scientific reports

Check for updates

OPEN

Disparate co-evolution and prevalence of sulfadoxine and pyrimethamine resistance alleles and haplotypes at *dhfr* and *dhps* genes across Africa

Nina F. D. White[1], Georgia Whitton[1], Varanya Wasakul[2], Lucas Amenga-Etego[3,7], Antoine Dara[4,7], Voahangy Andrianaranjaka[5,7], Milijaona Randrianarivelojosia[5,7], Olivo Miotto[2,7], Umberto D'Alessandro[6], Abdoulaye Djimdé[4,7], Cristina V. Ariani[1], Richard D. Pearson[1✉] & Alfred Amambua-Ngwa[6,7✉]

Sulfadoxine-pyrimethamine (SP), despite emergence of mutations in *dhfr* and *dhps* genes associated with lower treatment efficacy, is still recommended for preventive malaria treatment. Therefore, it is important to understand the evolution of *P. falciparum dhfr* and *dhps* genes. We used the MalariaGEN Pf7 dataset to describe haplotype frequencies across 22 African countries, including changes over time in The Gambia, Mali, Ghana, and Kenya. We show that the triple mutant of *dhfr*, N51I/C59R/ S108N, has remained the dominant haplotype across the continent with limited evidence of additional mutations. There is greater variation for *dhps* in terms of haplotype diversity and spatial heterogeneity of haplotypes found across Africa. Although samples from Madagascar have low genetic differentiation from samples from mainland East Africa at the whole genome level, we show that *dhps* K540E is highly differentiated between the two populations, being at very low frequency in Madagascar (4%). Whole genome data reveal 12 SNPs which are also highly differentiated between Madagascar and East Africa, including *aat1* and a possible novel drug resistance locus approximately 20 kb 3′ of *mdr1*. We highlight the value of longitudinal sampling and whole genome sequence data for understanding the heterogeneity and ongoing changes in anti-malarial drug resistance genetic markers.

Keywords *Plasmodium falciparum*, Malaria, Genomic surveillance, Sulfadoxine-pyrimethamine (SP) resistance

As malaria continues to cause significant morbidity and mortality, especially in Africa, monitoring of current interventions and implementation of new ones are needed. Malaria particularly affects children and pregnant women, who contributed to most of the over 240 million infections and 600,000 deaths reported in 2022[1]. To prevent morbidity and mortality in these vulnerable groups, malaria-endemic countries in Africa are employing Sulfadoxine-Pyrimethamine (SP) in various methods. SP is used for intermittent preventive treatment during pregnancy (IPTp), for perennial malaria chemoprevention (PMC) in infants or, in combination with amodiaquine (AQ) for seasonal malaria chemoprevention (SMC) in children 3–59 months old. SP is deployed across 15 countries in the Sahel region of West Africa, and in Mozambique and Uganda in East Africa[2]. SP has been previously used as the first-line treatment for malaria. In addition, AQ is a partner drug in one of the currently available artemisinin-based combination therapies (ACTs) to treat patients with uncomplicated malaria. Although SP-AQ remains effective for chemoprevention, resistance to SP and AQ separately has long

[1]Genomic Surveillance Unit, Wellcome Sanger Institute, Hinxton, UK. [2]Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. [3]West African Center for Cell Biology of Infectious Pathogens, University of Ghana, Accra, Ghana. [4]Malaria Research and Training Centre, Faculty of Pharmacy, Université des Sciences, des Techniques et des Technologies de Bamako, Bamako, Mali. [5]Institut Pasteur de Madagascar, Antananarivo, Madagascar. [6]Medical Research Council Unit, The Gambia at London School of Hygiene and Tropical Medicine, Banjul, The Gambia. [7]Pathogens Genomic Diversity Network Africa (PDNA), Sotuba, Bamako, Mali. ✉email: rp7@sanger.ac.uk; Alfred.Ngwa@lshtm.ac.uk

been established, with SP discontinued as a first-line therapy since the early 2000s in most endemic countries[3]. Hence the continuous use of SP in IPTp, PMC and SMC may further select resistant *P. falciparum* strains that could render the drugs even less effective in the future.

Resistance to SP has been attributed to mutations in *P. falciparum* dihydrofolate reductase (*dhfr*) and dihydropteroate synthetase (*dhps*) genes for pyrimethamine and sulfadoxine components, respectively[4]. In African malaria parasite populations, resistance-associated haplotypes in these genes were imported from Asia and spread by transmission[5,6]. For *dhfr*, the following mutations; N51I, C59R, S108N, and I164L form the resistant haplotype, with I, R, N mutations at 51, 59, and 108 reported at high frequencies across the African continent[7]. Sulfadoxine resistance-associated haplotypes of *dhps* at positions S436A/F, A437G, K540E, A581G, and A613S/T are more complex, with heterogeneous combinations and distribution across the African continent[7,8]. SP resistance results from different combinations of mutations across both the *dhfr* and *dhps* genes. For example, the quintuple mutant, a combination of *dhfr* N51I/C59R/S108N and *dhps* A437G/K540E, confers a larger risk of treatment failure than the individual *dhfr* triple or *dhps* double mutants[9]. Beyond these, *dhfr*-I164L and *dhps*-A581G mutations have been associated with a higher risk of treatment failure in East Africa[10,11]. Alongside *dhps*-A613S/T, these two mutations in combination with the quintuple mutant are said to confer 'super-resistance' to SP[12]. The prevalence of both rare and common haplotype combinations, and their effect on chemoprevention is relevant for strategizing drug intervention policies toward control and elimination. As haplotype prevalence data do not directly inform on patient outcomes, they should be complemented by in vitro assays or therapeutic efficacy studies to validate haplotype associations with antimalarial resistance.

Malaria Molecular Surveillance (MMS) of malaria parasites for markers of resistance and diversity can thus complement current epidemiological approaches against malaria and inform the optimal deployment of interventions. Several studies have employed microsatellite and SNP typing approaches to map the prevalence, origins, and dispersal of SP-resistance markers across Africa and the globe[6,10,11,13]. In the last decade, with advances in genome sequencing of *P. falciparum*, the Pathogen (*Plasmodium*) Diversity Network Africa (PDNA) consortium in collaboration with MalariaGEN embarked on genomic sequencing of malaria parasites across 20 African countries. This report presents the prevalence of genetic markers of SP resistance across both PDNA sites and other African countries. It shows the different rates and directions of selection of these alleles over 30 years in four African countries, especially for the *dhps* markers of sulfadoxine resistance.

Results

Whole-genome sequences from the MalariaGEN Pf7 dataset were used in this study[14]. Pf7 contains 8,492 QC-passed *Plasmodium falciparum* samples from 21 African countries which were selected for analysis. Amino acid haplotypes for *dhfr* and *dhps* were determined for each sample by applying all mutations discovered in the sample to the 3D7 reference sequence, then translating the nucleotide sequence of concatenated exons. Only homozygous haplotypes were analysed (see Methods). Frequencies of individual mutations over time were generated using the allele frequency method, which considers only homozygous loci in the calculation (see Methods).

Variation in *dhfr*

The most frequent haplotype of *dhfr* was the triple mutant (N51I/C59R/S108N), found in 80% of samples (Fig. 1, Supplementary Table S1). The second most frequent haplotype, found in 10% of samples, was the 3D7 reference genome sequence - which is thought to be the wild type for *dhfr*. The third and fourth most common haplotypes were C59R/S108N and N51I/S108N, respectively. These two haplotypes had distinct geographical spreads, C59R/S108N was more common in West and East Africa, while N51I/S108N was more common in Central and North-East Africa (Fig. 1). The other potential combination of the N51I, C59R, and S108N mutations, N51I/C59R, was not observed. All three mutations N51I, C59R, and S108N were seen in isolation (Supplementary Table S1). However, the N51I-only and C59R-only haplotypes were rare, each recorded in only two and one sample(s), respectively. The S108N-only haplotype was more common, being found in 25 samples (0.4%). The I164L mutation was very rare in this dataset and seen only twice: once together with the N51I/C59R/S108N triple mutant and once with N51I and S108N (Fig. 2b, Supplementary Table S1). In addition to the five most common haplotypes shown in Fig. 1, there were 14 additional *dhfr* haplotypes, each seen in either one or two samples, and having a combined frequency of 0.25% (Supplementary Table S1).

