Spatio-temporal distribution of antimalarial drug resistant gene mutations in a *Plasmodium falciparum* parasite population from Kilifi, Kenya: A 25-year retrospective study

Irene Omedo1-3, Brian Bartilol1, Domtila Kimani1, Sónia Gonçalves3, Eleanor Drury3, Martin K. Rono1, Abdirahman I. Abdi1, Jacob Almagro-Garcia3,4, Roberto Amato3, Richard D. Pearson3,4, Lynette Isabella Ochola-Oyier1, Dominic Kwiatkowski3,4, Philip Bejon1,2

1Bioscience department, KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya
2Nuffield Department of Medicine, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK
3Wellcome Sanger Institute, Hinxton, UK
4MRC Centre for Genomics and Global Health, Big Data Institute, University of Oxford, Oxford, UK

First published: 08 Feb 2022, 7:45
https://doi.org/10.12688/wellcomeopenres.17656.1
Latest published: 08 Feb 2022, 7:45
https://doi.org/10.12688/wellcomeopenres.17656.1

**Abstract**

**Background:** Antimalarial drug resistance is a major obstacle to sustainable malaria control. Here we use amplicon sequencing to describe molecular markers of drug resistance in *Plasmodium falciparum* parasites from Kilifi county in the coastal region of Kenya over a 25-year period.

**Methods:** We performed *P. falciparum* amplicon sequencing on 1162 malaria-infected blood samples collected between 1994 and 2018 to identify markers of antimalarial drug resistance in the *Pfcrt*, *Pfdhfr*, *Pfdhps*, *Pfmdr1*, *Pfexo*, *Pfkelch13*, *plasmepsin 2/3*, *Pfarps10*, *Pfia*, and *Pfmdr2* genes. We further interrogated parasite population structure using a genetic barcode of 101 drug resistance-unrelated single nucleotide polymorphisms (SNPs) distributed across the genomes of 1245 *P. falciparum* parasites.

**Results:** Two major changes occurred in the parasite population over the 25 years studied. In 1994, approximately 75% of parasites carried the marker of chloroquine resistance, CVIET. This increased to 100% in 1999 and then declined steadily, reaching 6.7% in 2018. Conversely, the quintuple mutation form of sulfadoxine-pyrimethamine resistance increased from 16.7% in 1994 to 83.6% in 2018. Several non-synonymous mutations were identified in the *Kelch13* gene, although none of them are currently associated with artemisinin resistance. We observed a temporal increase in the *Pfmdr1* NFD haplotype associated
with lumefantrine resistance, but observed no evidence of piperazine resistance. SNPs in other parts of the genome showed no significant temporal changes despite the marked changes in drug resistance loci over this period.

**Conclusions:** We identified substantial changes in molecular markers of *P. falciparum* drug resistance over 25 years in coastal Kenya, but no associated changes in the parasite population structure.

**Keywords**
amplicon sequencing, antimalarial drug resistance, genetic barcode, single nucleotide polymorphisms, spatio-temporal analysis, Plasmodium falciparum, population structure

---

Corresponding author: Irene Omedo (iomedo@kemri-wellcome.org)

**Author roles:** Omedo I: Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Bartilol B: Data Curation, Investigation, Methodology, Validation, Writing – Review & Editing; Kimani D: Data Curation, Investigation, Methodology, Project Administration, Writing – Review & Editing; Gonçalves S: Data Curation, Investigation, Methodology, Project Administration, Validation, Writing – Review & Editing; Drury E: Data Curation, Investigation, Methodology, Validation, Writing – Review & Editing; Abdi AI: Conceptualization, Resources, Writing – Review & Editing; Almagro-Garcia J: Data Curation, Methodology, Software, Writing – Review & Editing; Amato R: Data Curation, Methodology, Software, Writing – Review & Editing; Pearson RD: Methodology, Software, Validation, Writing – Review & Editing; Ochola-Oyier LI: Conceptualization, Resources, Writing – Review & Editing; Kwiatkowski D: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Review & Editing; Bejon P: Conceptualization, Methodology, Resources, Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome [204911; a Wellcome Trust collaborative grant awarded to Prof. Dominic Kwiatkowski].

