Genetic polymorphism and evidence of signatures of selection in the Plasmodium falciparum circumsporozoite protein gene in Tanzanian regions with different malaria endemicity Beatus M. Lyimo^{1,2*}, Catherine Bakari¹, Zachary R. Popkin-Hall³, David J. Giesbrecht⁴, Misago D. Seth¹, Dativa Pereus^{1,5}, Ramadhan Moshi¹, Ruth Boniface^{1,5}, Celine I. Mandara¹, Rashid Madebe¹, Jonathan J. Juliano³, Jeffrey A. Bailey⁴ and Deus S. Ishengoma^{1,6,7} ¹National Institute for Medical Research, Dar es Salaam, Tanzania. ²Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania ³University of North Carolina, Chapel Hill, NC, USA ⁴Brown University, Providence, RI, USA ⁵Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania. ⁶Harvard T.H Chan School of Public Health, Boston, MA, USA ⁷Faculty of Pharmaceutical Sciences, Monash University, Melbourne, Australia *Corresponding author: beatus.lyimo@nm-aist.ac.tz

32 Abstract

33 Background: In 2021 and 2023, the World Health Organization approved RTS, S/AS01 and R21/Matrix M malaria vaccines, respectively, for routine immunization of children in 34 African countries with moderate to high transmission. These vaccines are made of 35 Plasmodium falciparum circumsporozoite protein (Pfcsp) but polymorphisms in this 36 gene raises concerns regarding strain-specific responses and the long-term efficacy of 37 these vaccines. This study assessed the *Pfcsp* genetic diversity, population structure 38 and signatures of selection among parasites from areas of different malaria 39 40 transmission in mainland Tanzania, to generate baseline data before the introduction of 41 the malaria vaccines in the country.

Methods: The analysis involved 589 whole genome sequences generated by and as 42 43 part of the MalariaGEN Community Project. The samples were collected between 2013 and January 2015 from five regions of mainland Tanzania: Morogoro and Tanga 44 (Muheza) (moderate transmission areas), and Kagera (Muleba), Lindi (Nachingwea), 45 46 and Kigoma (Ujiji) (high transmission areas). Wright's inbreeding coefficient (F_{ws}), Wright's fixation index (F_{ST}), principal component analysis, nucleotide diversity, and 47 48 Tajima's D were used to assess within-host parasite diversity, population structure and 49 natural selection.

Results: Based on F_{ws} (< 0.95), there was high polyclonality (ranged from 69.23% in Nachingwea to 56.9% in Muheza). No population structure was detected in the *Pfcsp* gene in the five regions (mean F_{ST} = 0.0068). The average nucleotide diversity (π), nucleotide differentiation (K) and haplotype diversity (Hd) in the five regions were 4.19, 0.973 and 0.0035, respectively. The C-terminal region of *Pfcsp* showed high nucleotide

diversity at Th2R and Th3R regions. Positive values for the Tajima's D were observed in
the Th2R and Th3R regions consistent with balancing selection. The *Pfcsp* C-terminal
sequences had 50 different haplotypes (H_1 to H_50) and only 2% of sequences
matched the 3D7 strain haplotype (H_50).

59 **Conclusions:** The findings demonstrate high diversity of the *Pfcsp* gene with limited population differentiation. The Pfcsp gene showed positive Tajima's D values for 60 parasite populations, consistent with balancing selection for variants within Th2R and 61 62 Th3R regions. This data is consistent with other studies conducted across Africa and worldwide, which demonstrate low 3D7 haplotypes and little population structure. 63 64 Therefore, additional research is warranted, incorporating other regions and more recent data to comprehensively assess trends in genetic diversity within this important 65 66 gene. Such insights will inform the choice of alleles to be included in the future vaccines Keywords: Plasmodium falciparum, Circumsporozoite protein, malaria vaccine, genetic 67 68 diversity, signature of selection, Tanzania

- 69
- 70
- 71
- 72
- 73
- 74
- 75
- 76
- 77

78 Background

79 Malaria remains a major public health problem worldwide, with nearly half of the world's population at risk and lives in areas where malaria transmission occurs, but with the 80 81 majority of cases and deaths from Sub-Saharan (SSA) (1). In 2022, over 95% (233 82 million) of malaria cases and 96% of all deaths were reported from 29 countries in the World Health Organization (WHO) region of Africa (WHO – AFRO). The majority of the 83 deaths occurred in Nigeria (31%), the Democratic Republic of the Congo (12%), Niger 84 (6%) and the United Republic of Tanzania (4%) (2). In Tanzania, malaria is still a 85 86 leading cause of morbidity and mortality mainly among children and pregnant women, and about 93% of the population lives in malaria-transmission areas (3). However, in 87 recent years, the pattern of malaria in Tanzania has become heterogeneous, with a 88 89 consistently higher burden in north-western, western and southern regions where high prevalence in school children (reaching 82%) has been reported, while other regions 90 91 have consistently low to very low prevalence over the past 20 years (3).

92

The WHO Global Technical Strategy for Malaria 2016-2030 aims to reduce 90% of both 93 malaria incidences and mortality rates by 90% of the 2015 levels by 2030 and proceed 94 to the elimination targets (4). Malaria eradication which is projected by 2050 (5), will 95 greatly depend on the use of new and innovative tools as well as enhanced use of 96 97 current interventions including surveillance, monitoring and evaluation (SME), vector control, and case management [through parasitological testing with rapid diagnostic 98 tests (RDTs) and treatment with effective antimalarials (artemisinin combination 99 100 therapy- ACT)] (4). However, malaria control strategies face many challenges including

emerging drug and insecticide resistance and histidine-rich protein 2/3 (*Pfhrp2/3*) gene deletions, limited diagnostic capacity, socio-cultural hindrance and a lack of effective vaccines (4,6,7). Urgent measures are required to get malaria elimination efforts on track and progressing to the GTS targets by 2030.

105

106 Malaria parasites have a complex life cycle in both human and mosquito hosts, 107 including an asymptomatic stage (pre-erythrocytic), followed by a symptomatic blood 108 stage (erythrocytic) in humans and a sexual stage in the mosquito host. The parasite 109 also exhibits stages characterized by extensive genetic and antigenic diversity and 110 these factors which are associated with the biology of parasites and vectors may 111 present obstacles to anti-malarial control measures (8,9). As with other pathogens, 112 vaccines are among the most successful and cost-effective interventions in the history 113 of public health (10). For effective malaria control and elimination strategies, it is 114 important to develop a malaria vaccine that provides durable protection against clinical 115 malaria and prevents infection. Efforts to develop a malaria vaccine have been ongoing 116 since the 1980s and focused on vaccines targeting specific stages of the parasite life cycle such as pre-erythrocytic, erythrocytic and sexual stage vaccine candidates (11-117 118 13). Antigens expressed in the pre-erythrocytic (sporozoites and liver) stages represent 119 the ideal vaccine candidate to block the progression to clinical malaria through vaccine-120 based interventions (14,15).