For many of the countries, there were only data from one or two different years, making it impossible to determine trends. Time series analysis was possible for The Gambia, Mali, Ghana and Kenya as there were multiple time points (Fig. 2). All three mutations, i.e., N51I, C59R, and S108N, increased significantly in frequency between the first and final years sampled in each country (Fig. 2a, Supplementary Table S2). In Kenya, S108N rose in frequency earlier than N51I and C59R, as expected given the established step-wise accumulation of these three *dhfr* mutations on an S108N background. However, in West African countries, N51I, C59R, and S108N increased in frequency at approximately the same rate. The prevalence of S108N increased earlier in Kenya, followed by The Gambia and Ghana, and then Mali (Fig. 2b). N51I prevalence also increased earlier in Kenya and The Gambia, followed by Ghana and then Mali. The same trend was observed for C59R.

Variation in *dhps*

Compared with *dhfr*, there was a much greater diversity of haplotypes in *dhps*. The eight most frequent *dhps* haplotypes were all seen at a frequency greater than 1%, while there were a further 43 *dhps* haplotypes with a combined frequency of 3.5% (Fig. 3, Supplementary Table S3). Over half of these (24) were singleton haplotypes seen in only one sample (Supplementary Table S3). The most frequent *dhps* haplotypes showed distinct geographic patterns (Fig. 3). Six of the eight most common haplotypes contained the mutation A437G (Fig. 3). Four haplotypes with A473G and without K540E were found predominantly in West and Central Africa.
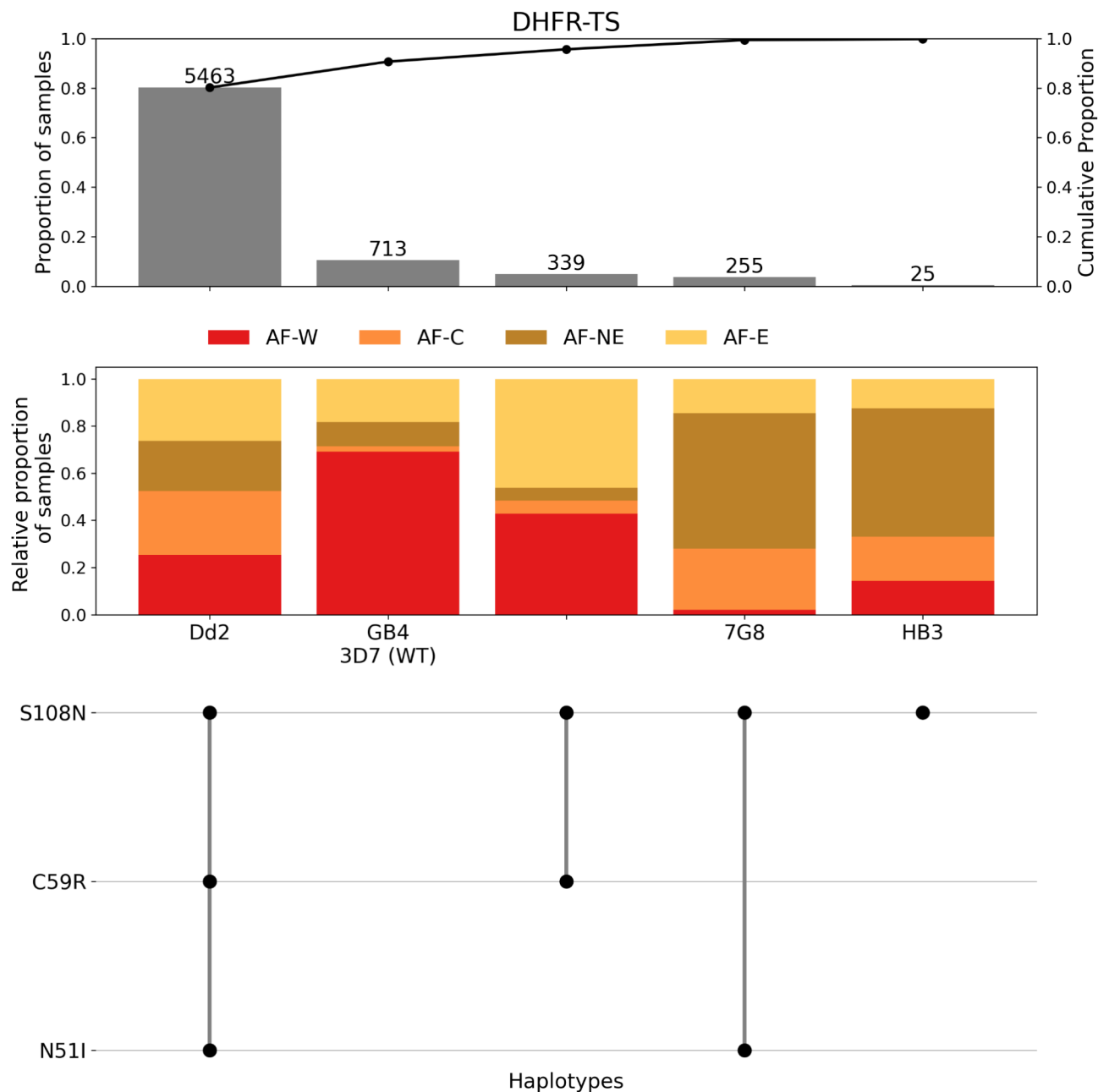
Fig. 1. Summary of the five most frequent *dhfr* haplotypes across 21 African countries. The frequency of each haplotype, in terms of sample count (number above each bar), frequency across the dataset (1st y-axis), and cumulative proportion of samples (2nd y-axis) is shown in the first panel. The proportion of each haplotype assigned to each of the major African subpopulations is shown in the second panel (AF-W = West Africa, AF-C = Central Africa, AF-NE = North-East Africa, AF-E = East Africa). The final panel illustrates the individual amino acid mutations contributing to each unique haplotype. The 3D7 wild type (WT) haplotype contains no mutations, so remains blank in the final panel.

