

1 **Stable *Plasmodium falciparum* merozoite surface protein-1 allelic diversity** 2 **despite decreasing parasitemia in children with multiple malaria infections**

3 **Authors**

4 Reuben M. Yaa^{a,b,c}, Kelvin M. Kimenyi^{a,d}, Palasciano H. Antonio^e, George Obiero^a, Lynette Isabella Ochola-
5 Oyier^{#,d}

6 **Affiliations**

- 7 a. Centre for Biotechnology and Bioinformatics, University of Nairobi, Nairobi, Kenya,
8 b. UK Dementia Research Institute, Imperial College London, UK,
9 c. Department of Brain Sciences, Imperial College London, London, UK,
10 d. KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya,
11 e. Department of Mathematics, Imperial College London, London, UK
12

13 #Author for correspondence

14 LiOchola@kemri-wellcome.org

16 **Keywords**

17 Malaria, *Pfmsp1*, diversity, microhaplotype, parasitaemia
18

19 **Abstract**

20 Individuals experiencing recurrent malaria infections encounter a variety of alleles with each new
21 infection. This ongoing allelic diversity influences the development of naturally acquired immunity and it
22 can inform vaccine efficacy. To investigate the diversity and infection variability, *Plasmodium falciparum*
23 merozoite surface protein 1 (*Pfmsp1*), a crucial protein for parasite invasion and immune response, was
24 assessed in parasites isolated from children with repeated malaria episodes. A total of eleven
25 microhaplotypes were observed across all malaria episodes. There were 4 prevalent microhaplotypes, E-
26 KSNG-L, Q-KSNG-L, Q-KSNG-F, and Q-KNNG-L, in the population. Conversely, microhaplotypes such as E-
27 KSSR-L, E-KNNG-L, E-KSSG-L, E-TSSR-L (3D7), Q-TSSR-L, E-TSSG-L, and E-KSNG-F were rare and maintained
28 at low frequencies. High allelic replacements were observed, however some individuals experienced
29 consecutive re-infections with the same microhaplotype. Notably, *Pfmsp1*₁₉ allelic diversity as measured
30 by haplotype diversity was stable, while nucleotide diversity decreased over time with decreasing

31 parasitemia. Parasite *Pfmsp1₁₉* allelic diversity remained stable over the multiple malaria episodes,
32 despite declining parasitaemia levels and ongoing immune development. In addition, our findings reveal
33 dynamic *Pfmsp1₁₉* allelic replacements across parasite infection episodes, with initial immune responses
34 to prevalent variants potentially waning over time, allowing for re-infections. Overall, our findings
35 emphasize that protective immunity to malaria involves complex immune responses targeting multiple
36 alleles that wane over time and thus the development of sterilizing immunity is a challenge.

37

38 **Introduction**

39 In an attempt to evade host immunity during infection, *Plasmodium falciparum* parasites regularly
40 replace merozoite antigen epitope conformations and thus have the option to use alternative invasion
41 pathways, disrupting complement activation (Herrera et al. 2015; Awandare et al. 2018; Larsen et al. 2018;
42 Oyong et al. 2018). At the genetic level, the changes may arise from point mutations leading to single
43 nucleotide polymorphisms (SNPs), insertions/deletions of one or more base residues as well as meiotic
44 recombination events of parental alleles, generating newer progeny (Mu et al. 2005; Miles et al. 2016).
45 This is further maintained by the phenomenon of balancing selection. This process stabilizes the
46 polymorphic circulation of immune targeted antigens resulting in multi-allelic circulation of these genes
47 especially in malaria-endemic regions (Ochola-Oyier et al. 2019). Effects of balancing selection have been
48 shown before in merozoite surface protein (MSP) 1 (Parobek et al. 2014), apical membrane antigen-1
49 (*ama1*) (Polley and Conway 2001), MSP3 (Polley et al. 2007), erythrocyte binding antigen-175 (EBA-175)
50 (Verra et al. 2006), MSP Duffy binding Ligand-1 and 2 (MSPDBL1 and MSPDBL2) (Tetteh et al. 2009; Ochola
51 et al. 2010), reticulocyte binding homologues-2 (Rh2) (Rayner et al. 2005; Reiling et al. 2010) and Rh5
52 (Ochola-Oyier et al. 2016).

53 Merozoite antigen alleles are maintained by frequency-dependent immune selection which shifts
54 allele frequencies over time. This is because the rare alleles are observed less and therefore rise to high
55 frequency (Conway et al. 1992). This pattern of selection supports allele-specific immunity, which has
56 repeatedly been shown to reduce overall vaccine efficacies that are based on low frequency or single
57 alleles in different malaria endemic settings (Takala et al. 2007a; Ogutu et al. 2009; Ouattara et al. 2013).
58 Of interest is *P falciparum* MSP1, which has been advanced over time as a vaccine candidate and it is a
59 suitable marker for genotyping parasite populations in antimalarial efficacy studies and clinical trials
60 (Mwingira et al. 2011). However, allelic replacements are associated with reduced efficacies on vaccine
61 formulations targeting this antigen (Rotman et al. 1999; John et al. 2004a; Chauhan et al. 2010; Wilson et
62 al. 2011).

63 *PfMSP1* is a predominant antigen on the surface of the asexual blood stage of the parasite that
64 plays an imperative role in erythrocyte invasion to cause malaria clinical symptoms. It is synthesized as a
65 large precursor during schizogony and subsequently processed via proteolytic cleavage into 5 fragments
66 of which the smallest is a 19kDa fragment (*PfMSP1₁₉*). This fragment has two epidermal growth factor
67 (EGF) domains, one located at the C-terminal and another at the N-terminal ends. The C-terminal interacts
68 with band 3, the erythrocyte receptor, to facilitate parasite erythrocyte invasion (Cowman and Crabb
69 2006). Inside the erythrocyte, the parasite multiplies and later egresses into the bloodstream following
70 the rupture of the erythrocyte, a process in which *PfMSP1₁₉* is also involved (Tan and Blackman 2021).
71 During egress, subtilisin-like (SUB1) parasite serine protease modifies the structure of *PfMSP1* to bind
72 spectrin, a component of the host erythrocyte cytoskeleton to facilitate egress (Das et al. 2015). Genetic
73 diversity studies of *PfMSP1* have highlighted that fewer polymorphisms are located at the 19kDa fragment
74 than the rest of the protein, a total of 6 polymorphic loci (Takala et al. 2007b). Probably, because of its
75 direct proximal interaction with its receptor. The 19kDa fragment is easily accessible to the host immune
76 system as evidenced by merozoite invasion and parasite growth inhibition with antibodies in *in vitro* and
77 mice experiments (Rotman et al. 1999; Cowman and Crabb 2006; Takala et al. 2007b). The fragment elicits
78 both humoral and cell mediated immune responses during exposure to natural infections (Egan et al.
79 1999; John et al. 2004b), particularly to the polymorphic amino acids at the second EGF-like domain (Egan
80 et al. 1995).

81 Allelic diversity of *PfMSP1* at C-terminal region, has been shown previously in malaria endemic
82 regions such as Kenya, Tanzania and Uganda (Egan et al. 1995; Langhorne et al. 2008; Kariuki et al. 2013;
83 Simpalipan et al. 2014). Similarly, significant epitope diversity through immune assays (Conway et al.
84 1992) have been demonstrated before in longitudinal studies. Though *PfMSP1* allelic diversity and
85 patterns have previously been investigated in different malaria endemic regions, it has not been assessed
86 in recurrent multiple infections in moderate to high malaria transmission regions. To achieve this, allelic
87 replacements and the distribution of C-terminal *Pfmsp1* microhaplotypes were determined over time in
88 multiple infections to describe *P. falciparum* infection diversity. Interrogating *Pfmsp1* microhaplotypes in
89 individual infections over time will shed light on the development of naturally acquired immunity with
90 repeated exposures to infection, while also providing valuable guidance for vaccine design and potential
91 optimal intervals for vaccine dosing.