121

122 The RTS,S/A01 is a pre-erythrocytic vaccine based on circumsporozoite protein (*Pfcsp* 123 of the 3D7 laboratory strain of *P. falciparum*) (9). It is the first malaria vaccine to be

124 endorsed by the WHO for routine immunization of children in moderate to high 125 transmission areas (16). The vaccine is made of a fragment of CSP antigen fused with 126 Hepatitis B surface antigen and AS01 adjuvant (17). It targets specific immunogenic 127 epitopes expressed on the surface of the *Pfcsp* gene and offers better protection against malaria infections (17,18). Clinical trials of the RTS,S vaccine which were 128 129 conducted in African countries including phase III trials involving children aged 5-17 130 months and 6-12 weeks from Mozambique, Gabon, Gambia, Ghana, Kenya and 131 Tanzania showed significant protection against natural falciparum infections (19–23). The overall efficacy of the vaccine administered at 0-, 1-, 2-month followed by a fourth 132 133 dose administered at 18 months was 36.3% (95% CI: 31.8-40.5) after an average of 48 134 months of follow-up (19-23).

135

136 The R21/Matrix-M vaccine is a next-generation pre-erythrocytic vaccine based on CSP, similar to the RTS,S/AS01 vaccine (24). The vaccine was created by combining the C-137 138 terminal segment of CSP from the *P. falciparum* strain NF54 (24,25). The utilized CSP 139 portion encompasses 19 NANP repeats sourced from the central repeat region, 140 recognized as a protective B cell epitope, along with the C-terminal region housing 141 essential T cell epitopes. It is the second malaria vaccine endorsed by the WHO, 142 succeeding the RTS,S/AS01 vaccine, and is recommended for use in children residing 143 in regions with low to moderate malaria prevalence (2). This vaccine has successfully 144 completed phase III clinical trials, involving children aged 5-36 months, at five sites in four countries across West Africa (Burkina Faso and Mali) and East Africa (Kenya and 145 146 Tanzania), each with varying malaria transmission patterns. Results from this study have demonstrated a high efficacy of more than 75% against clinical malaria in African children, addressing both seasonal and perennial transmission settings (26,27). However, a study from elsewhere has indicated that the CSP-based vaccine demonstrates improved efficacy against clinical malaria when infections match the 3D7 vaccine construct at epitope haplotypes and amino acid positions. This implies that the vaccine's efficacy may partially depend on the prevalence of the C-allele in a given geographical location (28,29).

154 PfCSP is predominantly a surface protein of sporozoite and plays a critical role in the invasion of hepatocytes (liver cells) (30,31). The gene encoding this protein is located 155 156 on chromosome 3 of *P. falciparum* within regions that encode epitopes recognized by 157 the human immune system. The C-terminal repeat region of the *Pfcsp* gene, which 158 codes for epitopes recognized by anti-CSP antibodies, contains tetrameric repeats that vary in both sequence and number of tetramers (32-34). For cell-mediated immunity, 159 160 the RTS,S/AS01 vaccine includes a fragment of the central NANP-NVDP repeats 161 polymorphic B-cell epitope region and a highly polymorphic C-terminal non-repeat 162 epitope region of *Pfcsp*, which covers CD4+ and CD8+ T-cell epitopes denoted as 163 TH2R and TH3R, respectively (9,35). Previous studies revealed high polymorphisms in 164 these regions within the C-terminal of the *Pfcsp* gene in natural parasite populations, 165 which might have resulted from natural selection by the host immune system (36,37).

166 Reports from previous studies of the population structure of *P. falciparum* genes coding 167 for vaccine antigens, including the *Pfcsp* showed variable levels of diversity and 168 geographic restriction of specific subgroups which may have an impact on the efficacy 169 of malaria vaccines based on this gene in specific geographic regions (36). In particular, 170 comparative diversity studies of the 3D7 *Pfcsp* gene showed that only 0.2% to 5.0% of 171 the vaccine strains matched the global *Pfcsp* gene (38). A larger number of these 172 studies described polymorphisms in the *Pfcsp* gene at the global level, with limited 173 local-specific studies.

Therefore, there is an urgent need to explore the extent of genetic diversity and natural 174 175 selection of the malaria vaccine antigen PfCSP in natural parasite populations in 176 Tanzania and other countries where the RTS,S/AS01 and R21/Matrix-M vaccines will 177 be deployed in the near future. This study investigated the genetic diversity and 178 population structure of the *Pfcsp* gene within five regions of Tanzania with varying malaria transmission intensities. In addition, the study aimed to determine whether 179 180 genetic diversity and signatures of selection in this important gene are present. The 181 findings present baseline data before the Ministry of Health rolls out any of the malaria 182 vaccines in an effort to eliminate malaria in Tanzania.

183 Materials and methods

184 Study area

185 Field studies that generated samples and data for this study were conducted between May 2013 and January 2015 as part of the Pathogen Diversity Network Africa (PDNA) 186 187 baseline surveys in 15 countries from Sub-Saharan Africa (SSA) (39–41). In Tanzania, the studies covered five districts from five regions with varying malaria transmission that 188 189 are located more than 800 km apart. The study districts included Muheza in the Tanga 190 region and Morogoro (moderate to low transmission area), Ujiji in Kigoma, Muleba in 191 Kagera and Nachingwea in Lindi (all with high transmission), as described elsewhere 192 (39) (Figure 1).

193 Study design and sampling

The samples were collected through a cross-sectional study (CSS), which was conducted in five districts of Morogoro, Muheza, Muleba, Nachingwea and Ujiji in 2013 while other samples were collected from two therapeutic efficacy studies (TES) which were conducted at Mkuzi Health Centre in 2013, and in Muheza designated district hospital and Ujiji health centre from May 2014 to January 2015 as described elsewhere (39). The TES were conducted according to the WHO protocol as previously described (42,43).

