suggest balancing selection. When examining full-length antigens, *AMA1* showed the strongest signature of balancing selection, with *CelTOS* displaying neutrality and weaker signs of balancing selection, *CSP* exhibiting neutrality with signs of purifying selection, and *MSP1* similarly showing neutrality with signs of purifying selection (Fig. 4c, Table 2). Conversely, *Pfs28*, *Pfs230*, *Pfs47*, *Pfs25*, *Rh5, Pfs48/45* (in decreasing order) showed clear signatures of purifying selection (Fig. 4c, Table 2). When inspecting antigens with partial regions of interest for vaccine development, we found that *AMA1* still demonstrated strong balancing selection, while *MSP1* and *CSP* showed moderate signs of balancing selection, contrasting with the full-length gene signatures (Fig. 4c, Table 2). On the other side of the spectrum, *Pfs230* and *Pfs48/45* showed strong signs of purifying selection, while *Pfs47* displayed neutrality with moderate levels of purifying selection (Fig. 4c, Table 2). Overall, there was no clear correlation between balancing selection signals with vaccine target type, although a general pattern was observed with the sexual stage antigens mostly showing evidence of purifying selection.

Protein structures and balancing selection on known epitopes confirm transmission-blocking antigens are not under immune selection

To facilitate the search for effective malaria vaccine targets, we applied three structure-based analyses to the structures of ten vaccine candidate proteins (Table 1): nucleotide diversity of a focal residue's spatial neighborhood (π*), *Tajima's D* of a focal residue's spatial neighborhood (D*), and summarization of experimentally solved antibody-antigen structures (Fig. 5). We then explored if there was a high concentration of sequence diversity and balancing selection signatures within certain domains or patches of each protein. Additionally, we sought to determine if these regions correlated with known epitopes.

The 3D nucleotide diversity, π*, of *CSP, AMA1, CelTOS, and MSP1$_{19}$* showed the overall extreme diversity of the molecules, which agrees with the 2D analysis. The 3D selection measure, *D**, however, had greater variability across the molecules. *CSP* C-terminal domain exhibited uniformly high diversity and selection, with 95% of spatial windows in regions of π* ≥ 0.01 (maximum π* = 0.046), and 85% and 23% of spatial windows under moderate to strong balancing (D* ≥ 1) and strong balancing selection (D* ≥ 2, maximum D* = 3.34) respectively. The structurally characterized epitopes had similar diversity and selection pressures as the full *CSP* C-terminal domain (Fig. S4). The *D**, however, could be artificially inflated for parts of the surface, including the conserved β-sheet epitope[74]. Due to the small size of *CSP* C-terminal domain, the 15Å spatial window implemented in biostructmap could include occluded residues on the opposite surface with high polymorphism.

The known epitope of *AMA1*, structurally characterized from antibody 1F9[75] displayed higher diversity and stronger selection signature than the full *AMA1* structure. Overall, *AMA1* was found to have 58% of spatial windows in regions of π* ≥ 0.01 (maximum π* = 0.058), and 67% were under moderate to strong balancing selection (maximum D* = 3.38), while the epitope of 1F9 shows 100% of their spatial windows in regions of high diversity and high selection. Additionally, we recovered the signature of *AMA1* displaying a distinctive side of the surface with heightened selection, and another side with no selection, termed the “silent face” of *AMA1*[76,77] (Fig. 5).

CelTOS showed a distinctive region of low diversity and selection corresponding to its inferred dimerization interface[78] (Fig. 5). The overall structure of *CelTOS* was found to have 66% of spatial windows with π* ≥ 0.01 (maximum π* = 0.041), and 55% of spatial windows under moderate to strong balancing selection (maximum D* = 2.01) (Fig. S4).

MSP1$_{19}$, *Pfs48/45,* and *Rh5* contained regions under moderate balancing selection (D* between 1-2). *MSP1$_{19}$* domain 1 displayed stronger selection (mean D* = 0.67) than domain 2 (mean D* = 0.34). Among the antibodies with characterized structures, 42D6 is the most potent and targets domains 1 and 2 at the same time[79]. One region in *Rh5* was found to be under selection, which is near its *Basigin* binding interface[80]. Interestingly, this region was shown to be targeted by infection-blocking antibodies[80-82]. For *Pfs48/45*, we found heightened diversity in domains 2 and 3 compared to domain 1, both of which are

binding sites of infection-blocking antibodies[83]. Only part of domain 2 was found to be under moderate balancing selection. Interestingly, none of the known epitopes overlap with the selected region.

Among the remaining four vaccine-candidate proteins encompassing the transmission-blocking vaccine class, two had regions under weak balancing selection ($0 \leq D^* < 1$): *Pfs230* (maximum $D^* = 0.66$) and *Pfs47* (maximum $D^* = 0.75$), and two, *Pfs25* and *Pfs28,* were found to be entirely under purifying selection ($D^* \leq 0$) with maximum D^* of (-0.79) and (-0.78), respectively (Fig. S4).

Discussion

An ideal malaria vaccine target would be conserved with low balancing selection and containing infection/transmission blocking epitopes that induce strain-transcending antibodies. Against these criteria, ten malaria vaccine antigens targeting sporozoite (n=1), merozoite (n=3) and sexual lifecycle stages (n=6), were assessed against newly generated Zambian parasite genomes, together with publicly available *P. falciparum* WGS spanning five other countries in Africa.

The frequency of unique haplotypes varied across merozoite antigens when considering their full lengths, ranging from a high of 89.3% for *MSP1* to a low of 6.8% for *Rh5*. Notably, transmission-blocking antigens, including *Pfs25*, *Pfs48/45*, *Pfs28*, and *Pfs47*, exhibited markedly lower frequencies of unique haplotypes (ranging from 0.9% to 12.3%) compared to the merozoite antigen *AMA1* (62.5%) and the pre-erythrocytic antigen *CSP* (41.2%). *Pfs230*, another transmission-blocking antigen, demonstrated a frequency of unique haplotypes similar to the highest recorded, *MSP1* (88%). The high number of haplotypes is expected due to the length of the protein (>3100 AAs and 14 domains) and many haplotypes are the result of rare polymorphisms[25]. *CelTOS*, another sexual stage antigen studied, displayed a moderate frequency of unique haplotypes at 33.9%. When focusing on specific regions targeted for vaccine development, the merozoite antigen *AMA1* (Domains I-III) exhibited the highest frequency of unique haplotypes (52.8%). In contrast, other antigens displayed lower frequencies: *Pfs230* (N-terminal prodomain and Domains I-III) at 24.5%, *CSP* (C-terminal) at 17.2%, *Pfs47* (Del2) at 6.1%, *MSP1* (C-terminal region: 19) at 2.8%, and *Pfs48/45* (C-terminal 6-Cys domain) at 2.1%. These findings suggest a lack of a strict pattern based on vaccine types, although transmission-blocking antigens

generally exhibited lower frequencies of unique haplotypes compared to other vaccine antigens. Additionally, the observed low frequency unique haplotypes may be attributed to the high malaria transmission intensity in these regions of Africa.

Antibodies with varying effectiveness have been structurally characterized for seven of ten antigens (see Table 1). Our genetic and structural analyses show that the human immune system plays an important role in shaping the haplotype and nucleotide diversity of *P. falciparum* antigens, thus influencing the potency and coverage of the antibodies. Based on the results, we identified five transmission-blocking antigens (*Pfs28*, *Pfs25*, *Pfs48/45*, *Pfs47*, and *Pfs230*) and one blood-stage antigen (*Rh5*) as promising candidates for developing broad-coverage vaccines. *Pfs28* and *Pfs25* demonstrate markedly lower haplotype and nucleotide diversity than other antigens (Table 2; see also [40,84]) because they are not expressed within human hosts and therefore they are not subjected to immune selection[85,86]. The drawback of no natural exposure to the human immune system from these two antigens is the lack of antibody-boosting effect upon subsequent natural infections. Nevertheless, the low diversity and no signs of selection warrant *Pfs28* and *Pfs25* as promising potential broad-coverage transmission blocking vaccines. Additionally, combining antibodies raised to these antigens results in synergistic transmission blocking activity, highlighting the possibility of a multi-component vaccine using *Pfs25* and *Pfs28*[86]. Other transmission-blocking (sexual stage) antigens, such as *Pfs48/45*, *Pfs47*, and *Pfs230*, also represent promising broad-coverage vaccine candidates, with the added advantage of antibody-boosting upon subsequent exposure, and synergistic transmission blocking activity when combining antibodies of *Pfs48/45* and *Pfs230*[87]. In addition, despite brief exposure to the immune system during merozoite invasion[88], *Rh5* seems to be a more promising immunogen due to a less diverse haplotype population and potent antibodies with cross-reactivity[80,81].

For the pre-erythrocytic antigen, *CSP* C-terminal domain was shown to be highly polymorphic in the current study [Fig. 5; see also [74]]. Notably, the 3D7 haplotype of *CSP* C-terminal domain represents only 3.7% of haplotypes circulating in our study (Fig. 2 and [29]). Earlier studies have shown that 3D7 haplotype-based RTS,S vaccine efficacy decreases with increased mutational distance from the 3D7 haplotype[33]. *CSP* C-terminal domain has two characterized B-cell epitopes, and both are capable of weak inhibition of hepatocyte invasion[74]. One overlaps with T-cell epitopes Th2R and Th3R and contains many

polymorphic residues, which limits the cross-reactivity of antibodies to other *P. falciparum* strains[74,89]. A second epitope on a conserved β-sheet of *CSP* C-terminal domain was shown to have broad cross-reactivity[74] but much weaker invasion inhibition than antibodies targeting NANP repeats[74,90]. Overall, the C-terminal domain of *CSP* might not be sufficient to elicit full protection for the entire parasite population, namely in generating broad cross-reactive antibodies capable of potent inhibition of infection.