92

93 **Methods**

94 Study design

95 We retrieved a total of 426 blood samples from 33 children (comprising 19 males and 14 females)
96 who experienced a minimum of 10 malaria episodes. These children were from a cohort that was originally
97 part of a larger integrated study on natural immunity to malaria established in 2005 (Mwangi et al. 2005),
98 in Junju, Kilifi County, Kenya, conducted under institutional ethical review (SERU 3149) with sampling done
99 from 2008-2013. A blood sample was obtained from every participant upon confirmation of a febrile
100 malaria episode. From the blood samples malaria parasitemia load was estimated using microscopy. In
101 this study, the blood samples were used to evaluate *PfMSP1* C-terminal coding region from the parasite
102 isolates.

103 DNA extraction, PCR and sequencing

104 Total genomic DNA from the 426 blood samples were extracted using the QIAamp Blood Mini Kit
105 (Qiagen). The 272bp *Pfmsp1* 19kDa coding region was amplified by Polymerase Chain Reaction (PCR) using
106 High Fidelity Taq polymerase (Sigma Aldrich, cat. no:11732641001) with *Pfmsp1*₁₉-F 5'-
107 CAATGCGTAAAAACAATGTCC-3' and *Pfmsp1*₁₉-R 5'-TTAGAGGAACTGCAGAAAATACCA-3' specific
108 primers pairs on cycling conditions as follows: 1 cycle at 94°C for 2 minutes, 9 cycles of 94°C for 30
109 seconds, 44°C for 30 seconds, 72°C for 2 minutes, 24 cycles of 94°C for 30 seconds, 44°C for 30 seconds,
110 72°C for 2 minutes + 5 seconds per cycle and a final step of 72°C for 2 minutes. The amplified PCR products
111 were separated by 2% (w/v) agarose gel electrophoresis in a buffer composed of 40mM Tris, 1mM EDTA
112 and 20mM Acetic acid (TAE), pH 8.2, for 40 minutes at 100V. PCR products were visualized using 1% SYBR
113 (v/v) safe stained agarose gels and cleaned using ethanol precipitation. Sequencing templates were
114 prepared using BigDye™ Terminator v3.1 cycle sequencing kit. A volume of 3µl of the purified products
115 was resuspended with, 4µl BigDye™ Terminator 3.1 ready reaction mix, 1µl of 10µM of *Pfmsp1*₁₉-F primer
116 used during fragment isolation and 3µl of deionized water to a total reaction volume of 10µl in 96 well
117 plate. Cycle sequencing of the amplicons was done using PCR as follows: 96°C for 1 minute, 25 cycles of
118 96°C for 10 seconds, 50°C with +1°C/second for 5 seconds and 60°C for 4 minutes. Sequencing reactions
119 were purified using ethanol/EDTA precipitation and reactions resuspended in Hi-Di™ formamide. The
120 plates were then analyzed using capillary electrophoresis on ABI 3500XL Genetic Analyzer outsourced
121 from Inqaba Biotechnical Industries (Pty), South Africa. Sequences were assembled, trimmed and edited
122 using Sequencher® 5.3 DNA analysis software (Gene Codes Corporations, Ann Arbo, MI USA) and CLC

123 sequence viewer version 7(QIAGEN). DNA sequence data and corresponding translated protein were
124 aligned to *P. falciparum* 3D7 *msp1* (PF3D7_0930300) *msp1* (PF3D7_0930300) reference sequence,
125 ASM276v2, using the MUSCLE alignment algorithm in the MEGA 11 program (Edgar 2004; Tamura et al.
126 2021). The sequences were deposited in the GenBank NIH genetic sequence database under accession
127 numbers (OQ821998 - OQ822147).

128 Data processing and statistical analyses

129 After standardizing the sequences to the same length (234bp) and excluding short sequences that
130 did not cover the segregating sites in either orientation, sequences were clustered using USEARCH v11
131 software (Edgar 2010) and Phyclust R package (Tzeng 2005) to identify microhaplotypes. This was followed
132 by determining microhaplotype frequencies. Phyclust applies grouping of microhaplotypes and
133 categorizes those to be retained above a cut-off point which is an optimal balance between the sample
134 size, microhaplotype number and frequencies (Tzeng 2005).

135 Microhaplotype sequences were extracted as an alignment and transformed to a DNABin object
136 using ape R package (Paradis et al. 2004). The object was transformed to a hamming distance matrix by
137 measuring pairwise distances of corresponding residues between microhaplotype pairs, while counting
138 differences between them and storing this on a symmetric matrix which was visualized as a heatmap
139 (Pinheiro et al. 2005).

140 Temporal microhaplotype distribution densities, frequencies, and patterns at the individual and
141 population levels were assessed. A likelihood ratio test was used to determine whether the haplotypes
142 were uniformly distributed in the population (Read and Cressie 2012). More specifically, modeling
143 microhaplotype occurrences using a multinomial distribution, the test evaluated uniform probabilities of
144 infections and compared them to probabilities computed via maximum likelihood estimation, providing
145 strong evidence in favour of the latter. The expected number of microhaplotypes per infection episode
146 was estimated as $p_i \times N_t$, where p_i denotes the maximum likelihood estimate of the probability of seeing
147 haplotype i and N_t the total number of observations at episode t . Where the expected versus observed
148 differed by a large amount, we tested if the discrepancies were significant using Holm's method to control
149 for the family-wise error rate (Holm 1977).

150 In order to characterize the distribution patterns of microhaplotype across the infections, we
151 employed kernel density probability estimates (Węglarczyk 2018), line graphs of microhaplotype
152 frequencies as proportions of the total number of patients and infection variability using the
153 microhaplotype data. The microhaplotype distribution densities and line graphs were compared between

154 prevalent and rare microhaplotypes. A population infection variability analysis was conducted to examine
155 the genetic diversity of the *Pfmsp1* locus using haplotype and nucleotide diversity indices. We assigned
156 microhaplotypes back to the patients to assess patient-microhaplotype distribution and proportions. The
157 time between infections for individuals was determined by calculating time elapsed between successive
158 infections.

159 To determine parasitemia and microhaplotype dynamics over time, the outcome of *Pfmsp1₁₉* PCR
160 on all the samples and genetic diversity of the microhaplotype was correlated using the spearman rank
161 method with malaria parasitemia. To do so, we only used data corresponding to the first 14 episodes,
162 since the amount of data collected beyond this episode was insufficient to appropriately carry out this
163 analysis. In addition, microhaplotypes that did not have respective parasitaemia data were excluded from
164 the analysis. The samples were grouped based on *Pfmsp1₁₉* PCR amplification status as either amplicon
165 present or absent. Parasitemia was determined by microscopic examination of blood films of *P. falciparum*
166 parasites by counting the number of parasites/200 white blood cells (Mwangi et al. 2005) (**Supplementary**
167 **Table 1**). The difference in parasitemia between the two groups was compared using the Wilcoxon rank-
168 sum test. We assessed the correlation of parasitemia and microhaplotype diversity as a measure of
169 infection on time dimension.