The single mutation haplotype A437G was most common of all *dhps* haplotypes, found in 36% of all samples. Two haplotypes containing A437G alongside K540E were restricted to the other side of the continent, in East and Northeastern Africa (Fig. 3). A437G/K540E was the dominant haplotype in East Africa (74%, 907/1,234) and found in (18% of all African samples) (Fig. 3). Of the eight most common *dhps* haplotypes, two did not contain A437G (Fig. 3). These were both more geographically widespread relative to the other six common haplotypes. The single mutation haplotype S436A, was found in 11% of samples and occurred in all four African populations from this study (though predominantly was found in West Africa). The wild type *dhps* sequence, with no mutations, was the fifth most common haplotype, being seen in 7.8% of samples (Fig. 3). Of note is that the reference 3D7 strain does not contain the wild type haplotype for *dhps*, though the lab strains HB3 and IT
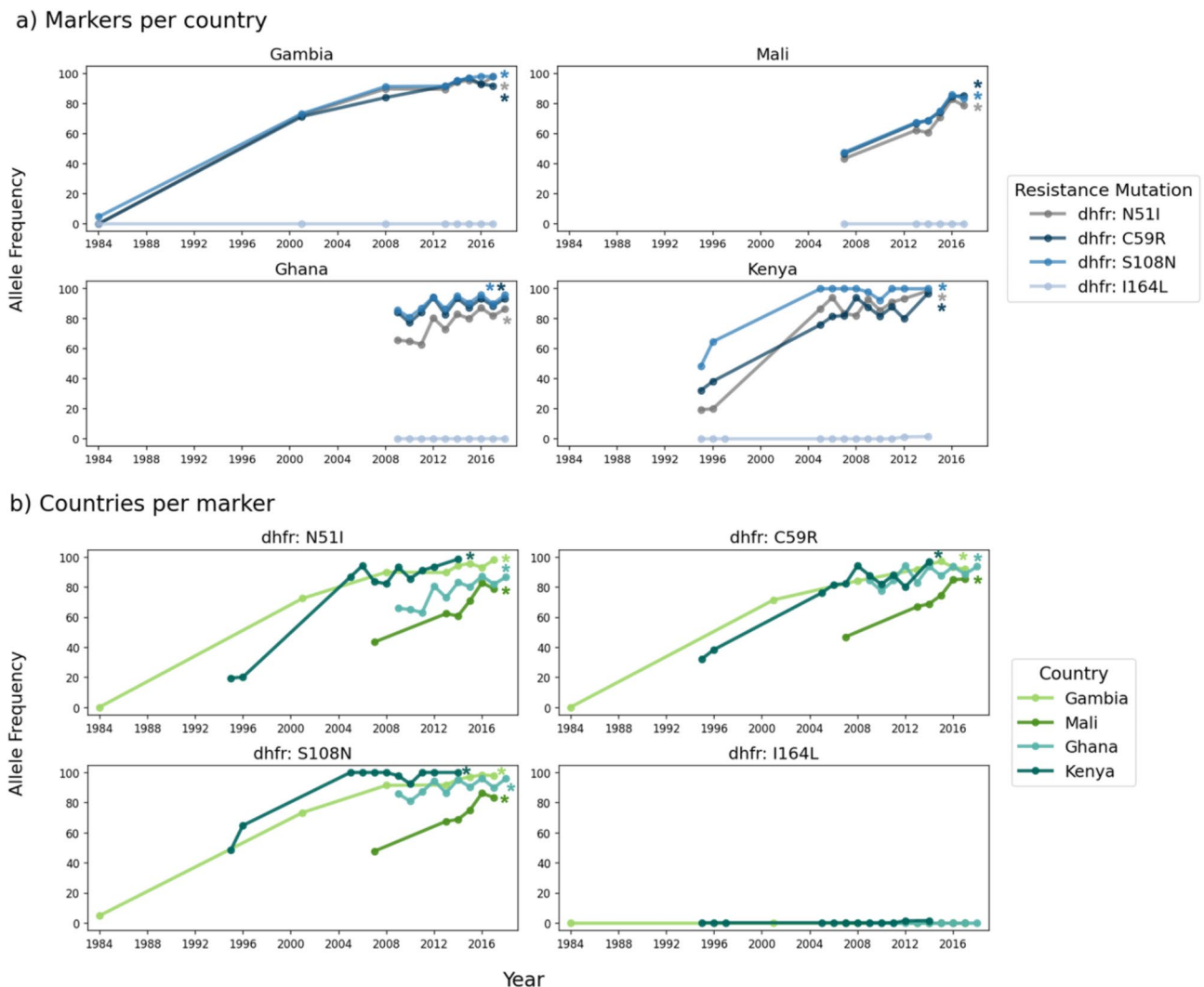
Fig. 2. Temporal analysis of *dhfr* allele frequencies. (**a**) All *dhfr* mutations within Gambia, Mali, Ghana, and Kenya, (**b**) comparison of allele frequencies for individual *dhfr* mutations across countries. Lines represent the allele frequency of that mutation originating from homozygous genotypes only. Allele frequencies were calculated only for years with at least 25 homozygous samples (see Methods for further details). Significant increasing trends over time (Fisher's exact test adjusted p-value < 0.05) are denoted with an asterisk (*). P-values were adjusted for multiple testing with a Benjamini and Hochberg correction[15].

have this haplotype. The wild type was most evenly distributed geographically of the eight most common *dhps* haplotypes.

We have data from samples collected in the 20th century from both the Gambia and Kenya. In both cases, the frequency of the A437G mutation has increased significantly since pre-2000 (Fig. 4, Supplementary Table S4). A437G reached near fixation in Kenya (2014) and a high of 80% in The Gambia. Between 2007 and 2017, A437G prevalence in Mali also increased significantly, while in Ghana it remained at a stable, high frequency between 2009 and 2018 and no significant increase was detected (Supplementary Table S4). Spatially, A473G has been more common in the south and east of West Africa than in the north and west. The frequency of the S436A mutation has increased significantly over time in The Gambia but has fallen slightly in both Mali and Ghana over time, with no significant change detected (Fig. 4b, Supplementary Table S4). The frequency of S436A has remained low in Kenya over time while the K540E mutation increased significantly here. No significant increases were detected in Ghana, but sampling only began in 2009 here.

Of the four countries for which we have time series data, the K540E mutation is seen almost exclusively in Kenya, where it reached close to fixation in recent years (Fig. 4b). It is interesting to note that in Kenya, the frequency of K540E has tracked almost exactly the frequency of A437G, suggesting co-selection in this geographical region. Although K540E is widespread across East Africa, it is very rare in both West Africa and Madagascar (Fig. 5b). Interestingly, despite such a high differentiation at K540E between East Africa and Madagascar, these populations are similar at a whole genome level (Supplementary Figure S1). It is important to note that data from only 24 isolates from Madagascar were available for this analysis.
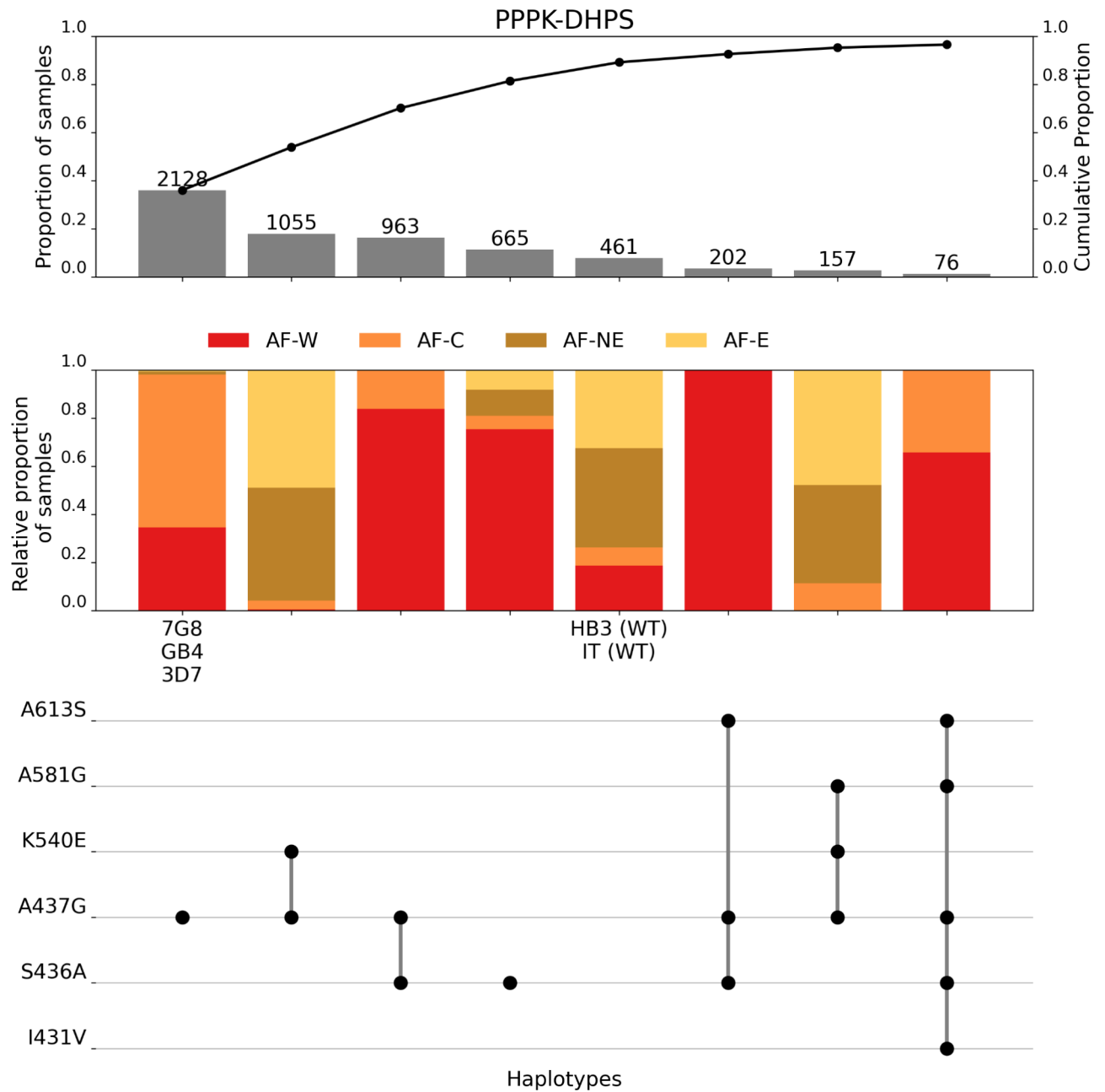
Fig. 3. Summary of the eight most frequent *dhps* haplotypes across 21 African countries. The frequency of each haplotype, in terms of sample count (number above the bar), frequency across the dataset (1st y-axis), and cumulative proportion of samples (2nd y-axis) is shown in the first panel. The proportion of each haplotype assigned to each of the major African subpopulations is shown in the second panel (AF-W = west Africa, AF-C = central Africa, AF-NE = northeast Africa, AF-E = east Africa). The final panel illustrates the individual amino acid mutations contributing to each unique haplotype. The HB3 and IT strains represent the wild type (WT) haplotype, so no mutations are shown here.