**Copyright:** © 2022 Omedo I et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Omedo I, Bartilol B, Kimani D et al. Spatio-temporal distribution of antimalarial drug resistant gene mutations in a *Plasmodium falciparum* parasite population from Kilifi, Kenya: A 25-year retrospective study [version 1; peer review: awaiting peer review] Wellcome Open Research 2022, 7:45 https://doi.org/10.12688/wellcomeopenres.17656.1

**First published:** 08 Feb 2022, 7:45 https://doi.org/10.12688/wellcomeopenres.17656.1
Introduction

Malaria is a major global health burden, and the spread of resistance to life-saving antimalarial drugs is of major concern. *Plasmodium falciparum* parasites have developed resistance to all major antimalarial drugs, including chloroquine and sulfadoxine-pyrimethamine, and more recently, artemisinin and its partner drugs. Chloroquine was widely used across sub-Saharan Africa as far back as the 1940s to treat uncomplicated *Plasmodium falciparum* malaria infections. Resistance to chloroquine was first detected in Kenya in 1978, but the drug continued to be used until 1998, when a change in treatment guidelines saw a shift to sulfadoxine-pyrimethamine (SP). Resistance to SP developed rapidly and by 2004, SP had to be replaced with a more effective alternative, artemether-lumefantrine. Following the withdrawal of chloroquine from frontline use, the molecular markers of resistance to the drug gradually reduced in frequency and the parasite reverted to chloroquine susceptibility.

Combination therapies based on artemisinin and its derivatives are currently the recommended first line treatments for malaria. In the past, resistance to antimalarials emerged in Southeast Asia and then spread to Africa. Artemisinin-resistant *P. falciparum* parasites have now emerged and spread across the Greater Mekong Subregion of Southeast Asia and although artemisinin-based combination therapies (ACTs) are still highly effective against parasites of African origin, mutations associated with artemisinin resistance have been identified in several African countries, including Rwanda, Uganda, and Tanzania. A recent study showing evidence of delayed parasite clearance in *P. falciparum*-infected patients treated with intravenous artesunate in Uganda has raised fears that the spread of artemisinin resistance across the continent is inevitable. Using an individual-based malaria transmission model, Slater et al. estimated that resistance to both artemisinin and its partner drugs would result in up to 78 million additional cases of malaria on the continent over a 5-year period. This illustrates the importance of introducing or strengthening existing surveillance and monitoring systems to facilitate early detection and response to emerging drug resistance.

Antimalarial drug resistance surveillance can include *in vitro* assessment of drug sensitivity in parasites, *in vivo* studies measuring therapeutic efficacy or treatment failure in patients, and detection of molecular markers associated with resistance to specific drugs. Therapeutic efficacy studies, the gold standard, are time- and resource-intensive, limiting their implementation as part of routine malaria control programmes, and *in vitro* assays lack standardization. Molecular assays, on the other hand, are easy to scale up and can be used to rapidly assemble point prevalence of drug resistant parasites in space and time. However, molecular markers (or even the genes involved) are not always known at the point of emergence of resistance. In the past, molecular markers of chloroquine and sulfadoxine-pyrimethamine resistance were confirmed only when resistance had risen to levels too high for the markers to be useful in guiding treatment policies. However, the emergence of artemisinin resistance and the associated mutations in the *Kelch13* gene were identified at a much earlier stage. A unique population sub-structure was also linked to the selection of these parasites, implying that monitoring the parasite population structure over time may allow early warning of the emergence of resistance, even before a definite molecular marker is identified. Multiple studies have analysed the population structure of *P. falciparum* parasites in Africa, using both limited numbers of single nucleotide polymorphisms (SNPs) and whole genome sequence data. However, the ideal number, or identity, of SNPs required to pick up important changes such as emerging drug resistance in these parasite populations is currently unclear.

Here, we describe the spatial and temporal distribution of antimalarial drug resistance markers in *Plasmodium falciparum* parasites collected in 1994–2018 across Kilifi county, Kenya. We also describe the genetic structure of the parasite population using a set of 101 ‘genetic barcode’ SNPs not directly associated with drug resistance, but which are spaced throughout the genome, and analyse whether the specific temporal and spatial patterns observed in the drug resistance markers can be picked up using these genome-wide distributed barcode SNPs.

Methods

Ethics statement

Ethical approval for this study was obtained from Kenya Medical Research Institute’s Scientific and Ethics Review Unit (SSC no. 2669). Written informed consent for parasite genotyping was obtained from parents or guardians of study participants at the time of sample collection.

Study site and population

Kilifi county lies along Kenya’s southern coast. It is largely rural and the population relies mainly on subsistence farming. Malaria transmission occurs throughout the year, with peaks during and immediately following the two rainy seasons in April–August and October–December. The 25-year period (1994–2018) covered by this study saw well documented changes in malaria transmission intensity. Venous blood samples were collected from 1245 children under 15 years of age who tested positive for malaria. The majority of these children lived within the Kilifi health and demographic surveillance system, an 891 km² area whose set up and operation has been described previously, and their homesteads had been geospatially mapped as part of ongoing long-term epidemiological surveillance.