170

171 **Results**

172 Microhaplotype classes and associated specific patterns in the population

173 We genotyped the *Pfmsp1₁₉* fragment from blood samples of 33 children (19 males and 14
174 females) with infections spread between 10 and 24 multiple episodes totaling to 426 infections. On
175 recruitment, the average age of the participants was (5.5 ± SD 1.8 years), and at the end of study was
176 (10.30 ± SD 1.8 years). A total of 64.8% (276/426) of the samples yielded *Pfmsp1₁₉* amplicons, of these, a
177 total of 65.2% (180/276) were sequenced and only 54.3% (150/276) generated the full length (234bp)
178 *Pfmsp1₁₉* C-terminal contigs that were used for the microhaplotype analysis. The median parasitemia
179 was significantly higher among in the samples for which *Pfmsp1₁₉* amplicons were generated, 160,000
180 parasites/μl (interquartile range (IQR) = 245,740) compared to those where no *Pfmsp1₁₉* amplicons were
181 generated, 15,800 parasites/μl (IQR = 113,640) (P < 0.0001). Throughout the entire period, we genotyped
182 an average of 5 samples (± SD 2.678) per individual, which yielded a total of 150 sequences. We identified
183 six distinct nucleotide polymorphisms at positions 4990, 5132, 5157, 5159, 5161, and 5206 in relation to
184 *Pfmsp1* reference gene coordinates. The nucleotide substitutions in the polymorphic sites resulted in non-

185 synonymous amino acid substitutions at codons 1644, 1691, 1699, 1700, 1701 and 1716 yielding a total
186 of 11 microhaplotypes (**Figure 1A**), all of which have been reported in previous studies (Roy et al. 2008;
187 Mwingira et al. 2011). Microhaplotypes E-KSNG-L, Q-KSNG-L, Q-KSNG-F and Q-KNNG-L corresponding to
188 FUP-Uganda PA, FVO Wellcome, Thai (T807) and Kenya strains respectively, were the dominant
189 microhaplotypes circulating in the population with proportion frequencies of 36% (54/150), 26% (39/150),
190 18% (27/150) and 4.7% (7/150), respectively and these microhaplotypes were classified as prevalent. Both
191 E-KSSR-L and E-KNNG-L haplotypes were circulating with frequencies of 4% (6/150), and together with the
192 remaining microhaplotypes, E-KSSG-L, E-TSSR-L (3D7), Q-TSSR-L, E-TSSG-L and E-KSNG-F with low
193 frequencies of <5%, were considered as rare circulating microhaplotypes in the population (**Figure 1B**).

194 At least 7 of the 33 patients exhibited infections with several different microhaplotypes ranging
195 from 7 to 10 over the entire infection period (**Figure 2A, Supplementary Figure 1A**). All patients were
196 infected with at least one or more of the predominant microhaplotypes. Specifically, 72.7% (24/33) were
197 infected with the E-KSNG-L predominant microhaplotype. This microhaplotype ranged from 22.2%-75%
198 per patient in relation to all genotyped re-infections per individual (**Supplementary Figure 1B**).

199 The microhaplotype hamming distance matrix classified the microhaplotypes into 3 groups based
200 on Pearson correlation measures (**Figure 1C**). The first larger group was composed of E-KSNG-L, E-KNNG-
201 L, Q-KSNG-L, Q-KSNG-F, E-KSNG-F and Q-KNNG-L and these were the prevalent microhaplotypes. The
202 second group was made up of a single microhaplotype, E-KSSG-L, and the third group included E-KSSR-L,
203 E-TSSG-L, E-TSST-L and Q-TSSR-L microhaplotypes which were the rare circulating microhaplotypes (**Figure**
204 **1C**).

205 Microhaplotype dynamics over time and across infections

206 We followed the pattern of the microhaplotypes across infections to investigate the allelic
207 replacements of the C-terminal of *Pfmsp1₁₉* over the course of multiple infections. At an individual level,
208 there were notably high random allelic replacements between re-infections. At least 39.4% (13/33) of the
209 patients were consecutively re-infected with the same microhaplotype of the prevalent alleles either 2 to
210 3 times across the infection (**Figure 2A**). With the exception of one case (1/14) the consecutive re-
211 infections occurred within a one-year timeframe (**Figure 2A, 3B**). Remarkably, all individuals experiencing
212 consecutive re-infections with the same microhaplotype apart from one (1/13), showed no recurrence of
213 those specific microhaplotypes in subsequent parasite infections (**Figure 2A**). The average interval
214 between infections for the entire period was 5.0 months, ranging from 1 and half weeks to around 25
215 months (**Figure 3A**).

216 At a population level there was a peak in microhaplotype density before the 10th infection
217 episode, with the rare alleles peaking slightly later just after episode 10 (**Figure 2B**). We checked the
218 pattern of the four most frequent microhaplotypes and rare microhaplotypes across the multiple
219 infections. Of these, microhaplotypes E-KSNG-L was observed throughout the infections, Q-KSNG-L, Q-
220 KNGG-L and Q-KSNG-F peaked once before episode 10. The rare microhaplotypes peaked after episode 8
221 (**Figure 2B, C**). We noted that the occurrences of the prevalent microhaplotypes was not influenced by
222 chance as the observed frequencies matched the calculated expected frequencies (**Supplementary Figure**
223 **2**).

224 Correlation of parasitemia, genotyping and microhaplotypes

225 The change in parasitemia levels was examined over time and correlated with microhaplotype
226 and nucleotide diversities across the infection episodes. During the early and middle stage infection
227 episodes (<8 episodes) parasitemia levels were notably high. However, as the infections progressed from
228 towards the later (>8) episodes, parasitemia levels exhibited a decreasing trend. Genetic diversity of the
229 locus remained stable over the course of infection episodes (**Figure 4**). These trends were found to be
230 positively correlated with parasitaemia, with a strong positive correlation coefficient of 0.7 for nucleotide
231 diversity and but not for microhaplotype diversity, 0.37.

232

233 Discussion

234 Given the dynamic interplay between host immune responses which continuously shape
235 merozoite antigen diversity (Ferreira Marcelo U. et al. 2004; Early et al. 2018; Ochola-Oyier et al. 2019;
236 Naung et al. 2022) by shifting allele frequencies and favoring the presence of rare alleles, we leveraged
237 the diversity of *Pfmsp1* in parasite isolates from children with multiple malaria infections (a proxy of
238 developing immunity) as a window into quantifying the perturbation on merozoite antigen allele diversity.
239 Our results revealed preserved genetic diversity in *Pfmsp1*, even as parasitaemia levels decreased over
240 multiple malaria infections in an individual. This suggests that long-term immune pressure does not alter
241 the overall allele diversity. Similar patterns especially for the *msp1-19* microhaplotypes have been
242 observed across Sub-Saharan Africa including in the Coast of Kenya, Western Kenya, Republic of Congo,
243 Uganda, Tanzania, Mali and Burkina Faso (Mwingira et al. 2011; Kang et al. 2012; Kariuki et al. 2013;
244 Ochola-Oyier et al. 2019; Baina et al. 2023) suggesting that in moderate to high transmission, *P. falciparum*
245 populations maintain a complex infection pattern that supports out-breeding while preserving survival
246 genetic diversity.

247 It is expected that following multiple exposure to a single allele, immunity develops and reduces
248 its frequency in subsequent infections. However, repeat infections with the prevalent alleles circulating
249 in the population was common, with over a third of the children showing consecutive infections with the
250 same allele and only 6 children not showing a repeat infection with the same allele. The limitation in these
251 6 children is the low number of genotyped samples. Importantly, these findings highlight the complexity
252 of the parasite's genetic diversity that needs to be determined in the light of other polymorphic antigens.
253 Though the *Pfmsp1* alleles are the same in repeat infections, this region of the antigen is limited in
254 diversity to demonstrate distinct allelic changes. Other genetically diverse antigens may show differences
255 between each infection such as *ama1* or block 2 of *msp1* and block 3 of *msp2*. What was unique in our
256 data was the reduction in parasitemia in later infections, which did not lead to a reduction in haplotype
257 diversity. Though it is clear that following several malaria infections, the host can control parasitemia due
258 to the development of immunity (usually antibody-mediated) (Pohl and Cockburn 2022) with age; there
259 appears to be a lapse in the allele-specific immunity memory, such that over time antibody levels targeted
260 to a particular allele are low or wane over a short period (about <1 year) when re-infections occur with
261 the same allele. It is surprising that with continued exposure and increasing age re-infections with the
262 same alleles occur, while it is expected that exposure to several malaria infections should increase the
263 antibody repertoire to the several antigens and polymorphic combinations presented over time. However,
264 what is evident is the difference between anti-parasite immunity that develops to control parasitemia
265 levels following several infections, and increasing age, in an individual and the immunity that develops to
266 clear allelically different epitopes of the same antigens. The control of parasitemia following several
267 malaria infections is similar to previous findings in Uganda that observed lower parasite densities with
268 increasing age and in high malaria transmission areas (Rodriguez-Barraquer et al. 2018). Thus,
269 emphasizing the difference in the immunity that controls parasitemia and that which could lead to sterile
270 immunity. This latter process is not achieved for malaria, based on this data where repeated exposure by
271 the same prevalent allele occurs that is not completely cleared, potentially due to an ineffective immune
272 response that is not protective, akin to the original antigenic sin hypothesis (Good et al. 1993). There may
273 be some cross-reactivity in responses between alleles, however this too is likely to be limiting since re-
274 infections occur with the prevalent alleles circulating in the population. The re-infections with the same
275 allele allows the maintenance of their high prevalence in the population and thus the genetic diversity of
276 the infections is unaltered. The high haplotype diversity is sustained while the nucleotide diversity in
277 contrast dropped with the parasitemia levels. The nucleotide diversity is likely to reduce as re-infections
278 occur with the same allele, the average nucleotide differences between sequences across the population