The blood-stage antigens, *AMA1* (Domains I-III) and *MSP1*$_{19}$, are more diverse than the transmission-blocking antigens, with a significant portion of their surface under at least moderate balancing selection (Fig. 5). These results are consistent with previous studies investigating across Africa and globally, with work revealing balancing selection and high haplotype diversity[25,39,91]. Additionally, the 3D7 haplotype was not dominant for these antigens (Fig. 2). Many polymorphic residues in *AMA1* 1F9's epitope found in our study were shown to abrogate 1F9 binding (H 200 [E,R,L,V], D 204 [K,N], I 225 [N])[75], implying that a single vaccine is unlikely to confer protection to multiple divergent haplotypes that exist in natural parasite populations. Not included in the present structural analysis is a highly flexible loop of D2, which has a conserved epitope (targeted by 4G2 monoclonal antibody[92]). Among the antibodies with characterized structures for *MSP1*$_{19}$, 42D6 is reported to be the most potent and targets domain 1 and 2 simultaneously[79]. We found three polymorphic residues in the epitope of 42D6, two of which (Pf3D7 E1671K, L1692F) have been confirmed not to significantly impact binding, while the effect on binding and potency of the third residue (Pf3D7 G1623R) remains uncharacterized and thus unknown. In contrast, *Rh5* is uniquely well-conserved among blood-stage antigens, as previously reported[93]. Most potent antibodies to *Rh5* were found around the binding interface of *Rh5* and human protein *Basigin*[81]. These antibodies often target patches with no polymorphisms in our data with one exception (Pf3D7 C203Y), which has been shown to not affect binding[81]. Other epitope regions near *Rh5*'s interface with *PfCyRPA* were not effective in blocking merozoite invasion[82]. Therefore, targeting the *Rh5-Basigin* interface has the potential as an effective vaccine against merozoite invasion.

For transmission-blocking antigens, *Pfs47* was previously shown to be non-essential for *P. falciparum* life cycle, as mutants lacking this protein were capable of normal fertilization[94]. In contrast, both *Pfs48/45* and *Pfs230* are essential antigens for *P. falciparum* life cycle[95,96]. Potent transmission-blocking antibodies have been characterized for *Pfs48/45*[83,97,98], *Pfs230 (D1, D2)*[99-101], and *Pfs47*[102]. As

reported above, we only found signals of weak or no balancing selection in these three antigens where potent antibodies are binding. Polymorphic residues on domain 3 of *Pfs48/45* in our dataset were found not to affect binding affinity severely and the antibodies retained nanomolar affinity ranges[97]. In *Pfs230*, some of the potent transmission-blocking antibodies raised from natural exposure bound to regions with no polymorphism in our study[100], while other potent antibodies raised from human vaccine trials bound to polymorphic regions[101]. Among the polymorphic sites, (Pf3D7 G605S, D714N) were shown to affect binding, while (Pf3D7 E654K, E655V, and A699T) did not. The remaining set of polymorphic residues (Pf3D7 N616K, T656N, V727I) segregate at low frequencies and have not been tested for changes in binding affinities. *Pfs47* also contained polymorphic residues in the epitope of a potent antibody[102]. However, the influence of these polymorphisms has not been characterized. Immunization to both *Pfs28* and *Pfs25* have been shown to raise potent transmission-blocking antibodies[86,103-112]. The impact of polymorphic residues (Pf3D7 L63V, V132I) of *Pfs25* on the binding affinities of the antibodies has not yet been characterized.

While a part of transmission-blocking antigens, *CelTOS* is critical to liver invasion by sporozoites during the pre-erythrocytic stage[78]. The heightened balancing selection pressure and nucleotide diversity of *CelTOS* compared to the other transmission-blocking target antigens reflects its intense interaction with the human immune system. While no antibodies have been structurally characterized against *CelTOS* at the time of this study, it has been shown that mice immunized with *CelTOS* can raise potent transmission-blocking antibodies[113]. However, with nearly the entire surface exposed regions of *CelTOS* under moderate balancing selection, it is unlikely to provide broad-coverage vaccinations.

We acknowledge that this study is not without limitations. Firstly, it was a secondary analysis of a broader study[43], and we did not intend to do an exhaustive study of all vaccine candidate genes. We chose to incorporate representative genes across the three classes of vaccine candidates with a focus on transmission-blocking antigens that are currently under clinical vaccine development. Identifying potential vaccine targets from the proteome is beyond our study scope. Secondly, our study used samples mainly from sub-Saharan Africa with a particular focus on Zambia where genomic data were lacking. However, we acknowledge the absence of representation from Asia or South America. While we recognize that a study with a globally representative set of samples would be more comprehensive, we believe that this

investigation is still useful for understanding vaccine development in the context of Africa and Zambia specifically, a country that remains highly malaria endemic. Lastly, the analyses of full-length and partial genes should not be directly compared as they incorporated varying numbers of samples (the partial genes typically included more samples than the full-length genes). The partial gene analyses for regions of interest and the corresponding 3D structural analyses are more relevant for informing vaccine development.

In conclusion, this study offers valuable insights into selecting *P. falciparum* vaccine antigens with regions under low balancing selection and the presence of infection/transmission-blocking epitopes, which are essential for informing the development of new malaria vaccines. Based on these findings, we recommend the following strategies for vaccine design: 1) a multi-stage vaccine encompassing anti-merozoite (*Rh5*) and transmission-blocking antigens (*Pfs25*, *Pfs28*, *Pfs48/45*, *Pfs230*), and 2) transmission-blocking vaccines targeting combinations of *Pfs25*, *Pfs28*, and *Pfs48/45*. These approaches hold promise for the development of more effective malaria vaccines in the future.

Contributors

I.I.C, Q.H. and G.C. conceived and designed the study. I.I.C, B.K.B, S.X., Q.H and G.C. generated and collected data. S.X., I.I.C., B.K.B., Q.H., and G.C. analyzed data and interpret results. J.T., Q.H. and G.C. supervised analysis. M.C.M., B.M., C.M., J.M., J.L.S. W.J.M., and D.J.B. contributed to samples/materials. I.I.C., B.K.B., Q.H., and G.C wrote the manuscript. All authors read, edited, and approved the manuscript for publication.

Declaration of interests

All authors declare no competing interests.

Data Sharing

The raw sequence data are available in the NCBI Sequence Read Archive under BioProject PRJNA932927. Key analysis scripts can be accessed at https://github.com/giocarpi/Pf_capture_div_code.

Acknowledgements

We would like to thank collaborators at the Zambia National Malaria Elimination Centre, the study participants, and the Southern and Central Africa International Centers of Excellence for Malaria Research (ICEMR). We would also like to extend our gratitude to the communities and researchers of malaria endemic countries that enabled the collection and availability of the *P. falciparum* genomes used in this study and made publicly available through the MalariaGEN *P. falciparum* Community. The authors would also like to thank Abebe Fola for his contributions in extracting genomic DNA from DBS samples.

Additional information

Correspondence should be addressed to Giovanna Carpi and Qixin He.

Figures and Tables

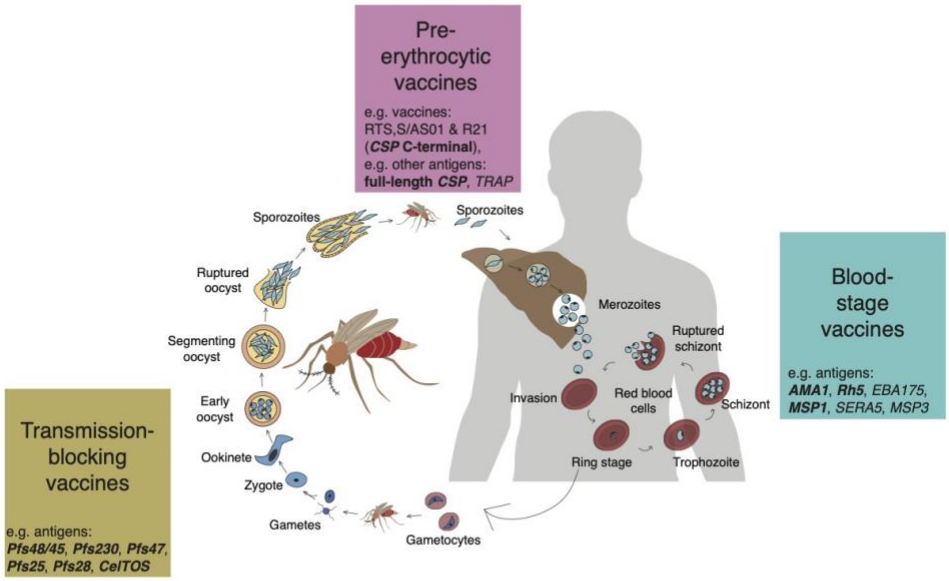
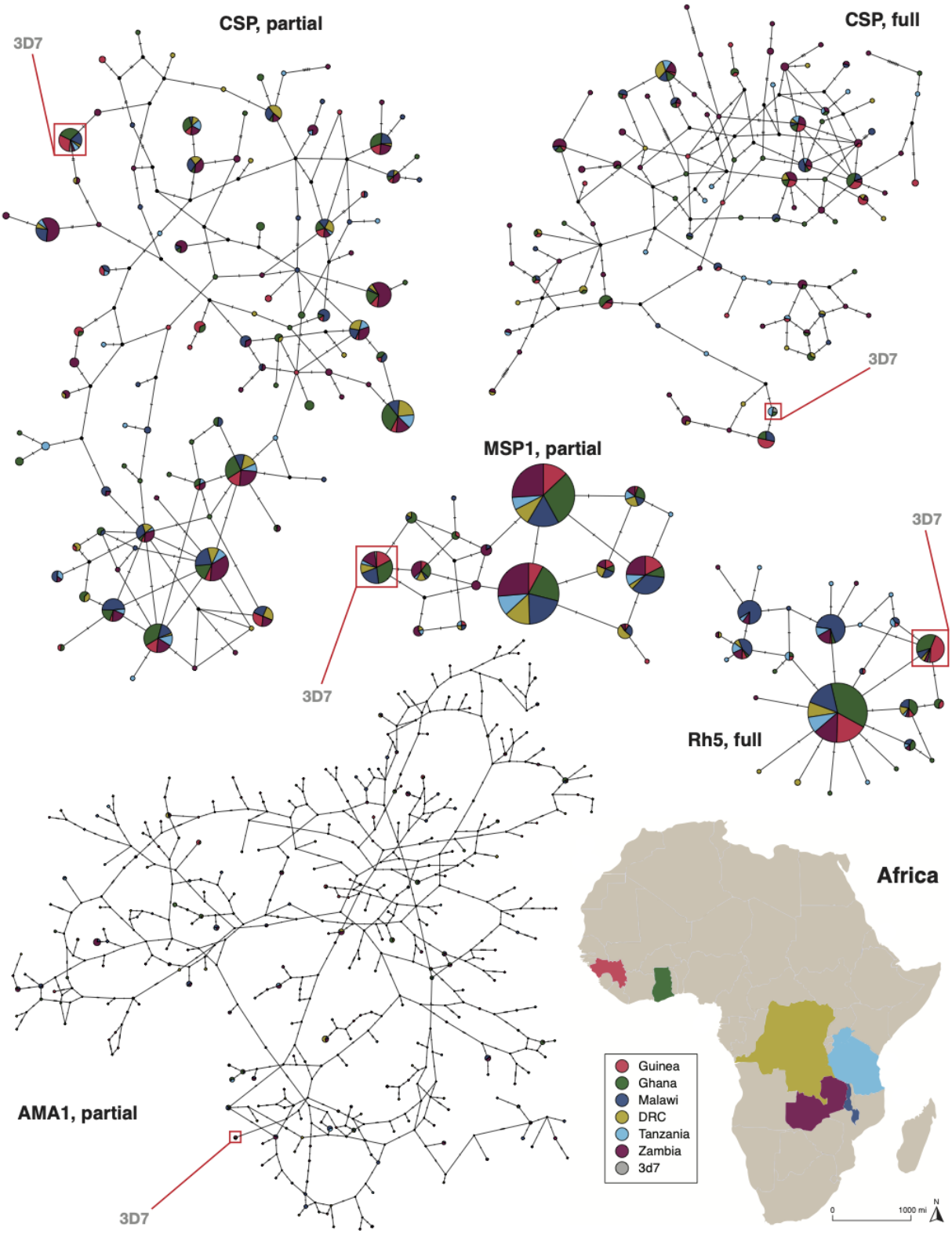


Figure 1. *Plasmodium falciparum* life cycle in the *Anopheles* mosquito vector and human host and corresponding targets for malaria vaccine development.