To further investigate the extent of differentiation at K540E between East Africa and Madagascar, pairwise F_{ST} values for QC-passed, genome-wide SNPs were generated between samples from these two subpopulations. Values across the genome ranged between 0 and 0.91, with a median of 0.0003 across all genome-wide variants (Table 1). The F_{ST} value for K540E was 0.748, which was greater than 99.99% of genomic positions analysed, with just 12 positions exhibiting a greater difference than 0.748 (Table 1, Supplementary Figures S2 and S3). Note that F_{ST} could be calculated for 763,535 positions. When investigating the 20 most differentiated SNPs between mainland East Africa and Madagascar, we observed SNPs in four genomic regions on chromosomes 5, 6, 8 and 14 (Table 1). These included SNPs encoding non-synonymous amino acid mutations within *aat1* and approximately 20 kb 3' of *mdr1*, which had even higher F_{ST} than the SNP encoding K540E (Table 1). These may suggest novel drug-resistance mutations.
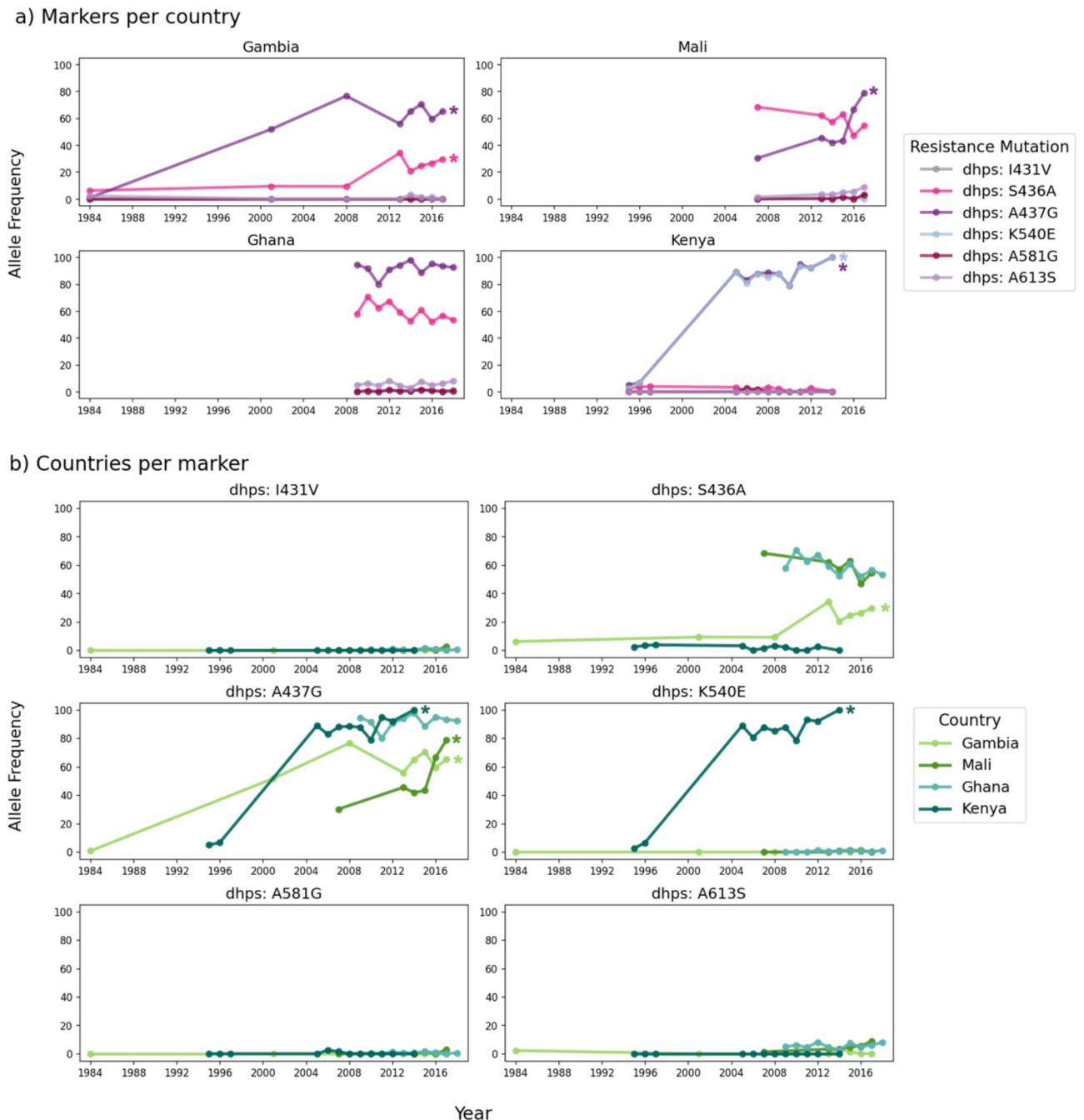
Fig. 4. Temporal analysis of *dhps* allele frequencies. (**a**) all *dhps* mutations within Gambia, Mali, Ghana, and Kenya are shown in the first four plots, (**b**) comparison of allele frequencies for individual *dhps* mutations across countries. Lines represent the allele frequency of that mutation originating from homozygous genotypes only. Allele frequencies were calculated only for years with at least 25 homozygous samples (see Methods for further details). Significant increasing trends over time (Fisher's exact test adjusted p-value < 0.05) are denoted with an asterisk (*). P-values were adjusted for multiple testing with a Benjamini and Hochberg correction[15].