279 will thus reduce. However, it will not alter the overall haplotype diversity the probability that two
280 randomly sampled alleles are different in the population. Furthermore, this population is unique in the
281 sense that these are children who are uncharacteristically infected several times over 5 years with malaria
282 and who have been shown to have a modified immune system high immune activation and inflammation,
283 TNF- α , IL-6, IL-10 and cell populations such as $\gamma\delta$ T cells were significantly higher in children with >8
284 malaria infections compared to the <5 infections group (Bediako et al. 2019). This skewed cytokine profile
285 may act in a way that the inflammatory immune response to some extent clears parasites controlling
286 parasitemia, however the IL-10 modulates the inflammatory response limiting inflammation and related
287 immunopathology, but it may interfere with robust protective allele-specific immune responses.

288 Of additional interest, was the time between infections that was on average 5 months for the
289 recurrent allelic infections, suggesting that immunity then wanes after 5 months and reinfections then
290 occur in this group of children who experience multiple malaria infections. This is likely to be a primer for
291 the design of multi-allelic blood stage vaccine combinations, as potentially giving an individual a cocktail
292 of alleles or alternating between alleles in follow up vaccine doses five or six months apart.

293 It is worth noting that our observations would have been more robust if complemented with other
294 invasion genes with higher levels of polymorphisms to support interpretability of our findings. In addition,
295 without immunological functional validation experiments, uncertainty persists regarding whether the
296 absence of consecutive same variants in late infections among certain individuals, and distribution
297 densities in later infections are a result of allelic-specific or antigenically dissimilar variant-transcending
298 immune responses. Further experiments in this direction are needed to clarify this observation.

299

300 **Conclusion**

301 Parasite *Pfmsp1₁₉* allelic diversity remains stable over the multiple malaria episodes despite
302 declining parasitaemia levels over time due to population-wide immune pressure. This suggests that
303 immune responses targeting *Pfmsp1* may not be the primary factor driving the observed reduction in
304 parasitaemia, underscoring the multifactorial nature of protective immune responses against malaria. In
305 addition, our findings reveal a high *Pfmsp1₁₉* microhaplotype replacement across multiple infections.
306 While shifts in alleles between infections indicate initial immune responses to variants, these responses
307 are not consistently maintained over time, allowing for re-infections with the same variant. In conclusion,
308 the study highlights that immunity to malaria is not achieved through complete sterilizing immunity but
309 rather involves a complex interplay of immune responses targeting multiple alleles from multiple
310 antigens.

311

312 **Data and code availability**

313 Data accessibility

314 The DNA sequences for the C-terminal of *Pfmsp1₁₉* were deposited in GenBank NIH genetic sequence
315 database and are available under the accession numbers: (OQ821998 - OQ822147) (**Supplementary Table**
316 **1**).

317 Code

318 Analysis and code for this study is available in the dedicated GitHub repository:

319 <https://github.com/mangiruben/pfMSP1-Malaria>

320

321 **Declarations**

322 Competing interests

323 The authors declare no competing interests.

324 Funding

325 The work was supported by the malaria capacity development consortium (MCDC) re-entry grant and a
326 Wellcome Trust Intermediate Fellowship (107568/Z/15/Z) to LIO-O, who is currently supported by a
327 Calestous Juma fellowship from BMGF INV036442. RMY was supported by a University of Nairobi (UoN)
328 scholarship administered through the Centre for Biotechnology and Bioinformatics (CEBIB). We extend
329 our gratitude to UoN and CEBIB.

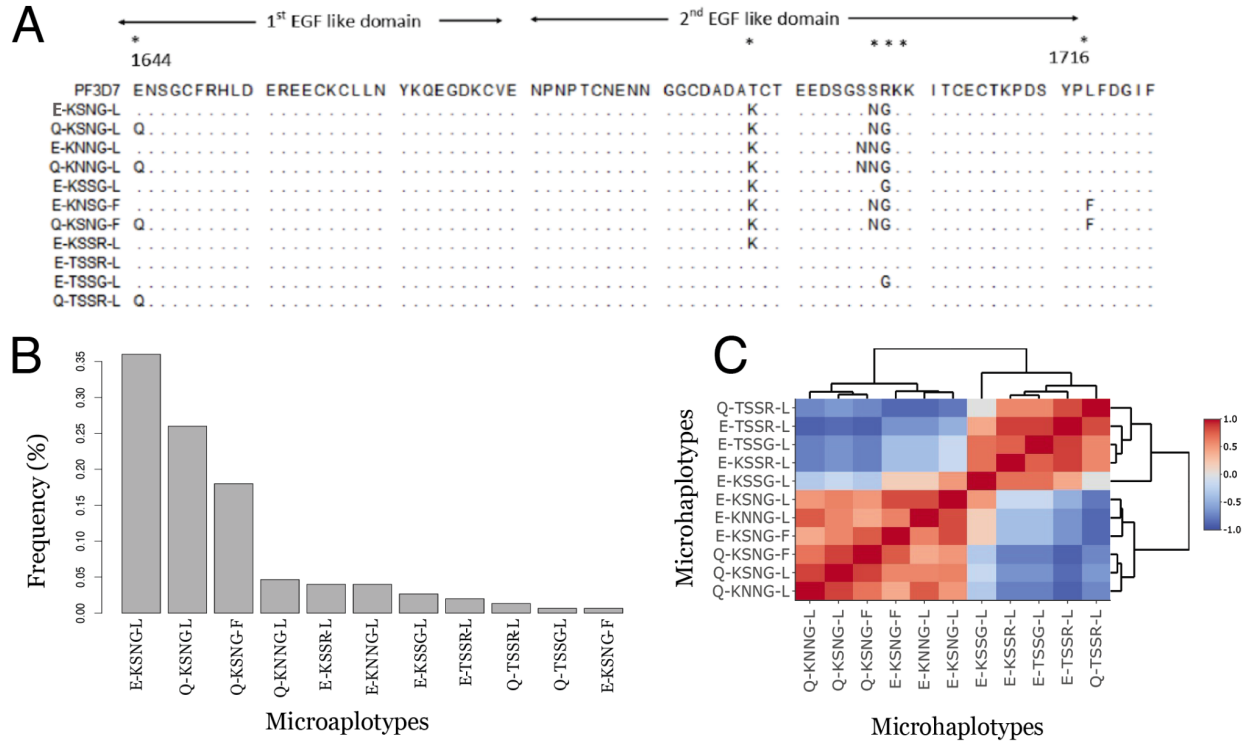
330 Author contributions

331 RMY performed all experiments, performed PCR and library preparation, data analysis and wrote the
332 manuscript. KMK did analysis and wrote the manuscript. PHA did analysis and wrote the manuscript. GO
333 oversaw the study and contributed to the manuscript. LIO-O conceived, designed, and oversaw the study
334 and wrote the manuscript. All authors read and reviewed the manuscript.