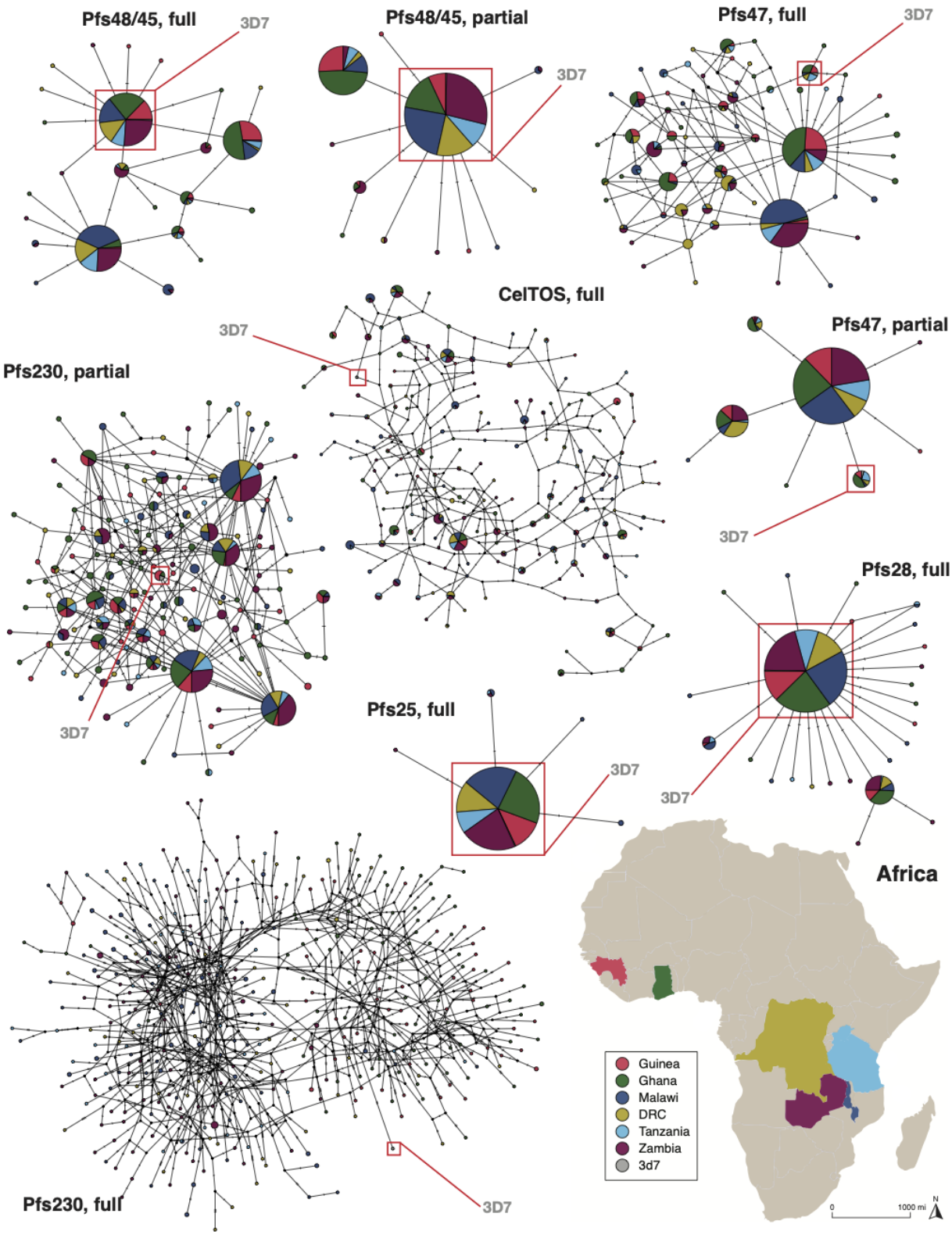
Vaccine strategies are broadly grouped by the *P. falciparum* life stage they target and include the categories: pre-erythrocytic, blood stage, and transmission-blocking. The pre-erythrocytic class of vaccines usually include candidates that target sporozoites or infected hepatocytes to prevent clinical initial infection and stop malaria parasites from re-entering the bloodstream. The blood stage vaccines target the parasite during replication in red blood cells in the human. Lastly, transmission-blocking vaccines target sexual and/or sporogonic development stages and aim to induce antibodies in humans that impair the parasite's development inside the mosquito midgut and thus subsequently prevent the transmission of sporozoites to humans. The antigens shown in bold were included in this study.



603

604

Figure 2. Haplotype network for malaria vaccine antigens (pre-erythrocytic and blood-stage proteins). Templeton, Crandall, and Sing (TCS) networks illustrate the diversity of the selected antigens. Circles represent unique nucleotide haplotypes and are scaled according to the frequency with which the haplotype was observed. The number of mutation differences that exist between each haplotype is depicted by the number of hatch marks on the network branches. Haplotype colors match the geographic origin of the samples depicted on the map (see map legend). The vaccine strain 3D7 (grey color, surrounded by a red box) is included for reference in each respective haplotype network.



613

614

Figure 3. Haplotype network for malaria vaccine antigens (transmission-blocking proteins). Templeton, Crandall, and Sing (TCS) networks illustrate the diversity of the selected antigens. Circles represent unique nucleotide haplotypes and are scaled according to the frequency with which the haplotype was observed. The number of mutation differences that exist between each haplotype is depicted by the number of hatch marks on the network branches. Haplotype colors match the geographic origin of the samples depicted on the map (see map legend). The vaccine strain 3D7 (grey color, surrounded by a red box) is included for reference in each respective haplotype network.

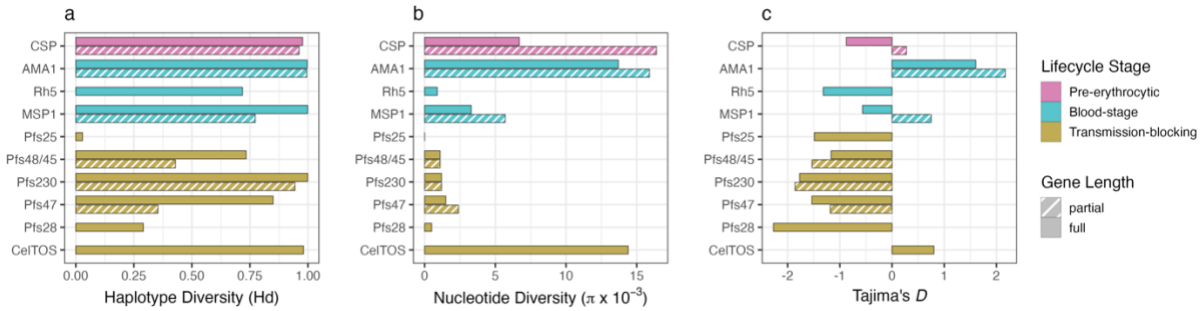


Figure 4. Results of haplotype diversity, nucleotide diversity, and Tajima's D values for each antigen across all countries included in this study. The horizontal barplot depicts the calculated values for the specified diversity parameters for each antigen across the different parasite populations. Three diversity parameters were included: haplotype diversity (panel a) and nucleotide diversity (panel b) (calculated using DNAsp), and Tajima's D statistic (panel c) (calculated using MEGA). Colors are selected to show the lifecycle stage of the parasite of the respective antigen (see legend for specific colors) and also correspond to the colors depicted in Fig. 1 of the *P. falciparum* parasite life cycle stages. The fully shaded-in bars indicate full-length genes, while the dashed bars denote partial length genes for genes with specific regions of interest for vaccine development.