201 Sample collection, processing, DNA extraction and Sequencing

Whole blood samples were collected from patients with malaria infections who were 202 203 initially screened with rapid diagnostic tests for malaria (mRDTs) and the infections 204 were confirmed using microscopy as described earlier (39). Human white blood cells 205 (WBCs) were depleted from whole blood samples using CF11 cellulose columns (44), 206 and parasite genomic DNA was extracted using QIAamp DNA blood midi kits (Qiagen 207 GmbH, Hilden, Germany) following the manufacturer's instructions. DNA samples were 208 shipped to the Wellcome Trust Sanger Institute in Hinxton, UK, for whole genome 209 sequencing (WGS) using the Illumina HiSeq platform as part of the MalariaGEN 210 Plasmodium falciparum Community Project (45). Illumina sequencing libraries (200 bp 211 insert) were aligned to the reference P. falciparum 3D7 genome, after which variant 212 calling was conducted via the customized GATK pipeline (45). Each sample was 213 genotyped for 797,000 polymorphic biallelic coding single nucleotide polymorphisms (SNPs) across the genome, ensuring a minimum of 5x paired-end reads coverage 214

- 215 across each variant per sample. The dominant allele was retained in the genotype file at 216 loci with mixed reads (reference/non-reference).
- 217

236

218 Sequence acquisition and pre-processing

219 WGS data of 589 P. falciparum samples collected from the five districts of Morogoro (n = 32), Muheza (n = 297), Muleba (n = 52), Nachingwea (n = 65) and Ujiji (n = 143) were 220 221 downloaded from the MalariaGEN Plasmodium falciparum Community (Pf7k) project in 222 variant call format (VCF) at https://www.malariagen.net/resource/34. The VCF file was 223 generated following standardized protocols (46). In brief, reads mapping to the human reference genome were discarded before all analyses, and the remaining reads were 224 225 mapped to the *P. falciparum* 3D7 v3 reference genome using Burrows-Wheeler aligner 226 (bwa) mem version 0.7.15 (47). Binary Alignment Map (BAM) files were created using 227 the Picard tools (48) CleanSam, FixMateInformation and MarkDuplicates version 2.6.0 228 and GATK v3 (49) base quality score recalibration. Potential SNPs and indels were 229 discovered by running GATK's HaplotypeCaller independently across each sample 230 BAM file and genotyping these for each of the 16 reference sequences (14 chromosomes, 1 apicoplast and 1 mitochondria) using GATK's CombineGVCFs and 231 232 GenotypeGCVFs. SNPs and indels were filtered using GATK's Variant Quality Score 233 Recalibration (VQSR). All variants with a variant quality score log-odds (VQSLOD) 234 score ≤0 were filtered out and functional annotations were applied using snpEff version 4.1. Genome annotation was performed using vcftools version 0.1.10 and masked if 235 they were outside the core genome. Only biallelic SNPs from chromosome 3

that

237 passed all VCF filters were retained. Then SNP variants for the entire *Pfcsp* were

extracted from chromosome 3 (position: 221,323–222,516) for further analysis.

239 **Population genetics analysis**

240 The analysis included determining the minor allele frequency (MAF) distribution for all

putative SNPs within *Pfcsp* using Plink1.9, and rare alleles (MAF \leq 0.01) were removed

from the analysis. Wright's inbreeding coefficient (F_{ws}) (50) was determined using R

243 package moimix (51). F_{ws} refers to the number of different parasite strains contained

within an infection, and it ranges from 0 to 1. A sample is classified as having multiple

infections (polyclonal) when $F_{ws} < 0.95$ and a single infection (monoclonal) when $F_{ws} \ge$

246 0.95. A Pearson chi-square test was used to determine the statistical significance of any

247 differences observed in Fws estimates between the populations, and a standard

threshold of p < 0.05 was considered statistically significant.

249 To assess gene flow between parasite populations, genetic differentiation was first 250 estimated using the Wright Fixation index (F_{ST}) (52) using Vcftools v0.1.5 and 251 population structure was determined using principal component analysis (PCA) as 252 implemented in PLINK1.9. $F_{ST} < 0.05$ indicates minimal population differentiation or gene 253 flow between population pairs. Haplotype diversity (the number of two randomly 254 selected strains within the population having different haplotypes) in *Pfcsp* was determined by studying the variants in the C-terminal region of the gene (221422-255 221583 on chromosome 3). A total of 229 FASTA DNA sequences were reconstructed 256

from the retained monoclonal VCF samples and the *Pfcsp* sequence of the 3D7 strain was included as a reference. The monoclonal VCF samples were obtained based on the F_{ws} results, where samples with $F_{ws} \ge 0.95$ were considered a single infection

260 (monoclonal). The average number of pairwise nucleotide diversity (K), haplotype

diversity (Hd) and nucleotide diversity were calculated using DnaSP version 6.12.03 (53). To assess the genetic relationships between *Pfcsp* C-terminal haplotypes in the five districts, the haplotype networking for 229 sequences of *Pfcsp* was analyzed using NETWORK version 10.2 with the Median-joining algorithm (54).

265 Nucleotide diversity (π) was used to measure the degree of polymorphism within the 266 parasite population, and it was analyzed by calculating the pairwise difference between all possible pairs of individuals in the samples using Vcftools v0.1.5 (55) in the sliding 267 windows, with a size of 1000 bp. The Tajima's D statistical test (56) was used to detect 268 269 evidence of balancing selection or purifying selection in the *Pfcsp*. Tajima's D was 270 calculated in *Pfcsp* monoclonal samples using Vcftools v0.1.5 in 1000 bp sliding 271 windows. Tajima's D test compares the average pairwise differences (π) and the total 272 number of segregating sites (S) (57) and it was assessed using MEGA software version 273 11.0.13 (58) and DnaSP version 6.12.03 (53).

274 Results

275 Within-host parasite diversity estimation and statistical analysis

276 Within-host diversity in the Pfcsp gene was assessed using the inbreeding coefficient 277 (F_{ws}) and theresults indicated high within-host diversity (polyclonality), varying from

278 69.23% in Nachingwea to 56.9% in Muheza. Nevertheless, the observed difference did

not reach statistical significance (*p*= 0.3596)(Figure 2)

280 Genetic differentiation and structure analysis

281 Genetic differentiation in the Pfcsp population was analyzed by calculating Wright's fixation index (F_{ST}). The mean F_{ST} across five districts—Morogoro, Muheza, Muleba, 282 Nachingwea and Ujiji was very low at 0.0068 (which is < 0.05) (Figure 3a). This 283 284 indicates low genetic differentiation among the parasite populations or a high exchange 285 of genetic materials despite the districts being geographically separated with a distance of over 800 km. Principal component analysis (Figure 3b) showed no population 286 structure between the five districts. The overall results suggest gene flow between the 287 288 populations.