335 **Acknowledgements**

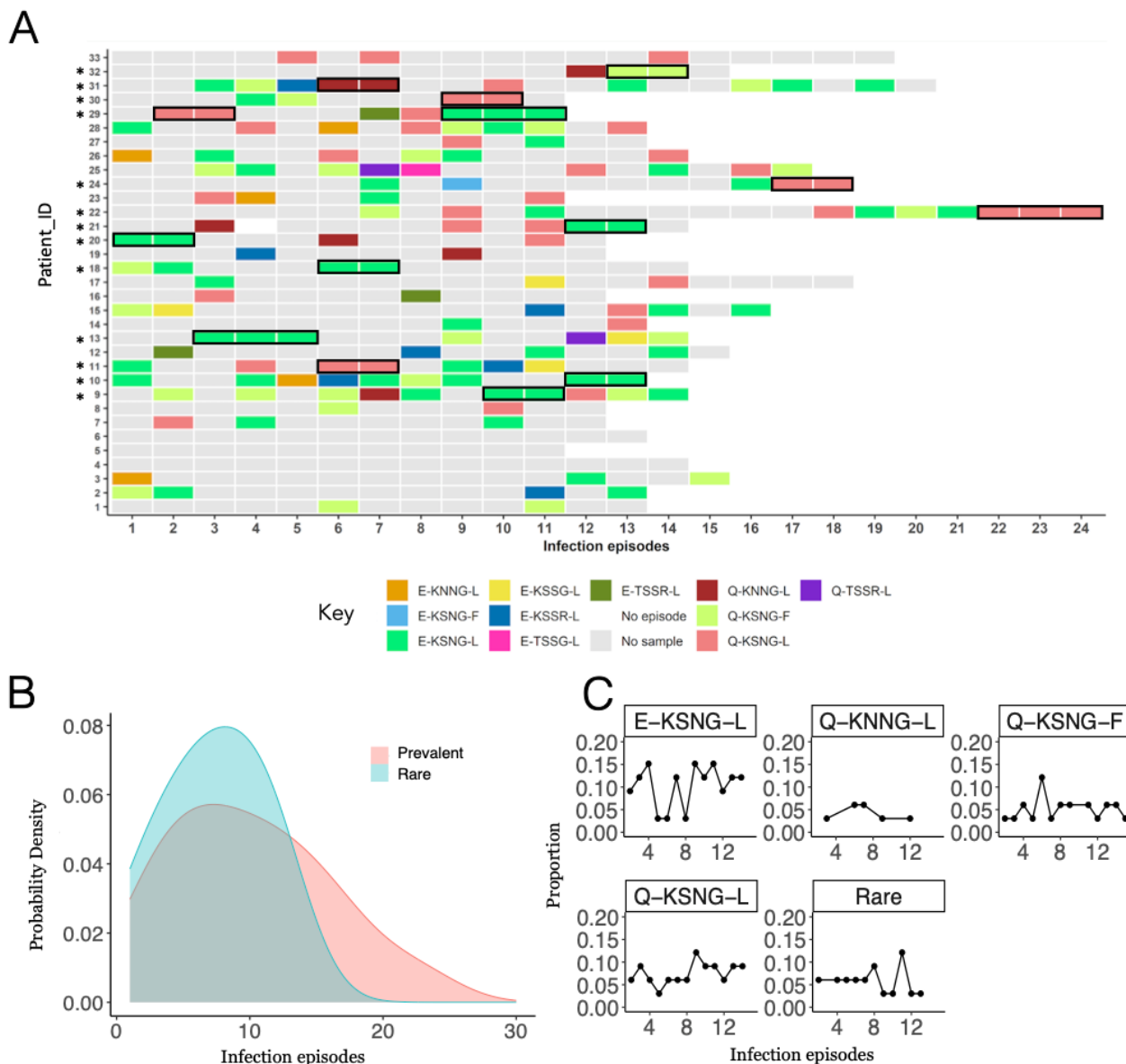
336 We thank Ann Owiti and Edwin Rono at Centre for Biotechnology and Bioinformatics (CEBIB), University
337 of Nairobi for overseeing PCR, library generation and sequencing. The Biosciences department at KEMRI-
338 Wellcome Trust Research Programme, Kilfi for the DNA samples from Junju cohort. We thank the previous
339 Directors of CEBIB and the Director of the Kenya Medical Research Institute for permission to publish this
340 article.

341 **Figures**



342

343 **Figure 1.** Dynamics of *P. falciparum* *msp19* microhaplotypes. **(A)** Amino- acid sequence alignment of 11
 344 identified microhaplotypes. Polymorphic sites are shown with an asterisk (*). The nucleotide positions
 345 relative to the start position of the *Pfmsp1* gene are shown below the asterisk. The dots in the alignment
 346 indicate the position corresponding to *P. falciparum* 3D7 with identical amino acid sequences. The
 347 epidermal growth factor (EGF)-like domains 1 and 2 are shown by arrows. The first polymorphism is
 348 located in the first EGF-like domain, whereas the second to the fifth polymorphism are located in the
 349 second EGF-like domain. **(B)** Microhaplotypes sorted by their abundance in the population. **(C)**
 350 Microhaplotypes clustered to groups based on the number of nucleotide differences between haplotypes.
 351 The dendrogram on the sides of the heatmap visually represents the relatedness of the microhaplotypes.
 352 In this context, the branches indicate distinct clusters formed through hierarchical clustering, highlighting
 353 groups of haplotypes with similar characteristics. Scale=Pearson correlation coefficients.