The six highest F_{ST} SNPs were found in the 6 kb region Pf3D7_05_v3:1139023–1,145,481. All were found at a frequency ≥ 90% in Madagascar, ≤ 1% in east Africa, and were in very high linkage disequilibrium (LD) with each other in Madagascar ($R^2 ≥ 0.88$) (Table 1, Supplementary Figure S4). This indicates that some force of selection has co-selected these loci in Madagascar. The 14 kb region, Pf3D7_06_v3:1201098–1,215,233, contained five SNPs which were all very high frequency in east Africa (≥ 69%) but low frequency in Madagascar (< 5%). The most highly differentiated of these was a SNP causing S258L in the gene *aat1*, presumed to be associated with drug resistance[16]. These five SNPs were in LD with each other in east Africa ($R^2 ≥ 0.22$) and, interestingly, in west Africa ($R^2 ≥ 0.17$) (Supplementary Figures S5 and S6). This suggests the presence of hitchhiker mutations, swept along with a causal variant. In the chromosome 8 region, in addition to the SNP causing K540E in *dhps*,
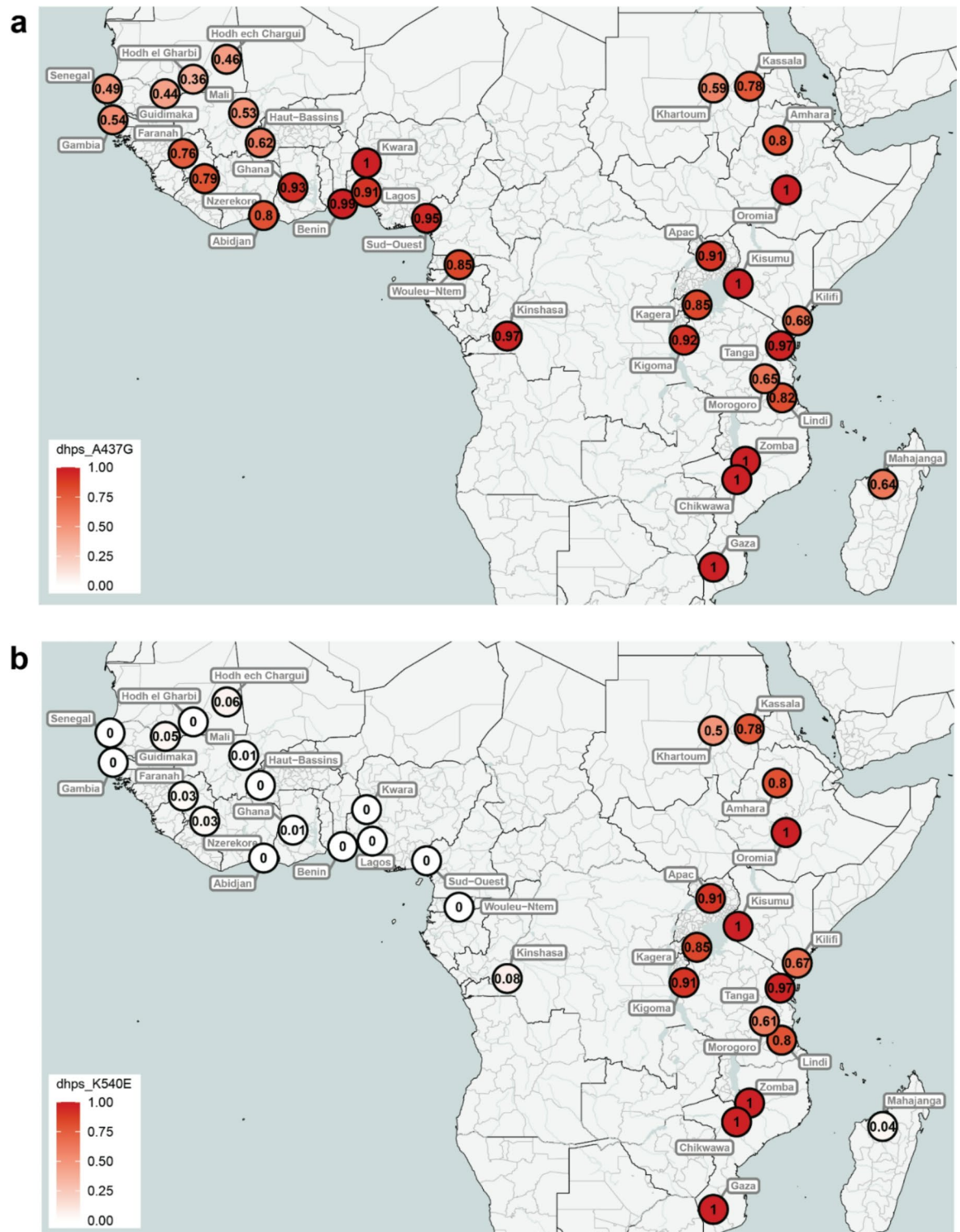
Fig. 5. Distribution of mutation frequencies for (**a**) *dhps* A437G and (**b**) *dhps* K540E. Data were aggregated at the administrative division 1 level. To improve visualisation and avoid overlapping markers for nearby locations, data from Senegal, Gambia, Ghana, Mali and Benin were combined at the country level for panel b. Colours represent mutation frequencies ranging from 0 (no mutation detected) to 1 (100% of the parasites in the location carrying the mutation). Generated using the mapMutationPrevalence function from the grcMalaria R package (version 2.0.0)[17].

Genomic Position	REF	ALTs	East Africa ALT Freq.	Madagascar ALT Freq.	F_{ST}	Gene ID	Gene name	Amino acid change	Mutation type
Pf3D7_05_v3:1141578	A	C	0.00	0.92	0.91	PF3D7_0527400			3' UTR variant
Pf3D7_05_v3:1145481	C	A	0.00	0.92	0.91	PF3D7_0527500	HIP		3' UTR variant
Pf3D7_05_v3:1139023	A	G	0.01	0.92	0.90	Intergenic region: PF3D7_0527300 - PF3D7_0527400			Intergenic region
Pf3D7_05_v3:1140984	T	G	0.00	0.90	0.89	PF3D7_0527400		Leu120Val	Non-synonymous
Pf3D7_05_v3:1140987	G	A	0.00	0.90	0.89	PF3D7_0527400		Asp121Asn	Non-synonymous
Pf3D7_05_v3:1140988	A	G	0.00	0.90	0.89	PF3D7_0527400		Asp121Gly	Non-synonymous
Pf3D7_06_v3:1215233	G	A	0.87	0.00	0.87	PF3D7_0629500	AAT1	Ser258Leu	Non-synonymous
Pf3D7_06_v3:1214259	C	G	0.84	0.00	0.84	PF3D7_0629500	AAT1		Intron variant
Pf3D7_06_v3:1214350	G	T	0.86	0.04	0.80	PF3D7_0629500	AAT1	Ile552Ile	Synonymous
Pf3D7_05_v3:1151093	C	A	0.00	0.81	0.81	Intergenic region: PF3D7_0527700 - PF3D7_0527800			Intergenic region
Pf3D7_08_v3:546841	T	A	0.78	0.00	0.78	PF3D7_0810700		Lys68Asn	Non-synonymous
Pf3D7_05_v3:1133912	T	A	0.00	0.77	0.76	PF3D7_0527200	USP14	Thr281Thr	Synonymous
Pf3D7_08_v3:549993	A	G	0.81	0.04	0.75	PF3D7_0810800	PPPK-DHPS	Lys540Glu	Non-synonymous
Pf3D7_05_v3:1119264	A	G	0.82	0.08	0.70	PF3D7_0526700		Tyr130His	Non-synonymous
Pf3D7_14_v3:2183418	A	T	0.05	0.79	0.72	PF3D7_1453200		Asn756Tyr	Non-synonymous
Pf3D7_05_v3:1114069	C	T, G	0.07	0.79	0.69	PF3D7_0526600		Ile4289Met (> G), Ile4289Ile (> T)	Non-synonymous (> G), Synonymous (> T)
Pf3D7_14_v3:2183653	C	A	0.01	0.73	0.70	PF3D7_1453200		Ser834Tyr	Non-synonymous
Pf3D7_14_v3:2183719	G	A	0.05	0.77	0.70	PF3D7_1453200		Arg856Gln	Non-synonymous
Pf3D7_06_v3:1206498	G	A	0.71	0.00	0.69	PF3D7_0629300	PL	Glu437Lys	Non-synonymous
Pf3D7_06_v3:1201098	T	C	0.69	0.00	0.69	PF3D7_0629200	DnaJ		5' UTR variant

Table 1. The 20 most differentiated single nucleotide polymorphisms between AF-E ($n = 1,515$) and Madagascar ($n = 24$), based on alternate allele frequency. REF = 3D7 reference strain nucleotide, ALT = alternate nucleotide.

a SNP around 3kbp away (Pf3D7_08_v3:546841) was also found at very high frequency in east Africa but was not seen in Madagascar. Again, we presume this is a hitchhiker mutation, co-selected with K540E in east Africa. Three SNPs on chromosome 14 were observed at high frequency in Madagascar and were strongly linked there too, with $R^2 > 0.74$ (Supplementary Figure S4).

Reliability of using single markers as proxies for full gene haplotypes

It has been suggested that the order in which mutations accumulated in *dhfr* was S108N, followed by N51I, followed by C59R[18,19]. If this were indeed the case, then it would be possible to use the marker C59R as a proxy for the N51I/C59R/S108N triple mutant (Supplementary Table S5). This study demonstrates that there are examples when, for example, C59R is seen in haplotypes that include S108N but not N51I. Therefore, C59R is certainly not a perfect proxy for the triple mutant. The reliability of the C59R marker as a proxy for the triple mutation can be assessed by determining what proportion of samples containing this mutation do indeed have the triple mutation. In all countries, this reliability score is over 80% (Fig. 6). However, reliability differs across countries, for example being only 85% in Kenya. If positions 51, 59 and 108 are genotyped, the vast majority of variation will be captured. Given that mutations at positions 51 and 59 are almost always found on the background of S108N, genotyping one, or both, of them should be sufficient to determine most of the variation within the gene.