The samples came from three main sources: 1) children who were either attending the outpatient clinic or were admitted to the inpatient children’s wards at the centrally located Kilifi County Hospital, 2) children under follow up as part of longitudinal cohort studies set up in two sub-locations; Ngarenza and Junju, in 1998 and 2005, respectively, and 3) children recruited into specific research studies investigating various aspects of malaria infections, including a) the role of *P. falciparum* variant antigens expressed on the infected red cell surface in generating protective immunity and in the pathogenesis of clinical malaria, b) the natural history of acquired immunity to malaria with particular reference to responses to *P. falciparum* variant antigens expressed on the infected red cell surface, and c) longitudinal cohort studies of the development of natural immunity to malaria.
Sample processing and DNA extraction

The samples used in this study were previously preserved in glycerol as white blood cell-depleted red blood cell (RBC) pellets and frozen in liquid nitrogen. Parasitized RBC (pRBC) pellets were retrieved from liquid nitrogen and thawed for one minute in a water bath set at 37°C. The thawed cells were then transferred into 50 mL Falcon tubes using Pasteur pipettes and re-hydrated using decreasing concentrations of sodium chloride (NaCl) (Sigma-Aldrich, St. Louis, Missouri, USA, cat no SLBQ5226V); 12% (200 µL), 1.8% (10 µL), and 0.9% (10 µL), under gentle vortexing, with 5 min incubation periods between each step. Samples were spun down using a benchtop centrifuge (Eppendorf, Hamburg, Germany, model no 5810R) set at 1800 rpm and 4°C for 5 min after each addition of 1.8% and 0.9% NaCl. To wash off the salts and glycerol, 10 mL RPMI (Sigma-Aldrich, cat no R0883) was added to each pRBC pellet, the solution centrifuged at 1800 rpm for 5 min and the supernatant discarded. Pellets were then re-suspended in 1 mL phosphate buffered saline (PBS) (Life Technologies, Carlsbad, California, USA, cat no 20012027), transferred into 1.5 mL DNA LoBind tubes (Eppendorf, Hamburg, Germany, cat no 0030108078) and centrifuged at 2000 rpm for 2 min. The supernatant was discarded and the remaining pellet volume recorded. Residual white blood cells were degraded using TURBO™ DNase (Thermo Fisher Scientific, Waltham, MA, USA, cat no AM2238). Briefly, each pRBC pellet was re-suspended in DNase enzyme and DNase buffer solutions equivalent to 0.1X of the pellet volume and 1X PBS equivalent to 0.8X of the pellet volume. The mixture was then incubated for 30 min at 37°C on a heating block set to shake at 550 rpm. To inactivate the DNase enzyme, the pellet was washed twice with 1 mL RPMI media containing 2.5 mM EDTA (Thermo Fisher Scientific, cat no AM9260G) and centrifuged at 1800 rpm for 3 min. The supernatant was discarded and the pellet volume topped up to 200 µL with 1X PBS. DNA was extracted using the QIAamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany, cat no 51306) following the manufacturer’s protocol.

DNA quantification

Total DNA in each sample was quantified on an Invitrogen Qubit 3.0 Fluorometer (Life Technologies, model no Q33216) using a Qubit dsDNA HS (High Sensitivity) Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA, cat no Q32853) according to the manufacturer’s instructions. Multi-species quantitative PCR was used to measure the concentration of human and Plasmodium DNA in each sample using Quantitect SYBR Green PCR Kit (QIAGEN, Hilden, Germany, cat no 204145). This assay targets platelet-derived growth factor receptor beta in the human genome and 18s ribosomal RNA in the Plasmodium genome. Briefly, 7.5 µL of PCR mastermix reagent was combined with 4 µL nuclease-free water and 0.75 µL each of the forward and reverse human (forward: 5’ -TTACACACTTCCGTACTTCATC-3’; reverse: 5’ -GCCAGGATGAAGTTGATTTACACAGTT-3’) and parasite (forward: 5’ -GTAATTTGGATGGATTTACACAGTT-3’; reverse: 5’ -TCACTTACAGGTCTTACTGCAAC-3’) primers. The PCR mixture was vortexed briefly and transferred to 96-well plates. 2 µL of sample DNA, water, or varying concentrations of human and parasite DNA standards (50 ng, 5 ng, 0.5 ng, 0.05 ng or 0.005 ng) were added into separate wells of the 96-well plates. The plates were sealed and DNA amplified under the following conditions: 1 activation cycle (95°C for 10 min), 40 amplification cycles (denaturation at 95°C for 30 sec, annealing at 60°C for 45 sec and extension at 72°C for 60 sec), 1 melting cycle (95°C for 15 sec, 60°C for 1 min, and 95°C for 30 sec), and a holding cycle (60°C for 15 sec). qPCR was conducted on the 7500 Real-Time PCR System (Applied Biosystems, Foster City, USA, model no 4351104).