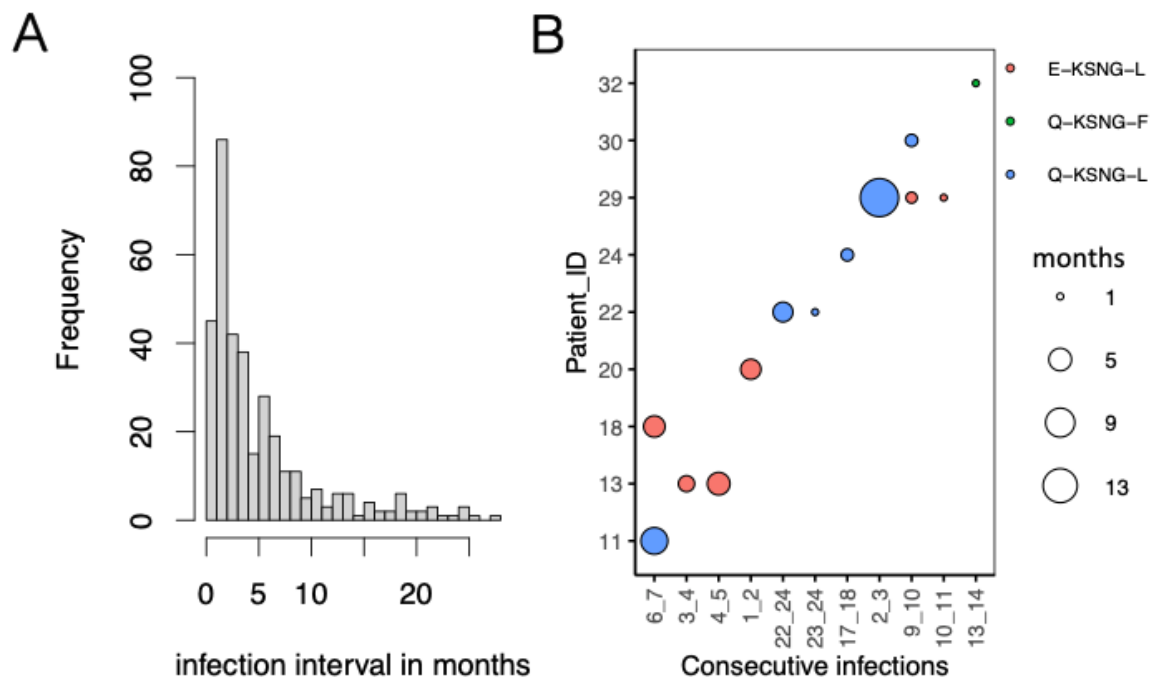


354

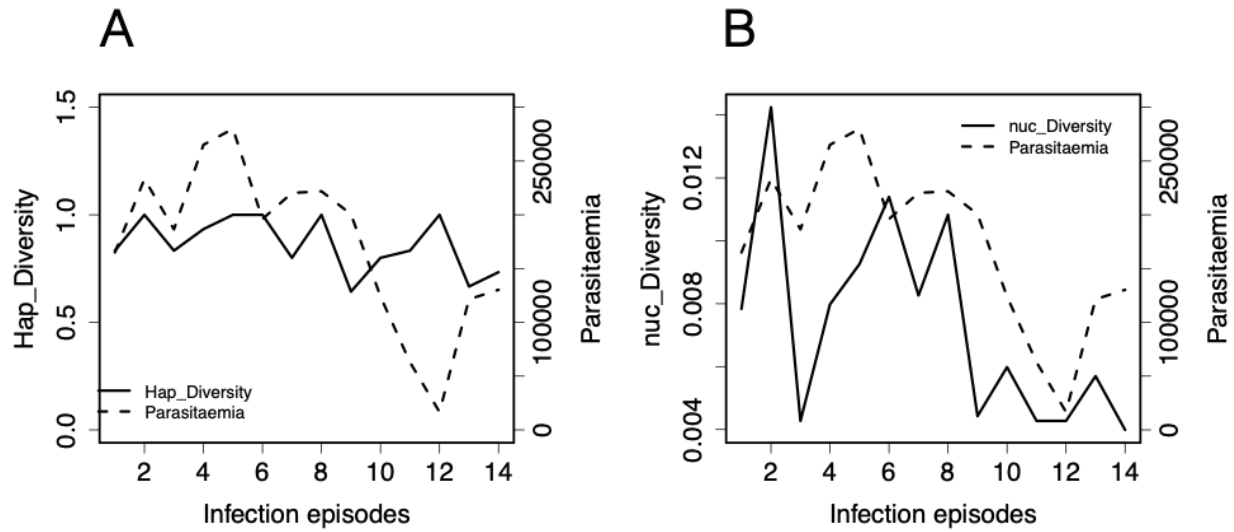
355

356 **Figure 2.** Microhaplotype patterns across the infections(A). The distribution of microhaplotypes in each
 357 patient across the malaria episodes. No sample - Samples not retrieved from the biobank, samples not
 358 genotyped or sequenced. No episode - No malaria episode was reported. Patient_IDs with asterisks
 359 represent cases that had 2-3 similar consecutive microhaplotypes outlined in black. (B) Density of
 360 prevalent ($\geq 5\%$ frequency) and rare ($<5\%$ frequency) microhaplotypes across the infection episodes based
 361 on **Figure 1B**. (C) Line graph of proportions of prevalent microhaplotypes (E-KSNG-L, Q-KSNG-L and Q-
 362 KSNG-F), and rare microhaplotypes across the first 15 infection episodes. Proportions are expressed from
 363 the total children genotyped at each infection episode.

364



365
366
367
368 **Figure 3:** Distribution of infection intervals. **(A)** Frequency histogram of the distribution of the time interval
369 between infections in months for all infections with genotype data. **(B)** Interval in months for individuals
370 with consecutive infections of the same microhaplotype. The size of the circles depicts the number of
371 months between infections with the same microhaplotype depicted by the color of the circle.



372

373 **Figure 4.** Parasitaemia correlations with measure of genetic diversity across infection episodes. **(A)**
374 Haplotype diversity (hap_diversity) fluctuates within a small range (between 0.6 and 1) across infection
375 episodes as parasitaemia (parasites/ μ l) reduces, the correlation between haplotype diversity and
376 parasitemia was low, 0.37. **(B)** Nucleotide diversity (nuc_diversity) reduces concurrently across infection
377 episodes with parasitaemia with a correlation of 0.7.

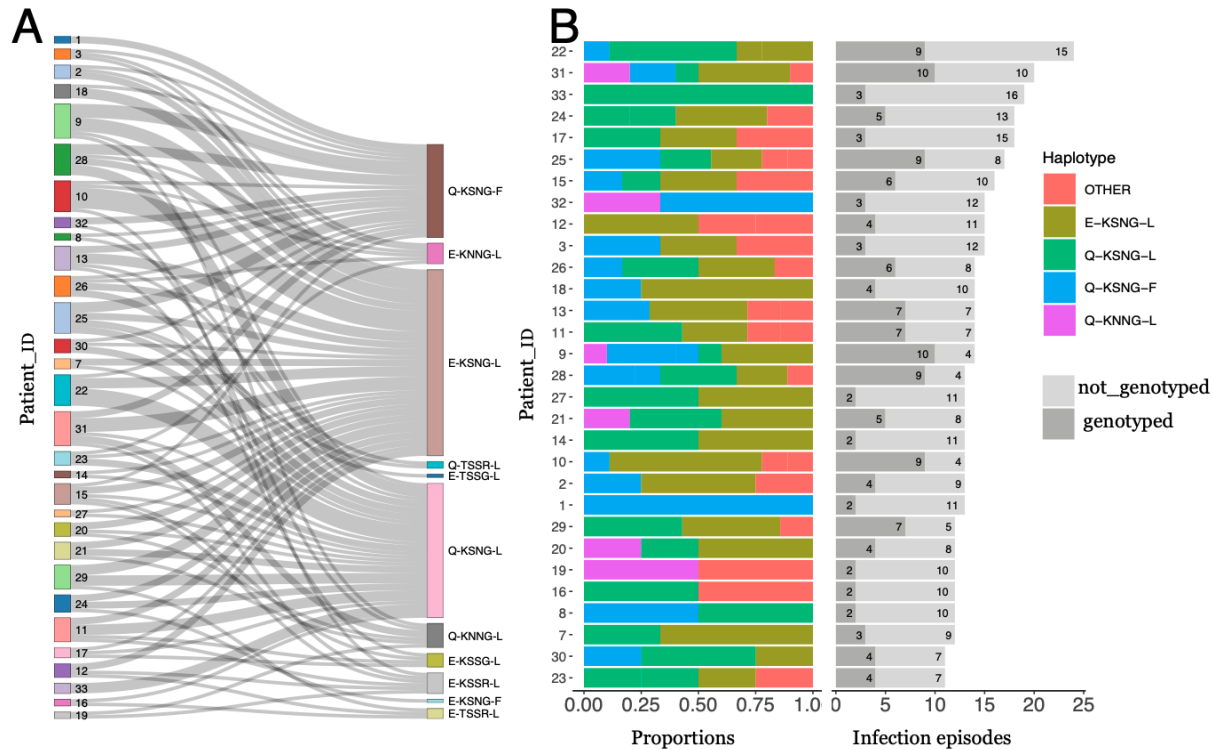
378 **Supplementary Materials**

379 **Supplementary Tables**

380 Table S1. Summary of study participants and blood samples

381 [Supplementary Table 1](#)(excel sheet)