Likewise, the combination of C59R and K540E markers are generally reliable markers of the quintuple multi-gene haplotype *dhfr* N51I/C59R/S108N, *dhps* A437G/K540E. However, the reliability of this combination of markers is lower in Kenya (89%) than in many other countries (Fig. 6). Finally and based on our results, while in Tanzania A581G seems to be a reliable marker of the multi-gene, sextuple haplotype *dhfr* N51I/C59R/S108N and *dhps* A437G/K540E/A581G, this was not the case across Africa. A581G in West African countries such as Ghana and Cameroon was never a proxy for this sextuple haplotype (Fig. 6). Therefore, a marker may be a reliable proxy of a haplotype in one region but not necessarily in other regions.

Discussion

SP resistance emerged approximately three decades ago and led to the discontinuation of SP as the sole first-line treatment for malaria in the early 2000s[3]. However, SP is used for chemoprevention during pregnancy, infancy and childhood. Thus, monitoring genetic markers in *dhfr* and *dhps* associated with SP resistance is informative for interventions relying on SP. This study describes the evolution of mutations associated to SP resistance across Africa, the continent with the highest malaria burden, where SP-based interventions are widely implemented.
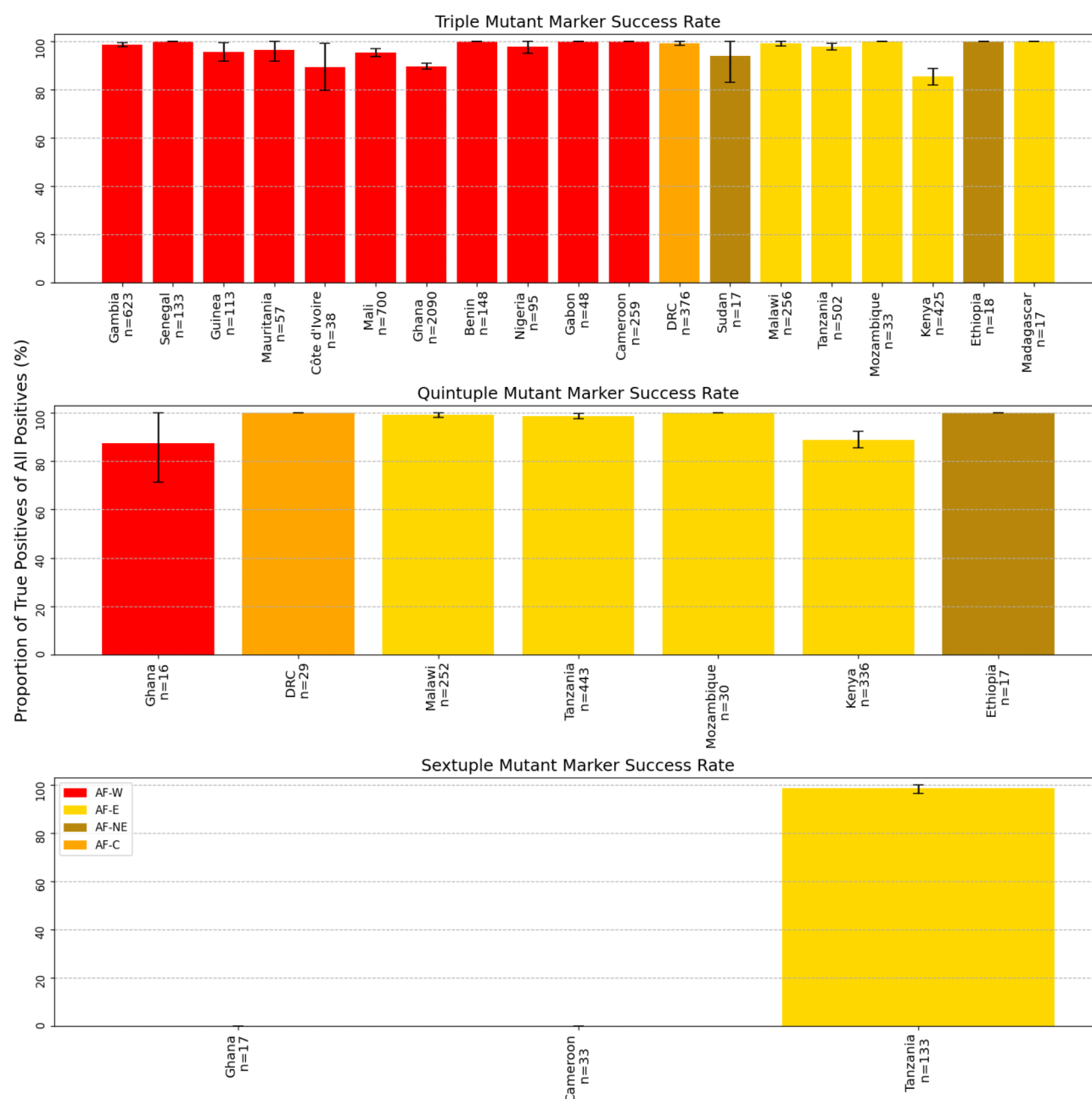
Fig. 6. Reliability of single mutation markers of haplotypes. Success rate of suggested markers to genotype for the triple, quintuple and sextuple multi-mutant haplotypes in all African countries where n > 15. Triple = *dhfr* C59R for *dhfr* C59R/N51I/S108N. Quintuple = *dhfr* C59R and *dhps* K540E for *dhfr* C59R/N51I/S108N, *dhps* A437G/K540E. Sextuple = *dhps* A581G for *dhfr* N51I/C59R/S108N, *dhps* A437G/ K540E/A581G. Bars are coloured according to major subpopulation (AF-W = west Africa, AF-C = central Africa, AF-NE = northeast Africa, AF-E = east Africa). Error bars represent 95% confidence intervals calculated as 1.96 x the standard error.

Prevalence of the triple mutant *dhfr* haplotype was high across Africa, present in over 80% of all samples analysed. The prevalence of the mutations *dhfr* N51I, C59R, and S108N in The Gambia, Ghana, Mali, and Kenya increased over 30 years to near-fixation, possibly due to the selection exerted by the widespread use of SP[7]. In some cases, like S108N in Kenya, alleles were close to fixation at the point of SP withdrawal in favour of artemisinin combination therapies in the early 2000s[3]. Despite the withdrawal of SP as first-line treatment and thus a decreased evolutionary pressure over the last 20–25 years, the parasite wild types did not increase. Rather, the high prevalence of the *dhfr* mutations suggest an ongoing and continuous selection for SP resistance, with the risk of impacting on the SP-based interventions.

Although the selective pressure exerted by the SP used for IPTp, PMC or SMC may explain in some countries the high prevalence of the triple mutant, this could also be due to the acquisition of other mutations

compensating the loss of fitness associated to the mutations. This is supported by the low but significant prevalence of haplotypes combining two of the three *dhfr* triple mutant loci. These haplotypes were C59R/S108N and N51I/S108N and were the third and fourth most common *dhfr* haplotypes, respectively. The accumulation of the triple mutant alleles seen in this study is concordant with the widely accepted step-wise order of S108N, then N51I and C59R[4]. In Kenya, the earlier fixation of S108N supports the theory of later introduction of N51I and C59R. In the Gambia, however, all three triple mutant alleles increased at the same rate, not in contrast to the stepwise accumulation hypothesis, but also not clearly demonstrating the event.

There was no joint N51I/C59R *dhfr* haplotype, suggesting that both N51I and C59R have arisen on an S108N background, which is the most frequent single mutant of *dhfr*. The N51I/C59R and C59R-only haplotypes have been described in Melanesia, and are thought to have originated independently from the dominant N51I/C59R/S108N triple mutant haplotype[20]. The *dhfr* C59R-only haplotype was first reported in Vietnam, though at low frequencies[21]. In Central and North-East Africa, C59R/S108N was rare, while N51I/S108N was found at over 25% in these populations. Therefore, N51I and C59R may have been introduced, emerged, or been selected differently in different populations on the S108N background. Selection of independently emerging double mutants has been shown in previous studies in Ghana and Kenya, with non-Southeast Asian indigenous haplotypes of double mutants on the S108N background[22]. Both double haplotypes (N51I/S108N and C59R/S108N) in West and East Africa may have spread by recent human migration or may be due to the breakdown of triple haplotype by recombination or reverse mutation. However, this is extremely difficult to establish. S108N was probably the earliest mutation in Africa, followed by multiple introductions of either or both N51I and C59R from Southeast Asia. Various studies of the fitness and epidemiology of combined mutations have suggested a stepwise evolution of mutants; N51/C59/S108N/I164 → N51I/C59/S108N/I164 or N51/C59R/S108N/I164 → N51I/C59R/S108N/I164 → N51I/C59R/S108N/I164L[18,19]. We identified only three isolates in East Africa carrying I164L. A deeper understanding of the evolutionary history of this gene in Africa will require extensive historical data from samples taken around the time of the initial rollout of SP.