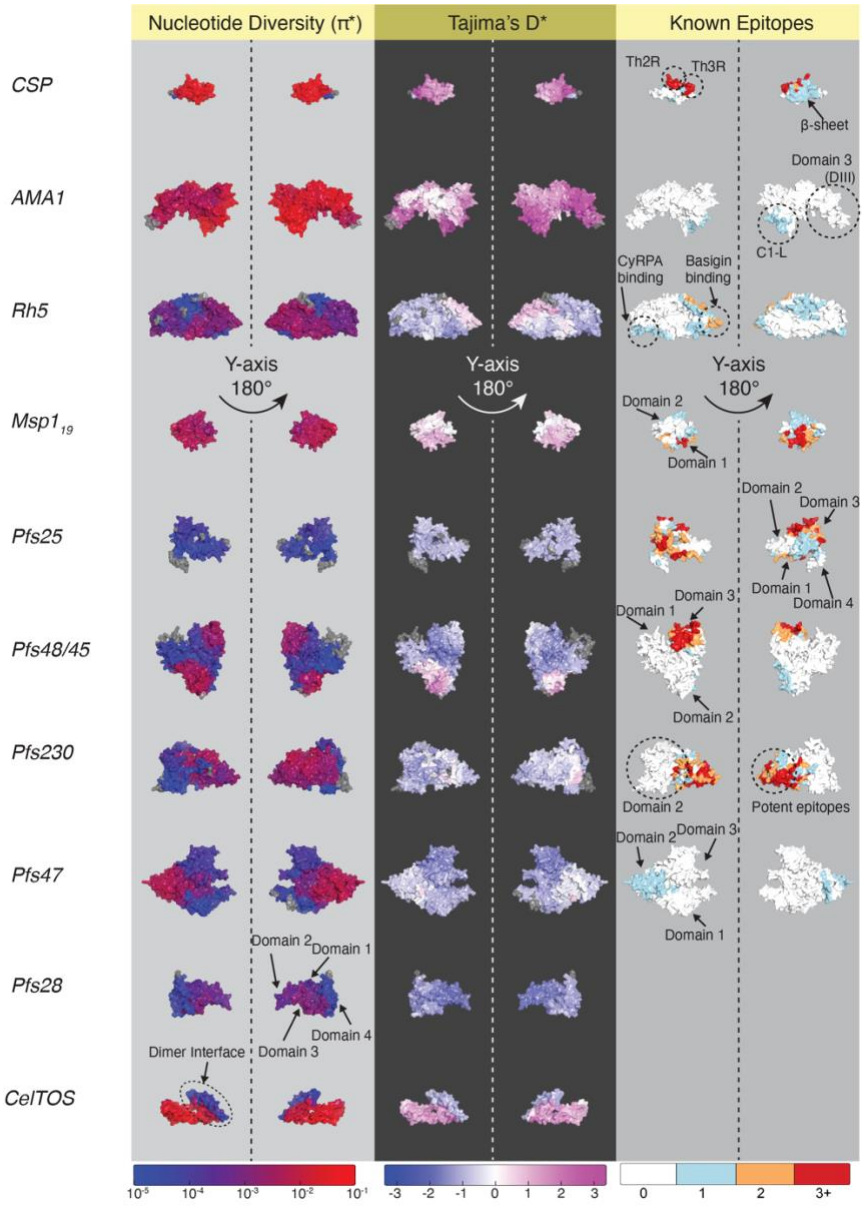


Figure 5. Vaccine candidate structural analyses of nucleotide diversity (π*), Tajima's D*, and antibody-antigen interactions (epitopes). A 3D sliding window of 15 Å was used for all calculations. Residues with no nearby SNP mapping residues are colored grey. Nucleotide diversity (π*) uses a log scale from extremely low diversity (10^{-5}) to extremely high diversity (10^{-1}). Tajima's D* varies from no balancing selection (<0) to strong balancing selection (>2) on a linear scale. In the column of known epitopes, the discrete color scale indicates the number of times a residue was part of an antibody-antigen structure deposited in RCSB. This column is additionally annotated with relevant domain, epitope, and protein-protein interaction surfaces.

Table 1. Vaccine candidate antigens.

P. falciparum antigens included in this study, with respective type of vaccine, lifecycle stage, full or partial gene used in analyses from NCBI gene database, and 3-D structure reference and antibody-antigen structures used in analyses from Research Collaboratory for Structural Bioinformatics (RCSB) database. The gene information of the proteins is provided in Table S1. All are relevant key candidates from malaria vaccine trials and are functionally important for malaria biology as they target different stages of the *Plasmodium falciparum* life cycle (See Figure 1). AF2 represents AlphaFold solved structures. "→" indicates we used AlphaFold to resolve loop regions instead of the solved structure (see Methods for details).

Type of vaccine (Parasite lifecycle stage)	Protein Name	Analysis (full/ partial gene), [nt]	Structure reference	AA start – AA end	Antibody-Antigen structures
Pre-erythrocytic (sporozoite)	*CSP*	full-length, partial (C-terminal), [829-1194]	7rxp ([74])	309 – 373	6b0s ([89]), [7rxi, 7rxj, 7rxl, 7rxp, 7s0x] ([74])
Blood-stage (merozoite)	*AMA1*	partial (Domains I-III), [298-1635]	2q8a ([75])	107 – 438	[2q8a, 2q8b] ([75])
	Rh5	full-length	4wat ([114])	146 – 513	[4u0r, 4u1g] ([80]), 5mi0 ([115]), [6rcu, 6rcv] ([81]), [7phu] ([82])
	MSP1	partial (C-terminal region: 19), [4822-5097]	8dfh --> AF2	1608 - 1699	[8dfh, 8dfi, 8dfg] ([79]), 1ob1 ([116]), 6xqw ([117])
Transmission-blocking (sexual and/or	*Pfs25*	full-length	6b0e ([107]) → AF2	23 – 193	[6azz, 6b0h, 6b0e, 6b08, 6b0a, 6b0g] ([107]), [6phb, 6phd, 6phf, 6phc] ([108]), 7txw ([118])

sporogonic development stages)	*Pfs48/45*	full-length, partial (C-terminal 6-Cys domain), [871-1284]	7zxg ([83]) → AF2	45 – 428	[7zwf, 7zwi, 7zxf, 7zwm, 7zxg] ([83]), 6h5n ([83]), 7uxl ([97]), 7unb ([119]), [6e63, 6e62] ([120])
	Pfs230	full-length, partial (N-terminal prodomain + Domains I-III), [1327-3396]	7ufw ([101]) → AF2	552 - 890	6ohg ([87]), 7jum ([99]), [7uvi, 7uvq, 7uvs, 7uvo, 7uvh] ([100]), [7u9w, 7ua2, 7ufw,7ua8, 7ubs, 7ucq, 7uc8, 7u9e, 7ui1] ([101])
	Pfs47	full-length, partial (Del2), [532-687]	None → AF2	22 – 419	Linear epitope (no structure solved) for amino acids 178-229 ([102])
	Pfs28	full-length	8e1z ([121])	24 - 178	N/A
	CelTOS	full-length	(*Pv*) 5tsz ([78]) --> AF2	35 - 164	N/A

685 **Table 2. Diversity Statistics Results.**

686 Summary of the samples from Africa included in our analyses per gene.

687 n=number of sequences, h=number of unique haplotypes, h_c=number of unique haplotypes from a

688 respective country alone, S=number of polymorphic (segregating) sites, k=average number of pairwise

689 nucleotide differences, Hd=haplotype (gene) diversity, π=nucleotide diversity, *D*=Tajima's neutrality test

690 statistic

691

Gene Name **(Gene Symbol)**	**Country**	**n**	**h**	**h_c**	**S**	**k**	**Hd**	**π**	***D***
(PF3D7_0304600) ***CSP*, full**	*All*	*243*	*100*	*70*	*63*	*7.321*	*0.977*	*0.0067*	*-0.877*
	DRC	25	17	7	33	8.140	0.923	0.0070	-0.258
	Ghana	51	34	12	30	6.921	0.976	0.0059	0.126
	Guinea	41	23	7	28	6.667	0.951	0.0058	0.064
	Malawi	32	18	4	29	7.486	0.958	0.0065	0.141
	Tanzania	20	16	9	27	7.147	0.974	0.0061	-0.237
	Zambia	74	50	31	50	8.061	0.985	0.0072	-0.703
(PF3D7_0304600) ***CSP*, partial** (C-terminal)	*All*	*623*	*107*	*68*	*38*	*5.987*	*0.962*	*0.0164*	*0.277*
	DRC	73	27	6	23	5.960	0.946	0.0163	0.797
	Ghana	138	44	18	28	5.685	0.953	0.0156	0.341
	Guinea	72	29	6	23	6.079	0.957	0.0166	0.865
	Malawi	125	40	12	26	5.757	0.964	0.0157	0.572
	Tanzania	50	23	8	24	5.783	0.951	0.0158	0.258
	Zambia	165	45	18	26	6.322	0.952	0.0173	1.080
(PF3D7_1133400) ***AMA1*, full**	*All*	*579*	*362*	*315*	*114*	*25.550*	*0.997*	*0.0137*	*1.606*
	DRC	70	60	50	82	25.598	0.993	0.0137	1.710
	Ghana	129	99	79	88	25.124	0.993	0.0134	1.762
	Guinea	71	60	45	85	24.690	0.992	0.0132	1.369
	Malawi	125	73	51	84	25.762	0.990	0.0138	2.101
	Tanzania	52	43	25	76	25.519	0.991	0.0137	1.812
	Zambia	132	89	65	87	25.635	0.987	0.0137	1.938
	All	*583*	*308*	*248*	*84*	*21.319*	*0.995*	*0.0159*	*2.171*