382 **Supplementary Figures**

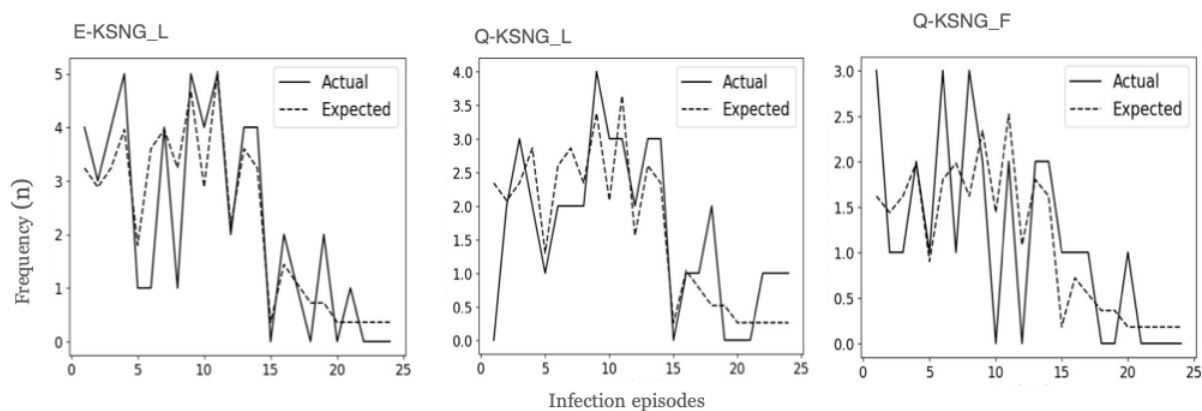


383

384 **Supplementary Figure 1:** Within host haplotype dynamics. **(A)** Haplotypes assigned back to the patient.
 385 Sankey plot showing the assignment of microhaplotypes each patient (Patient_ID) encountered across
 386 the multiple episodes. **(B)** The proportion of prevalent haplotypes (E-KSNG-L, Q-KSNG-L, Q-KSNG-F and Q-
 387 KNNG-L) in relation to all the rare variant microhaplotypes (OTHER) per patient (Patient_ID). The grey bar
 388 panel on the right shows the number of episodes of genotyped and not genotyped per patient.

389

390



391

392 **Supplementary Figure 2:** Difference between actual and expected microhaplotype frequencies across
393 infections. The expected number of occurrences was derived using probabilities computed from
394 maximum likelihood estimates.

395 REFERENCES

- 396 Awandare GA, Nyarko PB, Aniwah Y, Ayivor-Djanie R, Stoute JA. Plasmodium falciparum strains
397 spontaneously switch invasion phenotype in suspension culture. *Sci Rep*. 2018 Apr 10;8(1):5782.
- 398 Baina MT, Lissom A, Assioro Doulamo NV, Djontu JC, Umuhzoa DM, Mbama-Ntabi JD, et al. Comparative
399 study of Plasmodium falciparum msp-1 and msp-2 Genetic Diversity in Isolates from Rural and Urban
400 Areas in the South of Brazzaville, Republic of Congo. *Pathogens* [Internet]. 2023 May 22;12(5).
401 Available from: <http://dx.doi.org/10.3390/pathogens12050742>
- 402 Bediako Y, Adams R, Reid AJ, Valletta JJ, Ndungu FM, Sodenkamp J, et al. Repeated clinical malaria
403 episodes are associated with modification of the immune system in children. *BMC Med*. 2019 Mar
404 13;17(1):60.
- 405 Chauhan VS, Yazdani SS, Gaur D. Malaria vaccine development based on merozoite surface proteins of
406 Plasmodium falciparum. *Hum Vaccin* [Internet]. 2010 Sep 13;6(9). Available from:
407 <http://dx.doi.org/10.4161/hv.6.9.12468>
- 408 Conway DJ, Greenwood BM, McBride JS. Longitudinal study of Plasmodium falciparum polymorphic
409 antigens in a malaria-endemic population. *Infect Immun*. 1992 Mar;60(3):1122–7.
- 410 Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. *Cell*. 2006 Feb 24;124(4):755–66.
- 411 Das S, Hertrich N, Perrin AJ, Withers-Martinez C, Collins CR, Jones ML, et al. Processing of Plasmodium
412 falciparum Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite
413 Egress from RBCs. *Cell Host Microbe*. 2015 Oct 14;18(4):433–44.
- 414 Early AM, Lievens M, MacInnis BL, Ockenhouse CF, Volkman SK, Adjei S, et al. Host-mediated selection
415 impacts the diversity of Plasmodium falciparum antigens within infections. *Nat Commun*. 2018 Apr
416 11;9(1):1381.
- 417 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids
418 Res*. 2004 Mar 19;32(5):1792–7.
- 419 Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010 Oct
420 1;26(19):2460–1.
- 421 Egan AF, Burghaus P, Druilhe P, Holder AA, Riley EM. Human antibodies to the 19kDa C-terminal fragment
422 of Plasmodium falciparum merozoite surface protein 1 inhibit parasite growth in vitro. *Parasite
423 Immunol*. 1999 Mar;21(3):133–9.
- 424 Egan AF, Chappel JA, Burghaus PA, Morris JS, McBride JS, Holder AA, et al. Serum antibodies from malaria-
425 exposed people recognize conserved epitopes formed by the two epidermal growth factor motifs of
426 MSP1(19), the carboxy-terminal fragment of the major merozoite surface protein of Plasmodium
427 falciparum. *Infect Immun*. 1995 Feb;63(2):456–66.
- 428 Ferreira Marcelo U., da Silva Nunes Mônica, Wunderlich Gerhard. Antigenic Diversity and Immune Evasion
429 by Malaria Parasites. *Clin Vaccine Immunol*. 2004 Nov 1;11(6):987–95.
- 430 Good MF, Zevering Y, Currier J, Bilsborough J. “Original antigenic sin”, T cell memory, and malaria
431 sporozoite immunity: an hypothesis for immune evasion. *Parasite Immunol*. 1993 Apr;15(4):187–93.
- 432 Herrera R, Anderson C, Kumar K, Molina-Cruz A, Nguyen V, Burkhardt M, et al. Reversible Conformational
433 Change in the Plasmodium falciparum Circumsporozoite Protein Masks Its Adhesion Domains. *Infect
434 Immun*. 2015 Oct;83(10):3771–80.
- 435 Holm S. Sequentially rejective multiple test procedures. 1977.
- 436 John CC, O’Donnell RA, Sumba PO. Evidence that invasion-inhibitory antibodies specific for the 19-kDa
437 fragment of merozoite surface protein-1 (MSP-119) can play a protective role against blood-stage
438 The Journal of [Internet]. 2004a;. Available from:
439 <https://www.jimmunol.org/content/173/1/666.short>
- 440 John CC, O’Donnell RA, Sumba PO, Moormann AM, de Koning-Ward TF, King CL, et al. Evidence that

441 invasion-inhibitory antibodies specific for the 19-kDa fragment of merozoite surface protein-1 (MSP-
442 119) can play a protective role against blood-stage Plasmodium falciparum infection in individuals in
443 a malaria endemic area of Africa. *The Journal of Immunology*. 2004b;173(1):666–72.

444 Kang JM, Ju HL, Kang YM, Lee DH, Moon SU, Sohn WM, et al. Genetic polymorphism and natural selection
445 in the C-terminal 42 kDa region of merozoite surface protein-1 among Plasmodium vivax Korean
446 isolates. *Malar J*. 2012 Jun 18;11:206.

447 Kariuki SK, Njunge J, Muia A, Muluvi G, Gatei W, ter Kuile F, et al. Effect of malaria transmission reduction
448 by insecticide-treated bed nets (ITNs) on the genetic diversity of Plasmodium falciparum merozoite
449 surface protein (MSP-1) and circumsporozoite (CSP) in western Kenya. *Malar J*. 2013 Dec;12(1):295.

450 Langhorne J, Ndungu FM, Sponaas AM, Marsh K. Immunity to malaria: more questions than answers. *Nat*
451 *Immunol*. 2008 Jul;9(7):725–32.

452 Larsen MD, Quintana MDP, Ditlev SB, Bayarri-Olmos R, Ofori MF, Hviid L, et al. Evasion of Classical
453 Complement Pathway Activation on Plasmodium falciparum-Infected Erythrocytes Opsonized by
454 PfEMP1-Specific IgG. *Front Immunol*. 2018;9:3088.