289 Genetic diversity of *Pfcsp* C-terminal within parasite populations

290 To assess the extent of genetic diversity and similarity within and between the five 291 populations, an investigation into the diversity in the C-terminal region of Pfcsp 292 (221,422-221,583) was done and summarized in a median-join haplotype diversity 293 network (Figure 4). In total, 50 haplotypes were observed among the 230 Pfcsp sequences, including the 3D7 sequence. Of these, more haplotypes were identified in 294 295 Ujiji (24) and Muheza (27) while the other districts had fewer haplotypes between 8 and 296 14 (Morogoro = 8, Muleba, 14, Nachingwea = 14). The RTS, S vaccine haplotype (Pf3D7-type) was found in 2% (H 50) of all samples. Four haplotypes (8%) were shared 297 298 by *Pfcsp* sequences from all five regions, indicating genetic closeness between these populations. Haplotype 5 emerged as the most prevalent *Pfcsp* C-terminal haplotype, 299

representing 16.5% (38/230) of the isolates. Among the haplotypes, 50% appeared only
once (singleton haplotypes).

302 Nucleotide diversity and signatures of selection in the C-terminal non-repeat

303 region of the Pfcsp gene

The average number of pairwise nucleotide differences (K) was higher in Muleba (4.1) 304 and Uiiii (4.0) compared to other districts with 3.3 - 3.4 (in Morogoro = 3.4, Muheza = 305 306 3.4, and Nachingwea = 3.3). The highest haplotype diversity (Hd) was observed in 307 Muleba (0.976±0.008) and Ujijji (0.973±0.005) and the lowest was observed in 308 Morogoro (0.917±0.035). The highest nucleotide diversity (π) was observed in Muleba 309 (0.0038) and Muheza (0.0037) and the lowest was observed in Morogoro (0.0029) but 310 the differences were not significant (p > 0.05). Evidence of natural selection in the C-311 terminal non-repeat region of the *PfCcsp* gene was tested using Tajima's D and there 312 were slightly positive values of 1.3, 1.3, 0.76 and 0.16 in Muheza, Muleba, Nachingwea 313 and Ujiji, respectively; while Morogoro showed negative value of - 0.21. However, these 314 values suggest evidence of weak balancing in the population (**Table 1**).

The nucleotide diversity peaked at the Th2R and Th3R T-cell epitopes, while the connecting region between Th2R and Th3R was conserved. Nucleotide diversity showed slightly different values between populations according to their geographic origin (**Figure 5**), with Muleba showing a higher peak > 0.07 in Th2R epitopes than the other four districts. The extended haplotype homozygosity revealed some extended haplotypes from the focal SNP locus 221,554 in all populations, but no long range haplotypes extended beyond 221,554 (**Figure 6**)

322

323 Discussion

medRxiv preprint doi: https://doi.org/10.1101/2024.01.23.24301587; this version posted January 23, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

324 Tanzania is one of the malaria-endemic countries with a high number of cases and 325 deaths in SSA and malaria is still a major health problem despite intensified 326 interventions in the past two decades (2). Recent studies show that malaria 327 transmission in Tanzania has become heterogeneous with persistently high burden in the western, north-west and southern regions for over 20 years (59) but lower 328 transmission intensities in other regions attributed to the ongoing control efforts. The 329 330 Tanzanian NMCP is implementing targeted interventions to reduce the burden in areas 331 with high transmission and eliminate malaria in regions that have consistently reported low transmission over the past 20 years (60). Among the interventions, NMCP is 332 333 considering the deployment and use of malaria vaccines following their approval by 334 WHO in 2021 and 2023 (2,61). This study was therefore undertaken to generate 335 baseline data on the genetic diversity, population structure and signature of selection in 336 the Pfcsp gene before the roll-out of malaria vaccines in Tanzania. The data will 337 potentially be useful in monitoring the performance of the vaccines after their 338 introduction and use in Tanzania to look for shifts in variations - or new variations- that 339 may arise due to vaccine pressure.

In general, the study showed no clear differentiation in the patterns of the *Pfcsp* gene polymorphism despite the study sites being widely separated geographically (with over 800 km apart), which is consistent with global patterns of *Pfcsp* variations which have been reported elsewhere (62,63). The lack of clear differentiation was according to the pairwise index of differentiation analysis (F_{ST} below 0.05) and confirmed by PCA. Similar findings were reported in Ghana where *pfcsp* sequences collected from two sites located over 784.4 km apart showed a largely homogeneous population with F_{ST} below 0.05, and without any population structure as detected by PCA (62). This and other studies of the *Pfcsp* gene suggested a continued gene flow among parasites or human population mixing between the areas which might have contributed to the observed findings (62,64,65). High genetic diversity and a lack of population differentiation may have important implications related to the efficacy and emergence as well spread of vaccine-resistant parasites (66,67).

353 The C-terminal of the *Pfcsp* gene displayed high genetic diversity in the study areas. A 354 comparative sliding window plot analysis of nucleotide diversity in the C-terminal region 355 displayed predictable results of two peaks at the Th2R and Th3R regions, suggesting 356 that genetic variations were mainly concentrated in these regions. Differences were also 357 observed among parasites from different geographical locations. The overall nucleotide 358 diversity was 0.0035 with Muleba showing a high peak in Th2R epitopes. A similar 359 pattern of nucleotide polymorphism was found in other African countries including 360 Malawi, The Gambia, Nigeria, Ghana, Mali and Guinea (8,38,62).

361

The overall haplotype diversity in the C-terminal of the *Pfcsp* gene in Tanzanian isolates 362 363 was 0.937 which is similar to those previously observed in Cameroon, The Gambia, 364 Ghana, Senegal, Congo, Malawi and Guinea but higher than those observed in non-African countries (8,68). This suggests that the C-terminal of the *Pfcsp* gene in Africa 365 366 has a higher level of diversity compared to other parts of the world (37). The values of 367 Tajima's D in the C-terminal region of the *Pfcsp* exhibited slightly positive values which suggest evidence of weak balancing selection in response to host immune pressure on 368 369 this region (8). Similar results have been reported showing weak variant-specific

370 selection within this region, as shown by a study in Malawi (69). Evidence of balancing 371 selection has been reported in other malaria vaccine candidates such as the 372 extracellular domains of P. falciparum apical membrane antigen (PfAMA-1), Cell-373 Traversal protein for Ookinetes and Sporozoites (CeITOS), Thrombospondin Related Adhesion Protein (TRAP), Liver Stage Antigen 1 (LSA1) and merozoite surface proteins 374 375 (MSP1) (70-73). Evidence of recent directional selection (iHS) was not observed in all 376 study areas and extended haplotype homozygosity showed no extended haplotypes 377 from the focal SNP locus (221,554)(69,74).