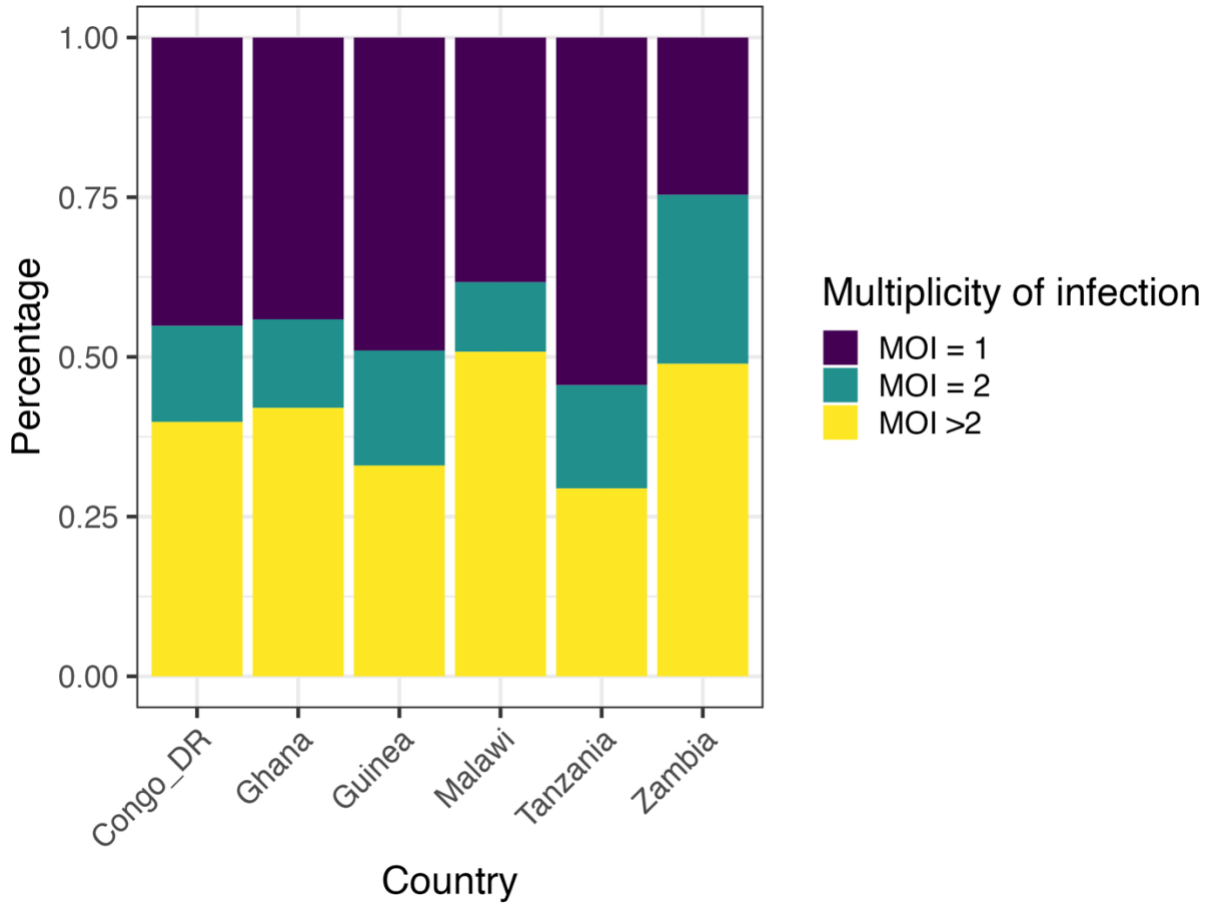
	DRC	71	54	36	64	21.384	0.985	0.0160	2.059
(PF3D7_1133400)	Ghana	131	89	63	68	20.715	0.991	0.0155	2.080
***AMA1*, partial**	Guinea	71	59	39	70	21.041	0.992	0.0157	1.522
(Domains I-III)	Malawi	125	69	39	66	21.477	0.989	0.0161	2.395
	Tanzania	52	43	21	61	21.548	0.991	0.0161	2.069
	Zambia	133	82	50	70	21.454	0.985	0.0160	2.127
(PF3D7_0424100)	*All*	*413*	*28*	*16*	*19*	*1.347*	*0.718*	*0.0009*	*-1.318*
***Rh5*, full**	DRC	28	7	2	7	1.013	0.624	0.0006	-1.312
	Ghana	112	12	4	9	0.692	0.530	0.0004	-1.456
	Guinea	66	9	0	6	0.814	0.616	0.0005	-0.854
	Malawi	116	10	1	7	1.637	0.785	0.0010	0.562
	Tanzania	41	14	6	11	1.715	0.790	0.0011	-1.006
	Zambia	50	11	3	9	1.312	0.682	0.0008	-0.972
(PF3D7_0930300)	*All*	*196*	*175*	*173*	*118*	*16.510*	*0.999*	*0.0033*	*-0.565*
***MSP1*, full**	DRC	20	18	17	61	17.100	0.989	0.0034	-0.022
	Ghana	37	35	35	81	16.480	0.997	0.0033	-0.554
	Guinea	22	22	22	67	17.403	1.000	0.0035	-0.212
	Malawi	52	40	39	67	15.778	0.989	0.0031	0.223
	Tanzania	10	10	9	49	15.467	1.000	0.0031	-0.523
	Zambia	55	52	51	71	15.727	0.998	0.0031	0.047
(PF3D7_0930300)	*All*	*612*	*17*	*4*	*8*	*1.585*	*0.773*	*0.0057*	*0.752*
***MSP1*, partial**	DRC	73	11	0	7	1.648	0.800	0.0060	0.357
(C-terminal region:	Ghana	139	9	0	6	1.571	0.729	0.0057	0.944
19)	Guinea	73	11	1	8	1.801	0.791	0.0065	0.241
	Malawi	124	12	1	7	1.613	0.782	0.0058	0.551
	Tanzania	49	9	1	7	1.480	0.751	0.0054	-0.153
	Zambia	154	11	1	7	1.436	0.764	0.0052	0.333
(PF3D7_1031000)	*All*	*552*	*5*	*3*	*4*	*0.029*	*0.029*	*0.0000*	*-1.489*
***Pfs25*, full**	DRC	67	1	0	0	0.000	0.000	0.0000	n/c
	Ghana	128	1	0	0	0.000	0.000	0.0000	n/c
	Guinea	67	1	0	0	0.000	0.000	0.0000	n/c
	Malawi	121	4	2	3	0.082	0.081	0.0001	-1.480
	Tanzania	47	2	0	1	0.043	0.043	0.0000	-1.109

	Zambia	122	3	1	2	0.033	0.033	0.0001	-1.351
(PF3D7_1346700)	*All*	*539*	*24*	*16*	*19*	*1.433*	*0.734*	*0.0011*	*-1.168*
***Pfs48/45*, full**	DRC	65	9	3	6	1.242	0.648	0.0009	-0.043
	Ghana	128	13	6	9	1.151	0.691	0.0009	-0.738
	Guinea	65	8	2	6	0.912	0.633	0.0007	-0.676
	Malawi	117	8	1	6	1.394	0.668	0.0010	0.525
	Tanzania	49	6	0	3	1.350	0.687	0.0010	2.090
	Zambia	115	12	4	8	1.205	0.671	0.0009	-0.473
(PF3D7_1346700)	*All*	*562*	*12*	*7*	*11*	*0.452*	*0.429*	*0.0011*	*-1.534*
***Pfs48/45*, partial** (C-terminal 6-Cys domain)	DRC	57	4	1	3	0.200	0.195	0.0005	-1.322
	Ghana	131	4	1	3	0.564	0.525	0.0014	0.042
	Guinea	67	5	2	4	0.594	0.544	0.0014	-0.621
	Malawi	118	4	1	3	0.299	0.292	0.0007	-0.814
	Tanzania	49	2	0	1	0.332	0.332	0.0008	0.647
	Zambia	130	7	2	6	0.225	0.215	0.0005	-1.718
(PF3D7_0209000)	*All*	*430*	*383*	*381*	*180*	*11.082*	*0.999*	*0.0012*	*-1.771*
***Pfs230*, full**	DRC	50	47	47	64	11.440	0.998	0.0012	-0.697
	Ghana	85	83	83	89	11.085	0.999	0.0012	-1.251
	Guinea	57	55	54	72	10.936	0.999	0.0012	-1.035
	Malawi	103	77	75	75	9.224	0.995	0.0010	-1.166
	Tanzania	37	37	36	46	9.216	1.000	0.0010	-0.587
	Zambia	98	87	86	68	9.389	0.995	0.0010	-0.934
(PF3D7_0209000)	*All*	*609*	*149*	*116*	*53*	*2.452*	*0.944*	*0.0012*	*-1.859*
***Pfs230*, partial** (N-terminal prodomain + Domains I-III)	DRC	68	34	15	23	2.575	0.953	0.0013	-1.438
	Ghana	134	64	39	29	2.897	0.963	0.0014	-1.330
	Guinea	74	38	21	18	2.702	0.964	0.0013	-0.796
	Malawi	129	37	11	18	2.079	0.916	0.0010	-1.029
	Tanzania	49	22	10	17	2.148	0.923	0.0010	-1.368
	Zambia	155	40	20	21	2.155	0.919	0.0011	-1.173
(PF3D7_1346800)	*All*	*522*	*64*	*38*	*33*	*2.022*	*0.850*	*0.0015*	*-1.536*
***Pfs47*, full**	DRC	60	20	6	12	2.267	0.925	0.0017	-0.341
	Ghana	124	30	13	19	1.982	0.808	0.0015	-1.224
	Guinea	67	16	3	10	1.663	0.772	0.0013	-0.561

	Malawi	118	25	8	16	1.468	0.652	0.0011	-1.396
	Tanzania	46	14	2	11	1.837	0.812	0.0014	-0.787
	Zambia	107	22	6	13	1.623	0.741	0.0012	-0.922
(PF3D7_1346800)	*All*	*555*	*8*	*4*	*7*	*0.380*	*0.354*	*0.0024*	*-1.184*
***Pfs47*, partial**	DRC	66	4	0	3	0.626	0.565	0.0040	-0.015
(Del2)	Ghana	132	5	1	4	0.461	0.419	0.0030	-0.703
	Guinea	69	5	1	4	0.394	0.370	0.0025	-1.119
	Malawi	121	5	1	4	0.160	0.142	0.0010	-1.509
	Tanzania	49	4	0	3	0.347	0.327	0.0022	-1.005
	Zambia	118	4	1	3	0.295	0.289	0.0019	-0.828
(PF3D7_1030900)	*All*	*531*	*25*	*21*	*23*	*0.313*	*0.291*	*0.0005*	*-2.272*
***Pfs28*, full**	DRC	62	3	1	2	0.236	0.232	0.0004	-0.757
	Ghana	125	7	5	6	0.340	0.326	0.0005	-1.506
	Guinea	66	6	3	5	0.314	0.298	0.0005	-1.608
	Malawi	119	10	6	9	0.280	0.263	0.0004	-2.028
	Tanzania	46	5	1	4	0.215	0.207	0.0003	-1.764
	Zambia	113	8	5	7	0.386	0.338	0.0006	-1.628
(PF3D7_1216600)	*All*	*560*	*190*	*131*	*42*	*7.891*	*0.981*	*0.0144*	*0.803*
***CelTOS*, full**	DRC	68	40	15	28	7.026	0.965	0.0128	0.639
	Ghana	131	70	38	29	8.331	0.979	0.0152	1.663
	Guinea	66	42	13	28	7.998	0.973	0.0146	1.138
	Malawi	122	55	25	30	7.848	0.974	0.0143	1.210
	Tanzania	49	33	8	30	7.571	0.969	0.0138	0.418
	Zambia	124	65	32	37	7.659	0.978	0.0140	0.353

Supplementary Material

Figure S1. Map of African endemic countries from which *P. falciparum* WGS were included in the study. Map showing the newly generated *P. falciparum* WGS from Zambia (outlined in black) and publicly available WGS data from other malaria endemic countries (MalariaGEN Pf3K) (shown in dark grey).