Samples with at least 5 ng/µL of total DNA and less than 80% human DNA contamination were selected for targeted amplicon sequencing.

Amplicon sequencing and drug resistance and barcode calling

Ten drug resistance-related genes (Pfcrtr, Pfâîbr, Pfâîhsps, Pfmdr1, Pfexo, Pfkelch13, Pfâî, Pfmdr2, Pfpar, and plasmepsin 2/3) and an additional 101 genome-wide distributed biallelic single nucleotide polymorphisms (SNPs) were typed on the Illumina Miseg platform using a standardized, in-house genotyping pipeline developed at the Wellcome Sanger Institute (see https://www.malariagen.net/resource/29 for details of the SPOTmalaria technical notes and detailed methods used to generate the data). The drug resistance markers were typed in 1217 samples, while the 101 genetic barcode SNPs were typed in 1245 samples (including the 1217 samples in which drug resistance markers were typed).

Genotypes at drug resistance-linked loci were called at each of the variant positions typed and the results reported as amino acids. The genotypes of genes carrying multiple SNPs of interest were concatenated to form haplotypes, which were then used to classify samples as resistant or sensitive, based on published genetic SNP markers and copy number variations and following the heuristic classification described at https://www.malariagen.net/sites/default/files/File6_Pf_6_resistance_cla
cification.pdf. Polygenomic samples carrying both sensitive and resistant haplotypes were classified as ‘undetermined’ and excluded from analysis of the spatio-temporal distribution of drug resistance. To reduce the number of polygenomic samples excluded from analysis, the ‘undetermined’ classification was applied on a gene by gene basis, such that samples which were polygenomic at one locus were excluded from analysis of that locus, but were included in the analysis of other loci where only one allele was detectable.

Variants at the 101 genome-wide SNP positions were concatenated to generate a genetic barcode unique to each parasite. Samples and SNPs with more than 10% genotyping failure rates were excluded from further analysis.

Identification of mixed species infections

Plasmodium species co-infections were detected by determining the parasite’s mitochondrial genome and differentiating the five Plasmodium species that infect humans (P. falci
paum, P. vivax, P. ovale, P. malariae, and P. knowlesi) in a single reaction.
P. ovale, P. malariae, P. vivax, and P. knowlesi) based on fixed SNP differences across the species. Detailed information on the sequencing protocol can be found at https://www.malariagen.net/resource/29.

Data analysis
All analyses were conducted in R statistical software (v4.0.2). The specific packages used are mentioned under the individual analyses.

Analysis of antimalarial drug resistant haplotypes. The numbers of parasites carrying haplotypes associated with resistance to piperaquine, artemisinin, chloroquine, sulfadoxine and pyrimethamine were determined. The numbers of parasites carrying markers of resistance to multiple antimalarial drugs were computed using the upset() function in the UpSetR package. Loci in four genes shown to favour the emergence of artemisinin resistance in Southeast Asian parasites were also genotyped and their frequencies in the Kilifi parasite population measured.

Annual frequencies of haplotypes associated with resistance to the antimalarial drugs were computed and temporal trends of the frequencies analysed using generalized additive models (GAM) implemented using the mgcv package.

Allele frequency distribution and parasite population structure. Frequencies of 98 genetic barcode SNPs that were successfully typed in more than 90% of the samples were computed to provide an overview of genetic diversity within the Kilifi parasite population. Simple linear regression was used to analyse the statistical significance of temporal variations in the frequencies of these barcode SNPs.

Existence of genetic structure in the parasite population was interrogated using principal component analysis (PCA). PCA was restricted to the 98 genetic barcode SNPs and 1142 samples in which more than 90% of the SNPs were successfully genotyped. For this analysis, unsuccessfully-typed variant positions in individual samples were replaced with the corresponding Pf3D7 reference alleles and PCA was performed using the base R function, prcomp(). The output, principal component (PC) scores, are a set of values representing new, uncorrelated genotypes of individual parasites. Three sets of 1142 PC scores (one for each sample) derived from the first three PCs were used as feature attributes, representing individual parasite genotypes.