378 The haplotype network analysis in the C-terminal showed a sequence connection 379 between the five areas, which suggests they were closely related. This relatedness is in 380 agreement with the F_{ST} and PCA results which showed limited separation in isolates 381 from the five districts. The current RTS.S, recombinant vaccine was constructed with 382 pfcsp of P. falciparum 3D7 strain, but in Tanzanian isolates, only 0.7% 3D7-matched 383 Tanzanian pfcsp haplotypes. Low matching of the 3D7 to the natural Pfcsp has been 384 reported in many other studies in Africa (8,38,64,68). This data is consistent with other studies conducted across Africa and worldwide, which demonstrate low 3D7 haplotypes 385 386 and little population structure. However, other studies of the genetic diversity and 387 protective efficacy of the RTS,S/AS01 vaccine suggested that high levels of mismatch between natural isolates and 3D7 may weaken vaccine efficacy (28,68). In addition, the 388 389 high degree of location-specific Pfcsp diversity observed in Tanzania might result in 390 differences in vaccine efficacy potentially reducing RTS,S/AS01 vaccine effectiveness. Monitoring of differential vaccine efficacy according to Pfcsp haplotypes during 391 392 RTS,S/AS01 implementation programmes will be valuable, particularly in high

transmission areas, where a post-vaccination expansion of non-vaccine haplotypes is likely to be observed which could lead to reduced vaccine efficacy and vaccine breakthrough infections (64). Going forward and ensuring the vaccine is efficacious in all endemic areas, the design of a malaria vaccine may consider developing a vaccine that works for specific regions based on the genetic architecture of the parasites found in those areas.

399 Conclusion

400 These findings demonstrate a significant level of genetic diversity in the *Pfcsp* gene 401 within the study areas. Although there is limited population differentiation and structure, there is a high degree of polyclonality, particularly in Morogoro, Muleba and 402 403 Nachingwea. Furthermore, there is no genetic structure across all the study districts, 404 indicating a high level of gene flow and genetic exchange within the *Pfcsp* gene. In 405 addition, the C-terminal region of the Pfcsp gene from Muleba and Muheza exhibited 406 higher nucleotide diversity compared to Morogoro, Nachingwea and Uijii. Additionally, 407 the Pfcsp gene displayed slight positive Tajima's D values in all five studied P. falciparum populations. These findings are consistent with the hypothesis of balancing 408 409 selection acting on the Th2R and Th3R regions of the gene. However, it is important to 410 note that further studies and continuous monitoring of the *Pfcsp* gene are necessary. 411 These future investigations should encompass other regions and incorporate more 412 recent data to comprehensively assess the current diversity and dynamics of the *Pfcsp* 413 gene. By expanding the understanding of the genetic variations within the *Pfcsp* gene, it creat an opportunity to improve the knowledge of the parasite's adaptability and 414 415 contribute to the development of effective malaria control strategies

416

417 Availability of data and materials

418 The datasets generated during the current study are available in the MalariaGEN

- 419 *Plasmodium falciparum* Community (Pf7k) Project
- 420 https://www.malariagen.net/resource/34.
- 421

422 Acknowledgement

Authors would like to thank the study participants, local health workers and technicians who participated in the sample collection. Whole genome sequencing was done at the Wellcome Sanger Institute as part of the MalariaGen *Plasmodium falciparum* Community Project.

427

428 Ethical approval

429 The study utilized sequence data downloaded from the open-source, MalariaGEN 430 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2928508/). The study protocols for the projects that collected the samples were reviewed and approved by the Tanzanian 431 Medical Research Coordinating Committee (MRCC) of the National Institute for Medical 432 433 Research (NIMR). Permission to conduct the study in Morogoro, Muheza – Tanga, 434 Muleba – Kagera, Nachingwea - Lindi and Ujiji - Kigoma was sought in writing and 435 obtained from the district and regional medical officers. Written informed consent was 436 obtained from patients or parents/guardians in the case of children. Appropriate information (about the study and the protocol/methods) in a language that was 437 438 understood by the parents/guardians of the study patients was compiled and provided

- 439 before consent was obtained. Permission to publish was sought and obtained from the
- 440 Director General of NIMR

441 **Authors' contributions**

- 442 DSI formulated the original idea, supervised data analysis and wrote the manuscript;
- 443 BL performed the analysis and wrote the manuscript
- 444 CB, ZPH and DG supported the data analysis, reviewed and edited the manuscript
- 445 RB, MS and CIM conceived the idea, implemented the field surveys, reviewed and
- 446 edited the manuscript
- 447 DP and RM reviewed and edited the manuscript
- 448 JJ, JB, and DSI critically reviewed the manuscript
- 449 All authors contributed to the article and approved the submitted version.
- 450

451 **Competing Financial Interests:**

- 452 The authors declare that they have no known competing financial interests or personal
- 453 relationships that could have influenced the work reported in this paper.
- 454

455

456 **Funding**

- 457 This publication uses data from the MalariaGEN *Plasmodium falciparum* Community
- 458 Project as described in 'An open dataset of *Plasmodium falciparum* genome variation in
- 459 7,000 worldwide samples.
- 460 This study was part of BL's postdoctoral fellowship which was supported, in whole, by
- the Bill & Melinda Gates Foundation [grant number 02202]. Under the grant conditions

- of the Foundation, a Creative Commons Attribution 4.0 Generic License has already
 been assigned to the Author Accepted Manuscript version that might arise from this
 submission. JJ also received funding from NIH K24AI134990.



- 470 Figure 1: Map of Tanzania showing *P. falciparum* prevalence as a color gradient, and471 the location of the study regions (75).



Figure 2: With in-host diversity in the samples from Morogoro, Muheza, Muleba, Nachingwea, and Ujiji populations. The differences were not statistically significant (p= 0.3596).



Figure 3: Population differentiation and structure. **a**. Weir and Cockerham's F_{ST} among

parasites from Morogoro, Muheza, Muleba, Nachingwea and Ujiji. b. The plot of the first
(PC1) and second principal components (PC2) of samples from Morogoro, Muheza,
Muleba, Nachingwea and Ujiji.