705

706 **Figure S2. Characterization of multiplicity of infections estimated by *Fws* among 1092 *P.***

707 ***falciparum* genomes across 6 African countries.** Colored bars represent the proportion of samples

708 with different levels of multiplicity of infection in each country based on *Fws*. MOI=1: *Fws* above 0.95;

709 MOI=2, *Fws* between 0.8 and 0.95; MOI>2, *Fws* below 0.8.

710

711

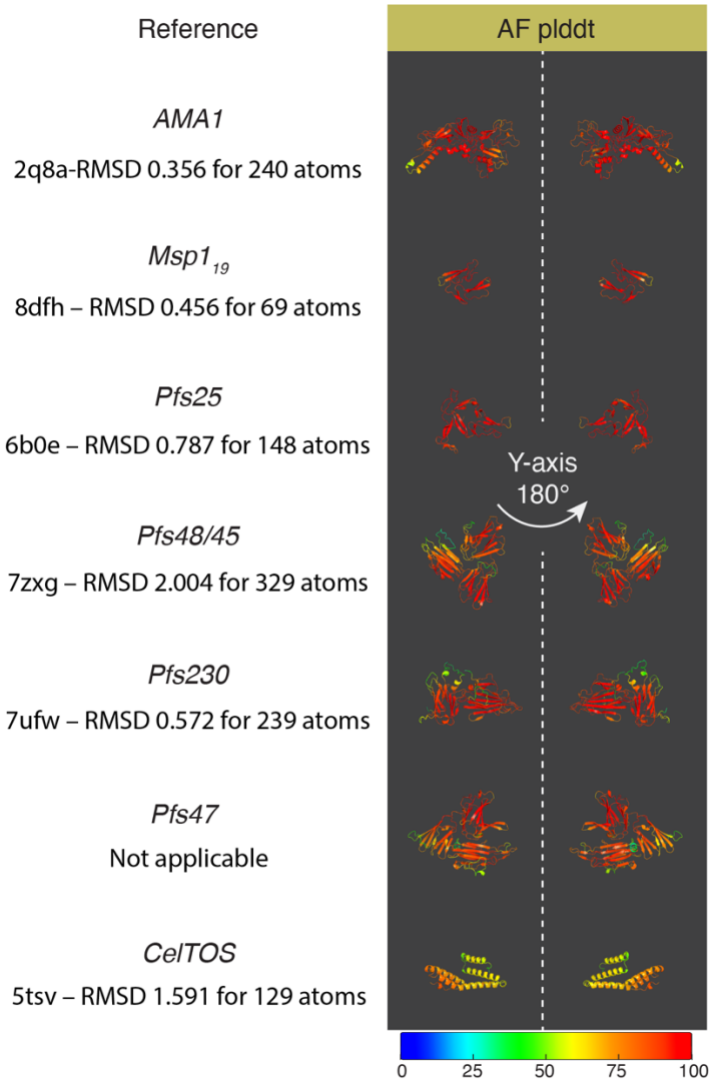


Figure S3. Vaccine candidates' AlphaFold2 model confidence. The color scale represents AlphaFold2's reported predicted local Distance Difference Test (pLDDT) for each structure model. See methods for AlphaFold2 implementation. Text under protein names indicates reference structure models and the agreement of alignments measured in Root Mean Squared Deviation (RMSD) over a certain number of atoms.

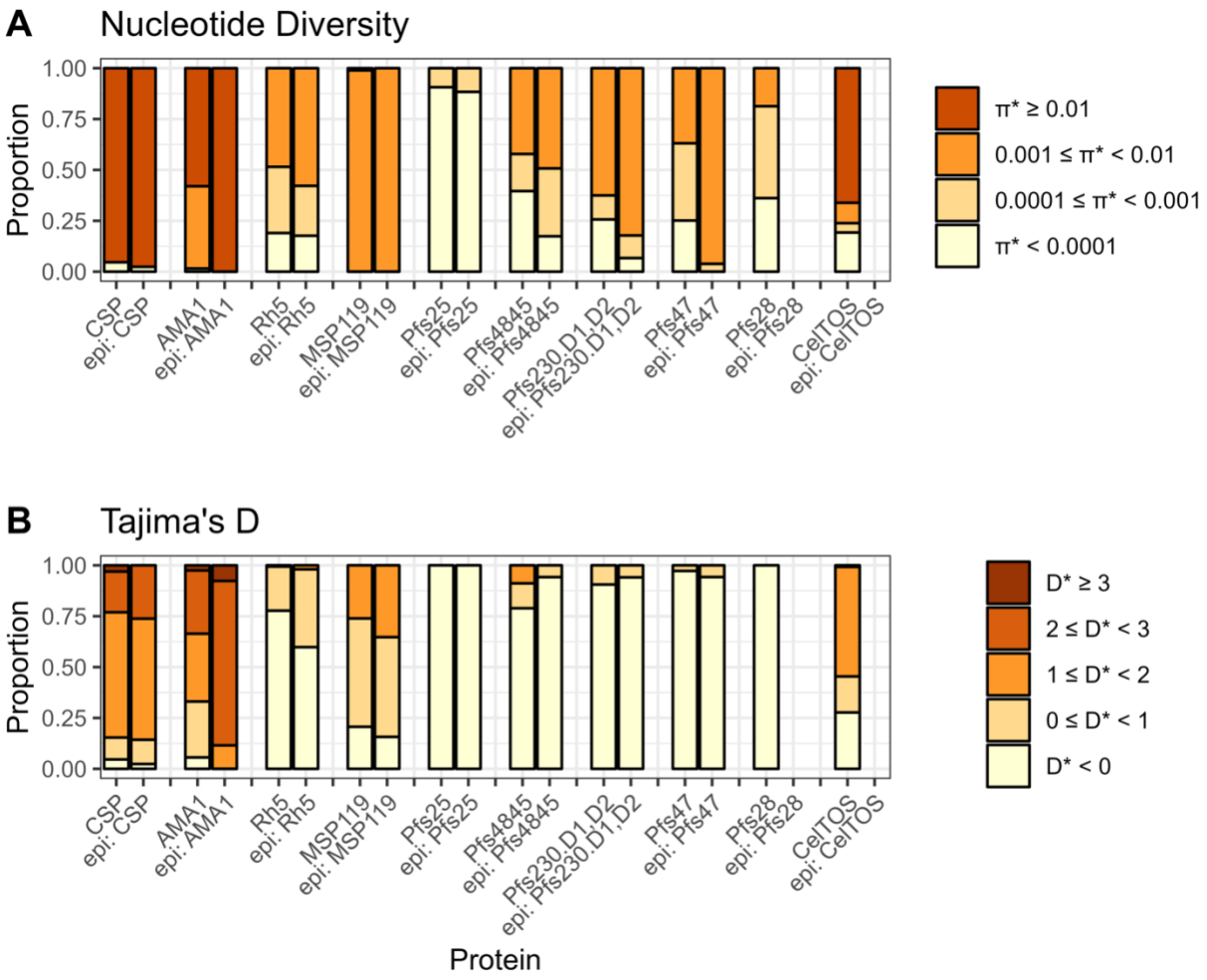


Figure S4. Spatially derived nucleotide diversity and Tajima's D summaries for whole protein structures versus known epitope residues. Known epitope residues, indicated as "epi: protein name", are calculated from solved antibody-antigen structures listed in Table 1. Color indicates the proportion of residues in each corresponding bin of nucleotide diversity. See methods for calculation of spatially derived statistics and calculation of epitope residues from solved structures. **A)** Summary of spatially derived nucleotide diversity (π*) values for whole protein structures versus known epitope residues. **B)** Summary of spatially derived Tajima's D (D*) values for whole protein structures versus known epitope residues.

733 **Table S1. Gene information of vaccine candidate antigens.**

734 Chromosome number and coordinates, relevant gene symbol and IDs of *P. falciparum* antigens from

735 NCBI gene database.

Gene Name	Chromosome Number	Chromosome Coordinates	Gene Symbol	Gene ID
CSP	3	(221323..222516, complement)	PF3D7_0304600	814364
AMA1	11	(1293856..1295724)	PF3D7_1133400	810891
RH5	4	(1082363..1084150, complement)	PF3D7_0424100	812437
MSP1	9	(1201812..1206974)	PF3D7_0930300	813575
Pfs25	10	(1253417..1254070, complement)	PF3D7_1031000	810560
Pfs45/48	13	(1876016..1877362, complement)	PF3D7_1346700	814212
Pfs230	2	(370438..379845)	PF3D7_0209000	812682
Pfs47	13	(1878875..1880194, complement)	PF3D7_1346800	814213
Pfs28	10	(1251151..1251807, complement)	PF3D7_1030900	810459
CelTOS	12	(659738..660286, complement)	PF3D7_1216600	811213

736

737

References

1. Duffy PE, Sahu T, Akue A, Milman N, Anderson C. Pre-erythrocytic malaria vaccines: identifying the targets. *Expert Review of Vaccines* 2012; **11**(10): 1261-80.
2. Marques-da-Silva C, Peissig K, Kurup SP. Pre-Erythrocytic Vaccines against Malaria. *Vaccines* 2020; **8**(3): 400.
3. Duffy PE, Patrick Gorres J. Malaria vaccines since 2000: progress, priorities, products. *npj Vaccines* 2020; **5**(1): 48.
4. Molina-Franky J, Cuy-Chaparro L, Camargo A, et al. Plasmodium falciparum pre-erythrocytic stage vaccine development. *Malaria Journal* 2020; **19**(1): 56.
5. WHO. WHO recommends groundbreaking malaria vaccine for children at risk. 2021. https://www.who.int/news/item/06-10-2021-who-recommends-groundbreaking-malaria-vaccine-for-children-at-risk (accessed February 11 2023).
6. Oxford. Oxford R21/Matrix-M™ malaria vaccine receives regulatory clearance for use in Burkina Faso. 2023. https://www.ox.ac.uk/news/2023-07-24-oxford-r21matrix-m-malaria-vaccine-receives-regulatory-clearance-use-burkina-faso (accessed September 9 2023).
7. Friedman-Klabanoff DJ, Berry AA, Travassos MA, et al. Low dose recombinant full-length circumsporozoite protein-based Plasmodium falciparum vaccine is well-tolerated and highly immunogenic in phase 1 first-in-human clinical testing. *Vaccine* 2021; **39**(8): 1195-200.
8. Miura K. Progress and prospects for blood-stage malaria vaccines. *Expert Review of Vaccines* 2016; **15**(6): 765-81.
9. Sauerwein RW, Bousema T. Transmission blocking malaria vaccines: Assays and candidates in clinical development. *Vaccine* 2015; **33**(52): 7476-82.
10. Coelho CH, Rappuoli R, Hotez PJ, Duffy PE. Transmission-Blocking Vaccines for Malaria: Time to Talk about Vaccine Introduction. *Trends in Parasitology* 2019; **35**(7): 483-6.
11. WHO. World malaria report 2022. 2022. https://www.who.int/publications/i/item/9789240064898 (accessed February 9 2023).
12. Laurens MB. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Hum Vaccin Immunother* 2020; **16**(3): 480-9.
13. Keating C. The history of the RTS,S/AS01 malaria vaccine trial. *The Lancet* 2020; **395**(10233): 1336-7.
14. WHO. WHO review of malaria vaccine clinical development. 2024. https://www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/who-review-of-malaria-vaccine-clinical-development (accessed February 15 2024).
15. Oxford. Oxford R21/Matrix-M™ malaria vaccine receives WHO recommendation for use paving the way for global roll-out. 2023. https://www.ox.ac.uk/news/2023-10-02-oxford-r21matrix-m-malaria-vaccine-receives-who-recommendation-use-paving-way-global (accessed February 9 2024).
16. Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 2002; **419**(6906): 498-511.
17. Cockburn IA, Seder RA. Malaria prevention: from immunological concepts to effective vaccines and protective antibodies. *Nature Immunology* 2018; **19**(11): 1199-211.
18. Kirchner S, Power BJ, Waters AP. Recent advances in malaria genomics and epigenomics. *Genome Medicine* 2016; **8**(1): 92.