The distribution of the *dhps* alleles and haplotypes was more heterogeneous than for *dhfr*. This heterogeneity highlights the differing evolutionary impact of SP use between regions, which should be considered in ongoing SP interventions. A combination of I431V, S436A, A437G, and A613S was dominant in West and Central Africa, while A437G, K540E and A581G were more common in East and North-East Africa. This supports the theory of multiple independent emergences of *dhps* mutant alleles in different African regions[23]. Within West Africa, the increasing prevalence of A437G in the Gambia seems to have "stalled" in 2016 at around 60%. Since 2016, A437G has been increasing and reached higher prevalence levels in Mali and Ghana. Such difference between countries on the western Atlantic coast and countries in the middle of West Africa has been shown for chloroquine resistance associated alleles[14]. Assessment of the *crt* K76T locus showed that, between 2013 and 2016, parasites in Ghana, Mali, and Côte d'Ivoire had reverted almost entirely to the wildtype, while the mutant types remained at higher prevalence in Gambia, Senegal and Mauritania. Thus, the local environment and possibly the history and scale up of SP-based interventions may play an important role in the dynamics of SP resistance alleles. Beyond this, there could be other effects impacting SP resistance haplotypes such as the host genetic background in the vector and the human host[24]. In Mali and Ghana, the prevalence of S436A has been declining while the rarer A613S mutation has recently emerged and spread. Combined with the emergence of I431V and the low but increasing frequencies of K540E and A581G, these additional mutations could seriously impact the efficacy of SP[13].

Given the concurrent increase of K540E and A437G in Kenya, it is perhaps surprising that K540E did not spread in West Africa, where A437G is highly prevalent (> 60%). In West Africa, a more commonly co-occurring mutation with A437G is S436A possibly because it provides a fitter phenotype. K540E was rare in Madagascar but common in the nearest continental mainland. Investigation of other highly differentiated SNPs between East Africa and Madagascar revealed several other loci outside of *dhfr* and *dhps* that may be involved in drug resistance. For example, the S258L mutation within *aat1* is associated with chloroquine resistance, with a fitness cost to the parasite[16]. The S258L mutation in *aat1* was also shown to have increased to near-fixation in the Gambia over a similar time span to this study (the years 1984–2014). The absence of the *aat1* S258L mutation in Madagascar also points to the regional differences in selective pressures across Africa. Similarly, the SNPs 3' of *mdr1* are notable as *mdr1* has speculatively been linked to amodiaquine resistance[25,26]. Further studies could make use of the Pf7 dataset to perform a continent-wide assessment of these loci. The exploration of loci beyond *dhfr* and *dhps* was possible thanks to whole-genome sequencing data. These also enabled the calculation of full-gene haplotypes for SP resistance genes, which revealed important variation in the reliability of certain proxy markers between different African regions. Further related work using whole genome data could include the investigation of intermediate haplotypes between the wild type and SP-resistance haplotypes for *dhps* and *dhfr*. For example, many studies have compared *dhfr* N51/C59/S108 vs N51I/C59R/S108N haplotypes and showed similar MS calls in the latter but not the former[27–29]. Using available WGS data to generate and then compare microsatellite loci around *dhfr* N51/C59/S108N, N51I/C59/S108N, and N51/C59R/S108N haplotypes could provide us with a more detailed picture on the evolution of SP resistance.

We provided an insight to SP resistance haplotype frequencies across 22 countries, but we were limited in our ability to describe temporal trends within these locations due to a lack of consistent sampling across the study period. Longitudinal data from four countries enabled us to highlight differing mutation dynamics within West Africa and between East and West Africa. Even within the locations for which several years of sufficient data existed, most samples were from before 2017, as the Pf7 resource extends only until 2018[14]. In the several years between then and now, advances in malaria control have been made but we are currently unable to draw conclusions about SP-resistance linked mutations from the Pf7 dataset for the 2020s[30]. While Pf7 is a vast resource, enabling continent-wide analyses, it is important to note the inherent biases of the dataset. Pf7 collates many individual studies, which may have been designed for a purpose other than general genomic surveillance.

As such, not all sampling in Pf7 will be stratified or fully representative of the parasite population in a country. Consequently, the trends described in this section may reflect overrepresented areas more than under sampled areas, on a local, country, and regional scale in Africa. Increasing the number of countries and years for which sufficient data exists requires longitudinal sampling. These data would increase the specificity of conclusions drawn about SP evolution in Africa, which is essential to ensure the continued efficacy of SP-based interventions such as SMC. In the absence of other openly available time series data from Africa, databases such as those of the Worldwide Antimalarial Resistance Network (WWARN - https://www.iddo.org/wwarn) could be explored. For example, since 2000, WWARN publishes information on the *dhfr* and *dhps* mutations in their SP Molecular Surveyor, collated across 77 countries. However, replicating the analyses of this study would not be possible with WWARN data because full gene haplotypes are not always represented. For example, *dhfr* multi-mutant haplotypes with I164L or the joint *dhps* 436 and 437 haplotypes are not included. Moving towards whole genome sequencing could increase the availability of shared *dhfr*, *dhps* haplotypes within such databases.

This study has highlighted the persistence and regional variability of SP resistance markers in *P. falciparum* across Africa, possibly driven by the continued use of SP for chemoprevention. The recent, near-fixation of the *dhfr* triple mutant haplotype in several countries and the wide diversity of *dhps* haplotypes across the continent underscore the ongoing selection pressure from SP use. The distinct regional patterns of allele frequencies illustrate the complex dynamics of SP resistance, shaped by local ecological factors, migration, and past antimalarial treatment policies. We have highlighted several avenues for further exploration, including tracing the evolutionary origins of the *dhfr* double mutant haplotypes and the assessment of several SNPs beyond *dhfr* and *dhps* in regions beyond East Africa. Continued molecular surveillance is an essential component of the suite of epidemiological tools which, together, can ensure the future effectiveness of SP-based interventions. For example, some National Malaria Control Programs like that of the Gambia plan to do therapeutic efficacy studies for SP, and this is likely the case for many countries implementing SMC. A recent cluster randomised SP-IPTi trial in Nigeria revealed no benefit on asymptomatic malaria and hospitalisations after months of intervention[31]. This was hypothesised to be linked to the emergence of SP-resistance mutations such as *dhps* I431V. Furthermore, whole-genome sequencing can play an important part in this surveillance by offering a view of resistance evolution beyond established markers. Additionally, moving towards longitudinal sampling would enhance our understanding of changes in SP resistance dynamics over time, informing targeted, region-specific strategies to combat malaria in Africa.

Methods

Analysis of genetic variation

Full gene haplotypes, generated from concatenated exon sequences, and their summaries across populations were generated using a custom Python script. For a temporal analysis of mutation frequencies, we investigated variation at four *dhfr* loci linked to SP resistance: 51, 59, 108, and 164 and six *dhps* loci: 431, 436, 437, 540, 581, and 613. We used the allele frequency for all *dhfr* and *dhps* markers. There are multiple ways in which mutation frequencies can be calculated (see Methods: calculating allele frequencies), but here we excluded from further analysis any sample which had a heterozygous mutation at any position within the coding region of the gene and calculated frequencies from homozygous haplotypes only. Custom Python scripts were used to compare marker allele frequencies between countries for which we had sufficient data and time points with at least 25 homozygous samples: the Gambia, Mali, Ghana, and Kenya. For each country and locus combination in the temporal analysis, a Fisher's exact test was used to determine whether the difference in derived allele frequency between the first and final sampled time point was significant.