519 Table 1: Diversity of the Pfcsp C-terminal region of samples included in the

520 network analysis

- 521
- 522

Population	n	н	S	k	Π	Hd	Tajima's D
Morogoro	10	8	15	3.4	0.0198	0.978±0.035	-0.235
Muheza	128	27	18	4.3	0.0027	0.925±0.005	0.12
Muleba	19	14	15	4.008	0.0037	0.970±0.016	0.44
Nachingwea	20	14	11	3.920	0.00033	0.953±0.008	0.41
Ujiji	52	24	17	4.008	0.0027	0.947±0.005	0.16

523

524 *n* number of sequences, *h* number of unique haplotypes, *s* number of segregating sites, 525 k average of pairwise nucleotides differences, π nucleotide diversity, *Hd* haplotype 526 diversity

527



529

Figure 4: Haplotype network analysis of the *Pfcsp* gene among parasites from Morogoro, Muheza, Muleba, Nachingwea and Ujiji.The size of each node indicated the proportion of the total haplotype frequencies. Each haplotype is donated as "H" with the vaccine strain 3D7 donated as "3D7" in Hap_50. The color of each node corresponds to the site of sample origin.

535

536

537



Figure 5: Nucleotide diversity of the C-terminal region of the *Pfcsp* gene indicated high diversity in both Th2R and Th3R, while the connecting region between Th2R and Th3R remained conserved.



Figure 6: Plots show Bifurcation diagram and Extended Haplotype Homozygosity (EHH). a) Bifurcation diagrams showing the breakdown of these extended haplotypes from increasing distances in the parasite population **b)** Plots of EHH showing extended haplotypes from a focal SNP locus (221,554).

578 Bibliography

- 579 1. WHO. World malaria report 2021. WHO; 2021.
- 580 2. WHO. World Malaria Report 2022. WHO [Internet]. 2022 [cited 2022 Dec 10]; Available
 581 from: https://apps.who.int/iris/rest/bitstreams/1484818/retrieve
- Mitchell CL, Ngasala B, Janko MM, Chacky F, Edwards JK, Pence BW, et al. Evaluating
 malaria prevalence and land cover across varying transmission intensity in Tanzania
 using a cross-sectional survey of school-aged children. Malar J. 2022 Mar 9;21(1):80.
- 585 4. WHO. GLOBAL TECHNICAL STRATEGY FOR MALARIA. WHO. 2021;
- 586 5. Feachem RGA, Chen I, Akbari O, Bertozzi-Villa A, Bhatt S, Binka F, et al. Malaria
 587 eradication within a generation: ambitious, achievable, and necessary. Lancet. 2019 Sep
 588 21;394(10203):1056–112.
- 589 6. Dhiman S. Are malaria elimination efforts on right track? An analysis of gains achieved
 590 and challenges ahead. Infect Dis Poverty. 2019 Feb 13;8(1):14.
- 591 7. Guyant P, Corbel V, Guérin PJ, Lautissier A, Nosten F, Boyer S, et al. Past and new
 592 challenges for malaria control and elimination: the role of operational research for
 593 innovation in designing interventions. Malar J. 2015 Jul 17;14:279.
- Huang H-Y, Liang X-Y, Lin L-Y, Chen J-T, Ehapo CS, Eyi UM, et al. Genetic
 polymorphism of Plasmodium falciparum circumsporozoite protein on Bioko Island,
 Equatorial Guinea and global comparative analysis. Malar J. 2020 Jul 13;19(1):245.
- 597 9. Casares S, Brumeanu T-D, Richie TL. The RTS,S malaria vaccine. Vaccine.
 598 2010;28(4880–94).
- 599 10. Koff WC, Gust ID, Plotkin SA. Toward a human vaccines project. Nat Immunol. 2014
 Jul;15(7):589–92.
- 11. Nadeem AY, Shehzad A, Islam SU, Al-Suhaimi EA, Lee YS. Mosquirix[™] RTS, S/AS01
 vaccine development, immunogenicity, and efficacy. Vaccines (Basel). 2022 Apr 30;10(5).
- Duffy PE, Patrick Gorres J. Malaria vaccines since 2000: progress, priorities, products. npj
 Vaccines. 2020 Jun 9;5:48.
- Le Roch KG, Chung DWD, Ponts N. Genomics and integrated systems biology in
 Plasmodium falciparum: a path to malaria control and eradication. Parasite Immunol.
 2012 Mar;34(2–3):50–60.
- Mohamed NS, AbdElbagi H, Elsadig AR, Ahmed AE, Mohammed YO, Elssir LT, et al.
 Assessment of genetic diversity of Plasmodium falciparum circumsporozoite protein in
 Sudan: the RTS,S leading malaria vaccine candidate. Malar J. 2021 Nov 10;20(1):436.
- Kurtovic L, Drew DR, Dent AE, Kazura JW, Beeson JG. Antibody Targets and Properties
 for Complement-Fixation Against the Circumsporozoite Protein in Malaria Immunity. Front
 Immunol. 2021 Dec 1;12:775659.
- 614 16. WHO. WHO Guidelines for malaria. WHO Geneva, Switzerland. 2022;

Stoute JA, Slaoui M, Heppner DG, Momin P, Kester KE, Desmons P, et al. A preliminary
evaluation of a recombinant circumsporozoite protein vaccine against Plasmodium
falciparum malaria. RTS,S Malaria Vaccine Evaluation Group. N Engl J Med. 1997 Jan
9;336(2):86–91.