19. Bonam SR, Rénia L, Tadepalli G, Bayry J, Kumar HMS. Plasmodium falciparum Malaria Vaccines and Vaccine Adjuvants. *Vaccines* 2021; **9**(10): 1072.
20. Coppel RL. Vaccinating with the genome: a Sisyphean task? *Trends in Parasitology* 2009; **25**(5): 205-12.
21. Pandey I, Quadiri A, Wadi I, Pillai CR, Singh AP, Das A. Conserved Plasmodium Protein (PF3D7_0406000) of Unknown Function: In-silico Analysis and Cellular Localization. *Infection, Genetics and Evolution* 2021; **92**: 104848.
22. Mu J, Awadalla P, Duan J, et al. Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. *Nature Genetics* 2007; **39**(1): 126-30.
23. Goodswen SJ, Kennedy PJ, Ellis JT. A guide to current methodology and usage of reverse vaccinology towards in silico vaccine discovery. *FEMS Microbiology Reviews* 2023; **47**(2).
24. Beeson JG, Kurtovic L, Dobaño C, et al. Challenges and strategies for developing efficacious and long-lasting malaria vaccines. *Science Translational Medicine* 2019; **11**(474): eaau1458.
25. Naung MT, Martin E, Munro J, et al. Global diversity and balancing selection of 23 leading Plasmodium falciparum candidate vaccine antigens. *PLOS Computational Biology* 2022; **18**(2): e1009801.
26. Olotu A, Fegan G, Wambua J, et al. Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among Young African Children. *New England Journal of Medicine* 2016; **374**(26): 2519-29.
27. CTP. 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* 2015; **386**(9988): 31-45.
28. de Vrieze J. First malaria vaccine rolled out in Africa—despite limited efficacy and nagging safety concerns. 2019. https://www.science.org/content/article/first-malaria-vaccine-rolled-out-africa-despite-limited-efficacy-and-nagging-safety.
29. Pringle JC, Carpi G, Almagro-Garcia J, et al. RTS,S/AS01 malaria vaccine mismatch observed among Plasmodium falciparum isolates from southern and central Africa and globally. *Scientific Reports* 2018; **8**(1): 6622.
30. Bell GJ, Agnandji ST, Asante KP, et al. Impacts of Ecology, Parasite Antigenic Variation, and Human Genetics on RTS,S/AS01e Malaria Vaccine Efficacy. *Current Epidemiology Reports* 2021; **8**(3): 79-88.
31. Preston MD, Campino S, Assefa SA, et al. A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. *Nature Communications* 2014; **5**(1): 4052.
32. Bailey JA, Mvalo T, Aragam N, et al. Use of Massively Parallel Pyrosequencing to Evaluate the Diversity of and Selection on Plasmodium falciparum csp T-Cell Epitopes in Lilongwe, Malawi. *The Journal of Infectious Diseases* 2012; **206**(4): 580-7.
33. Neafsey DE, Juraska M, Bedford T, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *New England Journal of Medicine* 2015; **373**(21): 2025-37.
34. Mangou K, Moore AJ, Thiam LG, et al. Structure-guided insights into potential function of novel genetic variants in the malaria vaccine candidate PfRh5. *Scientific Reports* 2022; **12**(1): 19403.
35. Zhang H, Wang L, Compans RW, Wang B-Z. Universal Influenza Vaccines, a Dream to Be Realized Soon. *Viruses* 2014; **6**(5): 1974-91.

36. Bullard BL, Weaver EA. Strategies Targeting Hemagglutinin as a Universal Influenza Vaccine. *Vaccines* 2021; **9**(3): 257.
37. Giuliani MM, Adu-Bobie J, Comanducci M, et al. A universal vaccine for serogroup B meningococcus. *Proceedings of the National Academy of Sciences* 2006; **103**(29): 10834-9.
38. 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. *Parasites & Vectors* 2018; **11**(1): 323.
39. Nirmolia T, Ahmed MA, Sathishkumar V, 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. *Malaria Journal* 2022; **21**(1): 62.
40. Sookpongthai P, Utayopas K, Sitthiyotha T, et al. Global diversity of the gene encoding the Pfs25 protein—a Plasmodium falciparum transmission-blocking vaccine candidate. *Parasites & Vectors* 2021; **14**(1): 571.
41. Ouattara A, Dwivedi A, Adams M, 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): 1090.
42. WHO. World malaria report 2023. 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023 (accessed February 3 2024).
43. Fola AA, He Q, Xie S, et al. Genomics reveals heterogeneous Plasmodium falciparum transmission and selection signals in Zambia. *Communications Medicine* 2024; **4**(1): 67.
44. Ministry of Health Zambia M. Zambia National Malaria Indicator Survey Report. 2015 2019. https://www.path.org/resources/zambia-natl-malaria-indicator-survey-mis-2018/ (accessed November 25 2020).
45. Ippolito MM, Gebhardt ME, Ferriss E, et al. Scientific Findings of the Southern and Central Africa International Center of Excellence for Malaria Research: Ten Years of Malaria Control Impact Assessments in Hypo-, Meso-, and Holoendemic Transmission Zones in Zambia and Zimbabwe. *The American Journal of Tropical Medicine and Hygiene* 2022; **107**(4_Suppl): 55-67.
46. Fola AA, Dorman J, Levy M, Ciubotariu I, Carpi G. Optimized HT gDNA extraction from Dried Blood Spot using QIAcube HT for Malaria Genomic Epidemiology Studies. 2020. https://www.protocols.io/view/optimized-ht-gdna-extraction-from-dried-blood-spot-j8nlkej31l5r/v1.
47. Choudhary S. pysradb: A Python package to query next-generation sequencing metadata and data from NCBI Sequence Read Archive [version 1; peer review: 2 approved]. *F1000Research* 2019; **8**(532).
48. Carpi G, Gorenstein L, Harkins TT, Samadi M, Vats P. A GPU-accelerated compute framework for pathogen genomic variant identification to aid genomic epidemiology of infectious disease: a malaria case study. *Briefings in Bioinformatics* 2022; **23**(5).
49. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics* 2013.
50. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**(9): 1297-303.
51. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**(5): 491-8.

874 52. Miles A, Iqbal Z, Vauterin P, et al. Indels, structural variation, and recombination drive
875 genomic diversity in Plasmodium falciparum. *Genome Res* 2016; **26**(9): 1288-99.
876 53. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the
877 effects of single nucleotide polymorphisms, SnpEff. *Fly* 2012; **6**(2): 80-92.
878 54. Manske M, Miotto O, Campino S, et al. Analysis of Plasmodium falciparum diversity in
879 natural infections by deep sequencing. *Nature* 2012; **487**(7407): 375-9.
880 55. Rausch KM, Barnafo EK, Lambert LE, et al. Preclinical evaluations of Pfs25-EPA and
881 Pfs230D1-EPA in AS01 for a vaccine to reduce malaria transmission. *iScience* 2023; **26**(7).
882 56. Duffy PE. The Virtues and Vices of Pfs230: From Vaccine Concept to Vaccine
883 Candidate. *The American Journal of Tropical Medicine and Hygiene* 2022; **107**(3_Suppl): 17-21.
884 57. Silk SE, Kalinga WF, Mtaka IM, et al. Superior antibody immunogenicity of a viral-
885 vectored RH5 blood-stage malaria vaccine in Tanzanian infants as compared to adults. *Med*
886 2023.
887 58. Schwartz L, Brown GV, Genton B, Moorthy VS. A review of malaria vaccine clinical
888 projects based on the WHO rainbow table. *Malaria Journal* 2012; **11**(1): 11.
889 59. Molina-Cruz A, Barillas-Mury C. Pfs47 as a Malaria Transmission-Blocking Vaccine
890 Target. *The American Journal of Tropical Medicine and Hygiene* 2022; **107**(3_Suppl): 27-31.
891 60. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. DnaSP 6: DNA Sequence
892 Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution* 2017; **34**(12):
893 3299-302.
894 61. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction
895 endonucleases. *Proceedings of the National Academy of Sciences* 1979; **76**(10): 5269-73.
896 62. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA
897 polymorphism. *Genetics* 1989; **123**(3): 585-95.
898 63. Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. *Genetics*
899 1981; **97**(1): 145-63.
900 64. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction.
901 *Methods in Ecology and Evolution* 2015; **6**(9): 1110-6.
902 65. Clement M, Snell Q, Walke P, Posada D, Crandall K. TCS: estimating gene genealogies.
903 Proceedings 16th International Parallel and Distributed Processing Symposium; 2002 15-19
904 April 2002; 2002. p. 7 pp.
905 66. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with
906 AlphaFold. *Nature* 2021; **596**(7873): 583-9.
907 67. Reva BA, Finkelstein AV, Skolnick J. What is the probability of a chance prediction of a
908 protein structure with an rmsd of 6 å? *Folding and Design* 1998; **3**(2): 141-7.
909 68. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in
910 proteins. *The EMBO Journal* 1986; **5**(4): 823-6.
911 69. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold:
912 making protein folding accessible to all. *Nature Methods* 2022; **19**(6): 679-82.
913 70. Guy AJ, Irani V, Richards JS, Ramsland PA. BioStructMap: a Python tool for integration
914 of protein structure and sequence-based features. *Bioinformatics* 2018; **34**(22): 3942-4.
915 71. Guy AJ, Irani V, Beeson JG, et al. Proteome-wide mapping of immune features onto
916 Plasmodium protein three-dimensional structures. *Scientific Reports* 2018; **8**(1): 4355.
917 72. Ramaraj T, Angel T, Dratz EA, Jesaitis AJ, Mumey B. Antigen–antibody interface
918 properties: Composition, residue interactions, and features of 53 non-redundant structures.
919 *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2012; **1824**(3): 520-32.