Maps of dhps A437G and K540E allele frequencies across Africa were generated using the mapMutationPrevalence function from the grcMalaria R package (version 2.0.0)[17]. Investigations of genetic dissimilarity between East Africa and Madagascar were performed using a neighbour-joining tree and a genome-wide comparison of F_{ST} and allele frequencies between these population subsets. The neighbour-joining tree was based on the same genetic distance matrix as calculated for Pf7[14], but a subset to include only Malagasy samples and a random subset from each African population using custom R scripts. Each African population (West, Central, Northeast, and East without Madagascar) was subset to 50 random individuals, while maintaining the contributing proportions of each country found in the full dataset. F_{ST} was computed from QC-passed SNPs and samples from Madagascar ($n = 24$) and from the remainder of East Africa ($n = 1,515$) using scikit-allel. The same was done for the genomic position encoding the K540E mutation (Pf3D7_08_v3:549993), and values were compared to genome-wide estimates. Linkage disequilibrium (R^2) was estimated for the 20 SNPs with highest F_{ST} using scikit-allel rogers_huff_r method, in the Madagascar, East African, and West African populations.

To assess the reliability of single markers as proxies for full gene haplotypes, we calculated the proportion samples with each single marker proxy, which also held the correct, associated full gene haplotype. This proportion was denoted as a marker "success rate". Proxy marker success rates were generated for countries with > 15 samples using custom Python scripts.

Calculating allele frequencies

Because some malaria samples contain parasite clones with different genotypes (sometimes referred to as complexity of infection or CoI), the genotype calls for a mutation can be heterozygous. When calculating frequencies, decisions need to be made about how to treat heterozygous calls. Assume we have R homozygous reference calls, A homozygous alternative calls and H heterozygous calls. One approach that is sometimes used is to calculate the carrier frequency. This is the proportion of samples that carry the alternate allele and is calculated as $\frac{(A + H)}{(R + A + H)}$. When considering drug-resistance mutations, it is sometimes assumed that if a patient

contains both sensitive and resistant parasites, the infection should be considered drug-resistant on the basis that in the presence of the drug, the resistant parasites are likely to out-compete the sensitive parasites, and therefore the carrier frequency is perhaps the most accurate way to estimate the proportions of infections that will be resistant. Population genetics, however, tends to be more concerned with allele frequencies in the population of all parasites. One unbiased estimate of the allele frequency is to simply ignore heterozygous genotype calls, and calculate the proportion of homozygous alternate calls as a proportion of all homozygous calls, i.e. $\frac{A}{(R + A)}$

. This is the method we have used in this paper, excluding heterozygotes on a locus-by-locus basis, rather than a sample-wise basis, to maximise the data available for analysis.

Data availability

References

1. WHO. World malaria report 2023. (2023).
2. WHO. Seasonal malaria chemoprevention with sulfadoxine–pyrimethamine plus amodiaquine in children: a field guide, second edition. (2023).
3. *Guidelines for the Treatment of Malaria*. (2006).
4. Plowe, C. V. et al. Mutations in plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of Pyrimethamine-Sulfadoxine use and resistance. *J. Infect. Dis.* **176**, 1590–1596 (1997).
5. Roper, C. et al. Intercontinental spread of Pyrimethamine-Resistant malaria. *Science* **305**, 1124–1124 (2004).
6. Mita, T. et al. Limited geographical origin and global spread of Sulfadoxine-Resistant Dhps alleles in plasmodium falciparum populations. *J. Infect. Dis.* **204**, 1980–1988 (2011).
7. Sridaran, S. et al. Anti-folate drug resistance in Africa: meta-analysis of reported dihydrofolate reductase (dhfr) and dihydropteroate synthase (dhps) mutant genotype frequencies in African plasmodium falciparum parasite populations. *Malar. J.* **9**, 247 (2010).
8. Amimo, F. et al. *Plasmodium falciparum* resistance to sulfadoxine-pyrimethamine in Africa: a systematic analysis of National trends. *BMJ Glob Health*. **5**, e003217 (2020).
9. Happi, C. T. et al. Polymorphisms in *Plasmodium falciparum Dhfr* and *Dhps* genes and age related in vivo sulfadoxine–pyrimethamine resistance in malaria-infected patients from Nigeria. *Acta Trop.* **95**, 183–193 (2005).
10. Mbogo, G. W. et al. Temporal changes in prevalence of molecular markers mediating antimalarial drug resistance in a high malaria transmission setting in Uganda. *Am. J. Trop. Med. Hyg.* **91**, 54 (2014).
11. Juliano, J. J. et al. Prevalence of mutations associated with Artemisinin partial resistance and sulfadoxine–pyrimethamine resistance in 13 regions in Tanzania in 2021: a cross-sectional survey. *Lancet Microbe* **0**, (2024).
12. Naidoo, I. & Roper, C. Mapping 'partially resistant', 'fully resistant', and 'super resistant' malaria. *Trends Parasitol.* **29**, 505–515 (2013).
13. Oguike, M. C. et al. Molecular determinants of sulfadoxine-pyrimethamine resistance in *Plasmodium falciparum* in Nigeria and the regional emergence of *Dhps* 431V. *Int. J. Parasitol. Drugs Drug Resist.* **6**, 220–229 (2016).
14. MalariaGEN et al. Pf7: an open dataset of plasmodium falciparum genome variation in 20,000 worldwide samples. *Wellcome Open. Res.* **8**, 22 (2023).
15. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
16. Amambua-Ngwa, A. et al. Chloroquine resistance evolution in plasmodium falciparum is mediated by the putative amino acid transporter AAT1. *Nat. Microbiol.* **8**, 1213–1226 (2023).
17. Verschuuren, T. D. et al. Genetic surveillance of Plasmodium falciparum reveals rapid population changes following first-line treatment policy revisions in the Greater Mekong Subregion. 06.06.24308535 Preprint at (2024). https://doi.org/10.1101/2024.06.06.24308535 (2024).
18. Lozovsky, E. R. et al. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci.* 106, 12025–12030 (2009).
19. Brown, K. M. et al. Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Mol. Biol. Evol.* **27**, 2682–2690 (2010).
20. Wang, P. et al. Resistance to antifolates in *Plasmodium falciparum* monitored by sequence analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. *Mol. Biochem. Parasitol.* **89**, 161–177 (1997).
21. Mita, T. et al. Independent evolution of pyrimethamine resistance in plasmodium falciparum isolates in melanesia. *Antimicrob. Agents Chemother.* **51**, 1071–1077 (2007).
22. Mita, T., Tanabe, K. & Kita, K. Spread and evolution of *Plasmodium falciparum* drug resistance. *Parasitol. Int.* **58**, 201–209 (2009).
23. Pearce, R. J. et al. Multiple origins and regional dispersal of resistant Dhps in African plasmodium falciparum malaria. *PLoS Med.* **6**, e1000055 (2009).
24. Mharakurwa, S. et al. Malaria antifolate resistance with contrasting plasmodium falciparum dihydrofolate reductase (DHFR) polymorphisms in humans and Anopheles mosquitoes. *Proc. Natl. Acad. Sci.* **108**, 18796–18801 (2011).
25. Holmgren, G. et al. Selection of *pfmdr1* mutations after amodiaquine monotherapy and amodiaquine plus Artemisinin combination therapy in East Africa. *Infect. Genet. Evol.* **7**, 562–569 (2007).
26. Humphreys, G. S. et al. Amodiaquine and artemether-lumefantrine select distinct alleles of the plasmodium falciparum mdr1 gene in Tanzanian children treated for uncomplicated malaria. *Antimicrob. Agents Chemother.* **51**, 991–997 (2007).
27. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in Southeast Asian malaria parasites. *Mol. Biol. Evol.* **20**, 1526–1536 (2003).
28. Roper, C. et al. Antifolate antimalarial resistance in Southeast Africa: a population-based analysis. *Lancet* **361**, 1174–1181 (2003).
29. Maïga, O. et al. A shared Asian origin of the Triple-Mutant Dhfr allele in plasmodium falciparum from sites across Africa. *J. Infect. Dis.* **196**, 165–172 (2007).
30. Impouma, B., Adidja, A., Mboussou, F., Cabore, J. & Moeti, M. Malaria vaccine introduction in Africa: progress and challenges. *Lancet* **405**, 521–524 (2025).
31. Adeleke, O. T., Oyenuga, A., Slusher, T. M. & Gbadero, D. A. Cluster-randomized controlled trial of intermittent preventive treatment in infancy using sulfadoxine–pyrimethamine (SP-IPTi): a pilot study in Nigeria. *J. Trop. Pediatr.* **69**, fmad001 (2023).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-98035-7.

Correspondence and requests for materials should be addressed to R.D.P. or A.A.-N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.