- Regules JA, Cummings JF, Ockenhouse CF. The RTS,S vaccine candidate for malaria.
 Expert Rev Vaccines. 2011 May;10(5):589–99.
- Asante KP, Abdulla S, Agnandji S, Lyimo J, Vekemans J, Soulanoudjingar S, et al. Safety
 and efficacy of the RTS,S/AS01E candidate malaria vaccine given with expandedprogramme-on-immunisation vaccines: 19 month follow-up of a randomised, open-label,
 phase 2 trial. Lancet Infect Dis. 2011 Oct;11(10):741–9.
- Bojang KA, Milligan PJ, Pinder M, Vigneron L, Alloueche A, Kester KE, et al. Efficacy of
 RTS,S/AS02 malaria vaccine against Plasmodium falciparum infection in semi-immune
 adult men in The Gambia: a randomised trial. Lancet. 2001 Dec 8;358(9297):1927–34.
- Aponte JJ, Aide P, Renom M, Mandomando I, Bassat Q, Sacarlal J, et al. Safety of the
 RTS,S/AS02D candidate malaria vaccine in infants living in a highly endemic area of
 Mozambique: a double blind randomised controlled phase I/IIb trial. Lancet. 2007 Nov
 3;370(9598):1543–51.
- RTS,S Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with
 or without a booster dose in infants and children in Africa: final results of a phase 3,
 individually randomised, controlled trial. Lancet. 2015 Jul 4;386(9988):31–45.
- Asante KP, Adjei G, Enuameh Y, Owusu-Agyei S. RTS,S malaria vaccine development:
 progress and considerations for postapproval introduction. VDT. 2016 Jun;25.
- 637 24. Collins KA, Snaith R, Cottingham MG, Gilbert SC, Hill AVS. Enhancing protective
 638 immunity to malaria with a highly immunogenic virus-like particle vaccine. Sci Rep. 2017
 639 Apr 19;7:46621.
- Ballou WR, Rothbard J, Wirtz RA, Gordon DM, Williams JS, Gore RW, et al.
 Immunogenicity of synthetic peptides from circumsporozoite protein of Plasmodium
 falciparum. Science. 1985 May 24;228(4702):996–9.
- 643 26. Datoo MS, Natama MH, Somé A, Traoré O, Rouamba T, Bellamy D, et al. Efficacy of a
 644 low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal
 645 administration to children in Burkina Faso: a randomised controlled trial. Lancet. 2021
 646 May 15;397(10287):1809–18.
- 647 27. Datoo MM, Dicko A, Tinto H, Ouédraogo J-B, Hamaluba M, Olotu A, et al. A Phase III
 648 Randomised Controlled Trial Evaluating the Malaria Vaccine Candidate R21/Matrix-M[™] in
 649 African Children. 2023;
- 850 28. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic
 diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. N Engl J Med. 2015
 852 Nov 19;373(21):2025–37.
- 653 29. Chaudhury S, MacGill RS, Early AM, Bolton JS, King CR, Locke E, et al. Breadth of
 654 humoral immune responses to the C-terminus of the circumsporozoite protein is

associated with protective efficacy induced by the RTS,S malaria vaccine. Vaccine. 2021
Feb 5;39(6):968–75.

- 857 30. Pinzon-Ortiz C, Friedman J, Esko J, Sinnis P. The binding of the circumsporozoite protein
 858 to cell surface heparan sulfate proteoglycans is required for plasmodium sporozoite
 859 attachment to target cells. J Biol Chem. 2001 Jul 20;276(29):26784–91.
- 860 31. Rathore D, Sacci JB, de la Vega P, McCutchan TF. Binding and invasion of liver cells by
 861 Plasmodium falciparum sporozoites. Essential involvement of the amino terminus of
 862 circumsporozoite protein. J Biol Chem. 2002 Mar 1;277(9):7092–8.
- Gandhi K, Thera MA, Coulibaly D, Traoré K, Guindo AB, Ouattara A, et al. Variation in the
 circumsporozoite protein of Plasmodium falciparum: vaccine development implications.
 PLoS ONE. 2014 Jul 3;9(7):e101783.
- Gandhi K, Thera MA, Coulibaly D, Traoré K, Guindo AB, Doumbo OK, et al. Next
 generation sequencing to detect variation in the Plasmodium falciparum circumsporozoite
 protein. Am J Trop Med Hyg. 2012 May;86(5):775–81.
- 669 34. Egan, JE, Hoffman SL, Gordon DM. Humoral immune responses in volunteers immunized
 670 with irradiated Plasmodium falciparum sporozoites . Cover The American Journal of
 671 Tropical Medicine and Hygiene The American Journal of Tropical Medicine and Hygiene.
 672 1993;
- 673 35. Plassmeyer ML, Reiter K, Shimp RL, Kotova S, Smith PD, Hurt DE, et al. Structure of the
 674 Plasmodium falciparum circumsporozoite protein, a leading malaria vaccine candidate. J
 675 Biol Chem. 2009 Sep 25;284(39):26951–63.
- Barry AE, Schultz L, Buckee CO, Reeder JC. Contrasting population structures of the
 genes encoding ten leading vaccine-candidate antigens of the human malaria parasite,
 Plasmodium falciparum. PLoS ONE. 2009 Dec 30;4(12):e8497.
- 37. Zeeshan M, Alam MT, Vinayak S, Bora H, Tyagi RK, Alam MS, et al. Genetic variation in
 the Plasmodium falciparum circumsporozoite protein in India and its relevance to RTS,S
 malaria vaccine. PLoS ONE. 2012 Aug 17;7(8):e43430.
- 682 38. Lê HG, Kang J-M, Moe M, Jun H, Thái TL, Lee J, et al. Genetic polymorphism and natural
 683 selection of circumsporozoite surface protein in Plasmodium falciparum field isolates from
 684 Myanmar. Malar J. 2018 Oct 12;17(1):361.
- 39. Ishengoma D, Shayo A, Mandara C, Baraka V, Madebe R, Ngatunga, et al. The Role of
 Malaria Rapid DiagnosticTests in Screening of Patients to be Enrolled in Clinical Trials in
 Low Malaria Transmission Settings. iMedPub Journals. 2016 Mar 16;
- 40. MalariaGEN, Ahouidi A, Ali M, Almagro-garcia J, Amambua-ngwa A, Amaratunga C, et al.
 An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples.
 Wellcome Open Research. 2021;1–22.
- 691 41. Ghansah A, Amenga-Etego L, Amambua-Ngwa A, Andagalu B, Apinjoh T, Bouyou-Akotet
 692 M, et al. Monitoring parasite diversity for malaria elimination in sub-Saharan Africa.
 693 Science. 2014 Sep 12;345(6202):1297–8.

694 42. Shayo A, Mandara CI, Shahada F, Buza J, Lemnge MM, Ishengoma DS. Therapeutic
695 efficacy and safety of artemether-lumefantrine for the treatment of uncomplicated
696 falciparum malaria in North-Eastern Tanzania. Malar J. 2014 Sep 20;13:376.