73. MalariaGEN. Genomic epidemiology of artemisinin resistant malaria. *eLife* 2016; **5**: e08714.
74. Beutler N, Pholcharee T, Oyen D, et al. A novel CSP C-terminal epitope targeted by an antibody with protective activity against Plasmodium falciparum. *PLOS Pathogens* 2022; **18**(3): e1010409.
75. Coley AM, Gupta A, Murphy VJ, et al. Structure of the Malaria Antigen AMA1 in Complex with a Growth-Inhibitory Antibody. *PLOS Pathogens* 2007; **3**(9): e138.
76. Bai T, Becker M, Gupta A, et al. Structure of AMA1 from Plasmodium falciparum reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proc Natl Acad Sci U S A* 2005; **102**(36): 12736-41.
77. Dutta S, Lee SY, Batchelor AH, Lanar DE. Structural basis of antigenic escape of a malaria vaccine candidate. *Proceedings of the National Academy of Sciences of the United States of America* 2007; **104**(30): 12488-93.
78. Jimah JR, Salinas ND, Sala-Rabanal M, et al. Malaria parasite CelTOS targets the inner leaflet of cell membranes for pore-dependent disruption. *Elife* 2016; **5**.
79. Patel PN, Dickey TH, Hopp CS, et al. Neutralizing and interfering human antibodies define the structural and mechanistic basis for antigenic diversion. *Nature Communications* 2022; **13**(1): 5888.
80. Wright KE, Hjerrild KA, Bartlett J, et al. Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature* 2014; **515**(7527): 427-30.
81. Alanine DGW, Quinkert D, Kumarasingha R, et al. Human Antibodies that Slow Erythrocyte Invasion Potentiate Malaria-Neutralizing Antibodies. *Cell* 2019; **178**(1): 216-28.e21.
82. Ragotte RJ, Pulido D, Lias AM, et al. Heterotypic interactions drive antibody synergy against a malaria vaccine candidate. *Nature Communications* 2022; **13**(1): 933.
83. Lennartz F, Brod F, Dabbs R, et al. Structural basis for recognition of the malaria vaccine candidate Pfs48/45 by a transmission blocking antibody. *Nature Communications* 2018; **9**(1): 3822.
84. Ochwedo KO, Onyango SA, Omondi CJ, et al. Signatures of selection and drivers for novel mutation on transmission-blocking vaccine candidate Pfs25 gene in western Kenya. *PLOS ONE* 2022; **17**(4): e0266394.
85. Kaslow DC, Quakyi IA, Keister DB. Minimal variation in a vaccine candidate from the sexual stage of Plasmodium falciparum. *Molecular and Biochemical Parasitology* 1989; **32**(1): 101-3.
86. Duffy PE, Kaslow DC. A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as falciparum malaria transmission-blocking vaccines. *Infection and Immunity* 1997; **65**(3): 1109-13.
87. Singh K, Burkhardt M, Nakuchima S, et al. Structure and function of a malaria transmission blocking vaccine targeting Pfs230 and Pfs230-Pfs48/45 proteins. *Communications Biology* 2020; **3**(1): 395.
88. Weiss GE, Crabb BS, Gilson PR. Overlaying Molecular and Temporal Aspects of Malaria Parasite Invasion. *Trends in Parasitology* 2016; **32**(4): 284-95.
89. Scally SW, Murugan R, Bosch A, et al. Rare PfCSP C-terminal antibodies induced by live sporozoite vaccination are ineffective against malaria infection. *Journal of Experimental Medicine* 2017; **215**(1): 63-75.

964 90. Pholcharee T, Oyen D, Flores-Garcia Y, et al. Structural and biophysical correlation of
965 anti-NANP antibodies with in vivo protection against P. falciparum. *Nature Communications*
966 2021; **12**(1): 1063.
967 91. Basu M, Maji AK, Mitra M, Sengupta S. Natural selection and population genetic
968 structure of domain-I of Plasmodium falciparum apical membrane antigen-1 in India. *Infection,*
969 *Genetics and Evolution* 2013; **18**: 247-56.
970 92. Collins CR, Withers-Martinez C, Hackett F, Blackman MJ. An Inhibitory Antibody
971 Blocks Interactions between Components of the Malarial Invasion Machinery. *PLOS Pathogens*
972 2009; **5**(1): e1000273.
973 93. Ghoshal S, Chowdhury P, Ray S, et al. Population genetic and biophysical evidences
974 reveal that purifying selection shapes the genetic landscape of Plasmodium falciparum RH
975 ligands in Chhattisgarh and West Bengal, India. *Malaria Journal* 2020; **19**(1): 367.
976 94. van Schaijk BCL, van Dijk MR, van de Vegte-Bolmer M, et al. Pfs47, paralog of the
977 male fertility factor Pfs48/45, is a female specific surface protein in Plasmodium falciparum.
978 *Molecular and Biochemical Parasitology* 2006; **149**(2): 216-22.
979 95. van Dijk MR, Janse CJ, Thompson J, et al. A Central Role for P48/45 in Malaria Parasite
980 Male Gamete Fertility. *Cell* 2001; **104**(1): 153-64.
981 96. Eksi S, Czesny B, Van Gemert G-J, Sauerwein RW, Eling W, Williamson KC. Malaria
982 transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating
983 male parasites and oocyst production. *Molecular Microbiology* 2006; **61**(4): 991-8.
984 97. Fabra-García A, Hailemariam S, de Jong RM, et al. Highly potent, naturally acquired
985 human monoclonal antibodies against Pfs48/45 block Plasmodium falciparum transmission to
986 mosquitoes. *Immunity* 2023; **56**(2): 406-19.e7.
987 98. Ko K-T, Lennartz F, Mekhaiel D, et al. Structure of the malaria vaccine candidate
988 Pfs48/45 and its recognition by transmission blocking antibodies. *Nature Communications* 2022;
989 **13**(1): 5603.
990 99. Coelho CH, Tang WK, Burkhardt M, et al. A human monoclonal antibody blocks malaria
991 transmission and defines a highly conserved neutralizing epitope on gametes. *Nature*
992 *Communications* 2021; **12**(1): 1750.
993 100. Ivanochko D, Fabra-García A, Teelen K, et al. Potent transmission-blocking monoclonal
994 antibodies from naturally exposed individuals target a conserved epitope on Plasmodium
995 falciparum Pfs230. *Immunity* 2023; **56**(2): 420-32.e7.
996 101. Tang WK, Coelho CH, Miura K, et al. A human antibody epitope map of Pfs230D1
997 derived from analysis of individuals vaccinated with a malaria transmission-blocking vaccine.
998 *Immunity* 2023; **56**(2): 433-43.e5.
999 102. Canepa GE, Molina-Cruz A, Yenkoidiok-Douti L, et al. Antibody targeting of a specific
1000 region of Pfs47 blocks Plasmodium falciparum malaria transmission. *npj Vaccines* 2018; **3**(1):
1001 26.
1002 103. Stowers AW, Keister DB, Muratova O, Kaslow DC. A Region of Plasmodium
1003 falciparumAntigen Pfs25 That Is the Target of Highly Potent Transmission-Blocking
1004 Antibodies. *Infection and Immunity* 2000; **68**(10): 5530-8.
1005 104. Kaslow DC. Immunogenicity of Plasmodium falciparum sexual stage antigens:
1006 Implications for the design of a transmission blocking vaccine. *Immunology Letters* 1990; **25**(1):
1007 83-6.

105. Barr PJ, Green KM, Gibson HL, Bathurst IC, Quakyi IA, Kaslow DC. Recombinant Pfs25 protein of Plasmodium falciparum elicits malaria transmission-blocking immunity in experimental animals. *Journal of Experimental Medicine* 1991; **174**(5): 1203-8.
106. Miura K, Keister DB, Muratova OV, Sattabongkot J, Long CA, Saul A. Transmission-blocking activity induced by malaria vaccine candidates Pfs25/Pvs25 is a direct and predictable function of antibody titer. *Malaria Journal* 2007; **6**(1): 107.
107. Scally SW, McLeod B, Bosch A, et al. Molecular definition of multiple sites of antibody inhibition of malaria transmission-blocking vaccine antigen Pfs25. *Nature Communications* 2017; **8**(1): 1568.
108. McLeod B, Miura K, Scally SW, et al. Potent antibody lineage against malaria transmission elicited by human vaccination with Pfs25. *Nature Communications* 2019; **10**(1): 4328.
109. Gozar MMG, Price VL, Kaslow DC. Saccharomyces cerevisiae-Secreted Fusion Proteins Pfs25 and Pfs28 Elicit Potent Plasmodium falciparum Transmission-Blocking Antibodies in Mice. *Infection and Immunity* 1998; **66**(1): 59-64.
110. Qian F, Aebig JA, Reiter K, et al. Enhanced antibody responses to Plasmodium falciparum Pfs28 induced in mice by conjugation to ExoProtein A of Pseudomonas aeruginosa with an improved procedure. *Microbes and Infection* 2009; **11**(3): 408-12.
111. Menon V, Kapulu MC, Taylor I, et al. Assessment of Antibodies Induced by Multivalent Transmission-Blocking Malaria Vaccines. *Frontiers in Immunology* 2018; **8**.
112. Gregory JA, Li F, Tomosada LM, et al. Algae-Produced Pfs25 Elicits Antibodies That Inhibit Malaria Transmission. *PLOS ONE* 2012; **7**(5): e37179.
113. Pirahmadi S, Zakeri S, A. Mehrizi A, et al. Cell-traversal protein for ookinetes and sporozoites (CelTOS) formulated with potent TLR adjuvants induces high-affinity antibodies that inhibit Plasmodium falciparum infection in Anopheles stephensi. *Malaria Journal* 2019; **18**(1): 146.
114. Chen L, Xu Y, Healer J, et al. Crystal structure of PfRh5, an essential P. falciparum ligand for invasion of human erythrocytes. *eLife* 2014; **3**: e04187.
115. Campeotto I, Goldenzweig A, Davey J, et al. One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proceedings of the National Academy of Sciences* 2017; **114**(5): 998-1002.
116. Pizarro JC, Chitarra V, Verger D, et al. Crystal Structure of a Fab Complex Formed with PfMSP1-19, the C-terminal Fragment of Merozoite Surface Protein 1 from Plasmodium falciparum: A Malaria Vaccine Candidate. *Journal of Molecular Biology* 2003; **328**(5): 1091-103.
117. Thouvenel CD, Fontana MF, Netland J, et al. Multimeric antibodies from antigen-specific human IgM+ memory B cells restrict Plasmodium parasites. *Journal of Experimental Medicine* 2021; **218**(4).
118. MacDonald NJ, Singh K, Reiter K, et al. Structural and immunological differences in Plasmodium falciparum sexual stage transmission-blocking vaccines comprised of Pfs25-EPA nanoparticles. *npj Vaccines* 2023; **8**(1): 56.
119. McLeod B, Mabrouk MT, Miura K, et al. Vaccination with a structure-based stabilized version of malarial antigen Pfs48/45 elicits ultra-potent transmission-blocking antibody responses. *Immunity* 2022; **55**(9): 1680-92.e8.
120. Kundu P, Semesi A, Jore MM, et al. Structural delineation of potent transmission-blocking epitope I on malaria antigen Pfs48/45. *Nature Communications* 2018; **9**(1): 4458.

121. Shukla N, Tang WK, Tolia NH. Structural analysis of Plasmodium falciparum ookinete surface antigen Pfs28 relevant for malaria vaccine design. *Scientific Reports* 2022; **12**(1): 19556.