455 Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, et al. Indels, structural variation, and
456 recombination drive genomic diversity in Plasmodium falciparum. *Genome Res*. 2016
457 Sep;26(9):1288–99.

458 Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GAT, et al. Recombination hotspots and population
459 structure in Plasmodium falciparum. *PLoS Biol*. 2005 Oct;3(10):e335.

460 Mwangi TW, Ross A, Snow RW, Marsh K. Case definitions of clinical malaria under different transmission
461 conditions in Kilifi District, Kenya. *J Infect Dis*. 2005 Jun 1;191(11):1932–9.

462 Mwingira F, Nkwengulila G, Schoepflin S, Sumari D, Beck HP, Snounou G, et al. Plasmodium falciparum
463 msp1, msp2 and glurp allele frequency and diversity in sub-Saharan Africa. *Malar J*. 2011 Apr 6;10:79.

464 Naung MT, Martin E, Munro J, Mehra S, Guy AJ, Laman M, et al. Global diversity and balancing selection
465 of 23 leading Plasmodium falciparum candidate vaccine antigens. *PLoS Comput Biol*. 2022
466 Feb;18(2):e1009801.

467 Ochola LI, Tetteh KKA, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and
468 polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic
469 genes in Plasmodium falciparum. *Mol Biol Evol*. 2010 Oct;27(10):2344–51.

470 Ochola-Oyier LI, Okombo J, Wagatua N, Ochieng J, Tetteh KK, Fegan G, et al. Comparison of allele
471 frequencies of Plasmodium falciparum merozoite antigens in malaria infections sampled in different
472 years in a Kenyan population. *Malar J*. 2016 May 6;15(1):261.

473 Ochola-Oyier LI, Wamae K, Omedo I, Ogola C, Matharu A, Musabyimana JP, et al. Few Plasmodium
474 falciparum merozoite ligand and erythrocyte receptor pairs show evidence of balancing selection.
475 *Infect Genet Evol*. 2019 Apr;69:235–45.

476 Ogutu BR, Apollo OJ, McKinney D, Okoth W, Siangla J, Dubovsky F, et al. Blood stage malaria vaccine
477 eliciting high antigen-specific antibody concentrations confers no protection to young children in
478 Western Kenya. *PLoS One*. 2009 Mar 5;4(3):e4708.

479 Ouattara A, Takala-Harrison S, Thera MA, Coulibaly D, Niangaly A, Saye R, et al. Molecular basis of allele-
480 specific efficacy of a blood-stage malaria vaccine: vaccine development implications. *J Infect Dis*.
481 2013 Feb 1;207(3):511–9.

482 Oyong DA, Kenangalem E, Poespoprodjo JR, Beeson JG, Anstey NM, Price RN, et al. Loss of complement
483 regulatory proteins on uninfected erythrocytes in vivax and falciparum malaria anemia. *JCI Insight*
484 [Internet]. 2018 Nov 15;3(22). Available from: <http://dx.doi.org/10.1172/jci.insight.124854>

485 Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language.
486 *Bioinformatics*. 2004 Jan 22;20(2):289–90.

487 Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of selection and
488 geospatial genetic diversity within two leading Plasmodium vivax candidate vaccine antigens. *PLoS*

- 489 Negl Trop Dis. 2014 Apr;8(4):e2796.
- 490 Pinheiro HP, de Souza Pinheiro A, Sen PK. Comparison of genomic sequences using the Hamming distance.
491 J Stat Plan Inference. 2005 Mar 1;130(1):325–39.
- 492 Pohl K, Cockburn IA. Innate immunity to malaria: The good, the bad and the unknown. Front Immunol.
493 2022 Aug 19;13:914598.
- 494 Polley SD, Conway DJ. Strong diversifying selection on domains of the Plasmodium falciparum apical
495 membrane antigen 1 gene. Genetics. 2001 Aug;158(4):1505–12.
- 496 Polley SD, Tetteh KKA, Lloyd JM, Akpogheneta OJ, Greenwood BM, Bojang KA, et al. Plasmodium
497 falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are
498 maintained by natural selection. J Infect Dis. 2007 Jan 15;195(2):279–87.
- 499 Rayner JC, Tran TM, Corredor V, Huber CS, Barnwell JW, Galinski MR. Dramatic difference in diversity
500 between Plasmodium falciparum and Plasmodium vivax reticulocyte binding-like genes. Am J Trop
501 Med Hyg. 2005 Jun;72(6):666–74.
- 502 Read TRC, Cressie NAC. Goodness-of-Fit Statistics for Discrete Multivariate Data. Springer Science &
503 Business Media; 2012.
- 504 Reiling L, Richards JS, Fowkes FJI, Barry AE, Triglia T, Chokejindachai W, et al. Evidence that the erythrocyte
505 invasion ligand PfRh2 is a target of protective immunity against Plasmodium falciparum malaria. J
506 Immunol. 2010 Nov 15;185(10):6157–67.
- 507 Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Kanya MR, Rosenthal PJ, Rek J, et al. Quantification
508 of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. Elife
509 [Internet]. 2018 Jul 25;7. Available from: <http://dx.doi.org/10.7554/eLife.35832>
- 510 Rotman HL, Daly TM, Long CA. Plasmodium: immunization with carboxyl-terminal regions of MSP-1
511 protects against homologous but not heterologous blood-stage parasite challenge. Exp Parasitol.
512 1999 Jan;91(1):78–85.
- 513 Roy SW, Ferreira MU, Hartl DL. Evolution of allelic dimorphism in malarial surface antigens. Heredity .
514 2008 Feb;100(2):103–10.
- 515 Simpalian P, Pattaradilokrat S, Siripoon N, Seugorn A, Kaewthamasorn M, Butcher RDJ, et al. Diversity
516 and population structure of Plasmodium falciparum in Thailand based on the spatial and temporal
517 haplotype patterns of the C-terminal 19-kDa domain of merozoite surface protein-1. Malar J. 2014
518 Feb 12;13:54.
- 519 Takala SL, Coulibaly D, Thera MA, Dicko A, Smith DL, Guindo AB, et al. Dynamics of polymorphism in a
520 malaria vaccine antigen at a vaccine-testing site in Mali. PLoS Med. 2007a Mar;4(3):e93.
- 521 Takala SL, Smith DL, Thera MA, Coulibaly D, Doumbo OK, Plowe CV. Short report: rare Plasmodium
522 falciparum merozoite surface protein 1 19-kDa (msp-1(19)) haplotypes identified in Mali using high-
523 throughput genotyping methods. Am J Trop Med Hyg. 2007b May;76(5):855–9.
- 524 Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol
525 Evol. 2021 Jun 25;38(7):3022–7.
- 526 Tan MSY, Blackman MJ. Malaria parasite egress at a glance. J Cell Sci [Internet]. 2021 Mar 8;134(5).
527 Available from: <http://dx.doi.org/10.1242/jcs.257345>
- 528 Tetteh KKA, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, Marsh K, et al. Prospective
529 identification of malaria parasite genes under balancing selection. PLoS One. 2009 May
530 15;4(5):e5568.
- 531 Tzeng JY. Evolutionary-based grouping of haplotypes in association analysis. Genet Epidemiol. 2005
532 Apr;28(3):220–31.
- 533 Verra F, Chokejindachai W, Weedall GD, Polley SD, Mwangi TW, Marsh K, et al. Contrasting signatures of
534 selection on the Plasmodium falciparum erythrocyte binding antigen gene family. Mol Biochem
535 Parasitol. 2006 Oct;149(2):182–90.
- 536 Węglarczyk S. Kernel density estimation and its application. ITM Web Conf. 2018;23:00037.

537 Wilson DW, Fowkes FJI, Gilson PR, Elliott SR, Tavul L, Michon P, et al. Quantifying the importance of MSP1-
538 19 as a target of growth-inhibitory and protective antibodies against Plasmodium falciparum in
539 humans. PLoS One. 2011 Nov 15;6(11):e27705.