- Mandara CI, Kavishe RA, Gesase S, Mghamba J, Ngadaya E, Mmbuji P, et al. High
 efficacy of artemether-lumefantrine and dihydroartemisinin-piperaquine for the treatment
 of uncomplicated falciparum malaria in Muheza and Kigoma Districts, Tanzania. Malar J.
 2018 Jul 11;17(1):261.
- Venkatesan M, Amaratunga C, Campino S, Auburn S, Koch O, Lim P, et al. Using CF11
 cellulose columns to inexpensively and effectively remove human DNA from Plasmodium
 falciparum-infected whole blood samples. Malar J. 2012 Feb 10;11:41.
- 70445.MalariaGEN. Pf7: an open dataset of Plasmodium falciparum genome. Wellcome Open705Research. 2023;8(22).
- 46. MalariaGEN, Ahouidi A, Ali M, Almagro-Garcia J, Amambua-Ngwa A, Amaratunga C, et
 al. An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide
 samples. Wellcome Open Res. 2021 Jul 13;6:42.
- Jung Y, Han D. BWA-MEME: BWA-MEM emulated with a machine learning approach.
 BioRxiv. 2021 Sep 1;
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a
 tool set for whole-genome association and population-based linkage analyses. Am J Hum
 Genet. 2007 Sep;81(3):559–75.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
 sequencing data. Genome Res. 2010 Sep;20(9):1297–303.
- Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis
 of *Plasmodium falciparum* diversity in natural infections by deep sequencing. Nature.
 2012 Jul 19;487(7407):375–9.
- 51. Lee S, Harrison A, Tessier N, Tavul L, Miotto O. Assessing clonality in malaria parasites
 from massively parallel sequencing data. F1000Research. 2015;4(1043).
- 52. Chakraborty R, Danker-Hopfe H. Analysis of population structure: A comparative study of
 different estimators of wright's fixation indices. Statistical methods in biological and
 medical sciences. Elsevier; 1991. p. 203–54.
- 725 53. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins
 726 SE, et al. Dnasp 6: DNA sequence polymorphism analysis of large data sets. Mol Biol
 727 Evol. 2017 Dec 1;34(12):3299–302.
- 54. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific
 phylogenies. Mol Biol Evol. 1999 Jan;16(1):37–48.
- 55. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.
- 732 56. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA

- 733 polymorphism. Genetics. 1989 Nov;123(3):585–95.
- Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993 Mar;133(3):693–
 709.
- 736 58. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis
 737 version 11. Mol Biol Evol. 2021 Jun 25;38(7):3022–7.
- Thawer SG, Chacky F, Runge M, Reaves E, Mandike R, Lazaro S, et al. Sub-national
 stratification of malaria risk in mainland Tanzania: a simplified assembly of survey and
 routine data. Malar J. 2020 May 8;19(1):177.
- MoHCDGEC. NATIONAL MALARIA STRATEGIC PLAN 2021-2025, TRANSITIONING
 TO MALARIA ELIMINATION IN PHASES. Ministry of Health, Community Development,
 Gender, Elderly and Childre. 2021;
- 61. WHO. WHO recommends R21/Matrix-M vaccine for malaria prevention in updated adviceon immunization. WHO. 2023;
- Aragam NR, Thayer KM, Nge N, Hoffman I, Martinson F, Kamwendo D, et al. Diversity of
 T cell epitopes in Plasmodium falciparum circumsporozoite protein likely due to proteinprotein interactions. PLoS ONE. 2013 May 7;8(5):e62427.
- Pringle JC, Wesolowski A, Berube S, Kobayashi T, Gebhardt ME, Mulenga M, et al. High
 Plasmodium falciparum genetic diversity and temporal stability despite control efforts in
 high transmission settings along the international border between Zambia and the
 Democratic Republic of the Congo. Malar J. 2019 Dec 4;18(1):400.
- Amegashie EA, Amenga-Etego L, Adobor C, Ogoti P, Mbogo K, Amambua-Ngwa A, et al.
 Population genetic analysis of the Plasmodium falciparum circumsporozoite protein in two distinct ecological regions in Ghana. Malar J. 2020 Nov 27;19(1):437.
- He Z-Q, Zhang Q-Q, Wang D, Hu Y-B, Zhou R-M, Qian D, et al. Genetic polymorphism of
 circumsporozoite protein of Plasmodium falciparum among Chinese migrant workers
 returning from Africa to Henan Province. Malar J. 2022 Aug 27;21(1):248.
- 66. Duffy CW, Ba H, Assefa S, Ahouidi AD, Deh YB, Tandia A, et al. Population genetic
 structure and adaptation of malaria parasites on the edge of endemic distribution. Mol
 Ecol. 2017 Jun;26(11):2880–94.
- Amambua-Ngwa A, Tetteh KKA, Manske M, Gomez-Escobar N, Stewart LB, Deerhake
 ME, et al. Population genomic scan for candidate signatures of balancing selection to
 guide antigen characterization in malaria parasites. PLoS Genet. 2012 Nov
 1;8(11):e1002992.
- Pringle JC, Carpi G, Almagro-Garcia J, Zhu SJ, Kobayashi T, Mulenga M, et al.
 RTS,S/AS01 malaria vaccine mismatch observed among Plasmodium falciparum isolates
 from southern and central Africa and globally. Sci Rep. 2018 Apr 26;8(1):6622.
- Bailey JA, Mvalo T, Aragam N, Weiser M, Congdon S, Kamwendo D, et al. Use of
 massively parallel pyrosequencing to evaluate the diversity of and selection on
 Plasmodium falciparum csp T-cell epitopes in Lilongwe, Malawi. J Infect Dis. 2012 Aug

- 772 15;206(4):580–7.
- 773 70. Polley SD, Conway DJ. Strong diversifying selection on domains of the Plasmodium
 774 falciparum apical membrane antigen 1 gene. Genetics. 2001 Aug;158(4):1505–12.
- 775 71. Nirmolia T, Ahmed MA, Sathishkumar V, Sarma NP, Bhattacharyya DR, Mohapatra PK, et
 776 al. Genetic diversity of Plasmodium falciparum AMA-1 antigen from the Northeast Indian
 777 state of Tripura and comparison with global sequences: implications for vaccine
 778 development. Malar J. 2022 Feb 22;21(1):62.
- 779 72. Ajibola O, Diop MF, Ghansah A, Amenga-Etego L, Golassa L, Apinjoh T, et al. In silico
 780 characterisation of putative *Plasmodium falciparum* vaccine candidates in African malaria
 781 populations. Sci Rep. 2021 Aug 10;11(1):16215.
- 782 73. Chenet SM, Branch OH, Escalante AA, Lucas CM, Bacon DJ. Genetic diversity of vaccine
 783 candidate antigens in Plasmodium falciparum isolates from the Amazon basin of Peru.
 784 Malar J. 2008 May 27;7:93.
- 785 74. Reeder JC, Wapling J, Mueller I, Siba PM, Barry AE. Population genetic analysis of the
 786 Plasmodium falciparum 6-cys protein Pf38 in Papua New Guinea reveals domain-specific
 787 balancing selection. Malar J. 2011 May 14;10:126.
- 788 75. DHS. Malaria Indicator Survey 2017 [Internet]. 2018 [cited 2024 Jan 15]. Available from:
 789 https://dhsprogram.com/pubs/pdf/MIS31/MIS31.pdf