Spatial analysis. To determine whether temporal variations in the proportions of parasites carrying antimalarial drug resistant haplotypes have a spatial component, we computed the proportions of antimalarial drug resistant haplotypes in parasites collected in the north and in the south of Kilifi county. Parasites were grouped into two-year bins prior to running the analysis, as some years had very few samples, making yearly analysis unfeasible. The proportions of parasites carrying resistant haplotypes in the north and in the south of the county were then computed for each year bin, and their temporal distributions analysed using generalized additive models.

Spatial autocorrelation was tested using Moran’s I spatial autocorrelation analysis. This metric measures the extent to which parasite pairs that are close to each other in geographic space share genetic similarity compared with parasite pairs that are more geographically distant. Moran’s I spatial autocorrelation coefficients were computed using the corrcov() function in the ncf package. PC scores derived from the first three PCs were used as feature attributes, representing individual parasite genotypes. Latitude and longitude coordinates were used to specify parasite sampling locations. Correlation coefficients were computed for all parasite pairs collected from the same location (i.e. at the same latitude/longitude point) and then within distance bands of 1 km (i.e. 1 km, 2 km, 3 km, etc), 2 km (i.e. 2 km, 4 km, 6 km, etc), and 5 km (i.e. 5 km, 10 km, 15 km, etc). Because most parasites were collected in close proximity to each other, the analysis was restricted to parasite pairs collected up to 20 km apart to reduce potential noise arising from too few parasite pairs sampled far apart. 1000 bootstrap resampling steps were used to determine the statistical significance of the observed correlation coefficients.

Results
We excluded 3.3% (40 of 1217) of antimalarial resistant samples which failed PCR amplification and carried out species identification in the remaining samples. Three samples contained non-Pf species only (1 Pm, 1 Po and 1 Pv/Po), while 14 were mixed species infections (8 Pf/Pm, 5 Pf/Po and 1 Pf/Pv/Pm). The three non-Pf samples and an additional 12 Pf samples that failed genotyping across all drug resistance loci were excluded, leaving 1162 samples (Dataset 1), which were used for subsequent analyses. The temporal distribution of these 1162 samples and the spatial distribution of 1003 samples with geocoded data are shown in Figure 1 and Fig. S1, respectively.

The Kelch13 gene was successfully sequenced in 1146 of the 1162 samples. 1122 of these samples contained clonal parasites, and of these, 97.5% (1094/1122) were wild type artemisinin-sensitive parasites, while the remaining 2.5% (28/1122) were mutants. In total, 24 samples were mixed infections consisting of parasites carrying both wildtype and
mutant alleles. Overall, 10 mutations: I354V, A486S, V487A, S522C, A578S, A582V, S600A, I634L, V666L and A676S, were identified. None of these mutations are currently associated with resistance to artemisinin.

Loci in four genes (arps10 (PF3D7_1460900), fd (PF3D7_1318100), crt (PF3D7_0709000), and mdr2 (PF3D7_1447900)), which form a unique genetic background favouring the emergence of artemisinin resistance in P. falciparum parasites, were successfully genotyped in 1142 samples. Mutations at six amino acid positions (V127M and D128Y/H in arps10, D193Y in fd, N326S and I356T in crt, and T484I in mdr2) were concatenated to form a haplotype representing this genetic background, which was previously identified in Southeast Asian parasites. There was little evidence of these mutations in the Kilifi parasite population, with all but two of the 1142 successfully typed samples carrying the wildtype haplotype (VDDNIT). Of the six amino acids making up this haplotype, only crt N326Y showed any variation (one parasite had a single mutation (VDDSIT), while another had both wildtype and mutant (VDD[N/S]IT) haplotypes).

We did not identify any mutations associated with resistance to the artemisinin partner drug, piperaquine, when we analysed the parasites for variations at amino acid position 415 in the putative exonuclease protein exo (PF3D7_1362500) in 1147 parasites that were successfully genotyped at this position, nor did we identify gene amplification when we analysed the plasmepsin2/plasmepsin3 breakpoint in 651 parasites in which the two genes were successfully typed.

Chloroquine resistance is primarily mediated by mutations in the chloroquine resistance transporter (pfcr, PF3D7_0709000) gene. Chloroquine resistant or sensitive haplotypes were identified by analysing mutations at amino acid positions 72–76 of the chloroquine resistance transporter protein in 1143 samples. Of these samples, 1057 contained monoclonal infections, while 86 were mixed infections containing 2 or more clones. Majority of the monoclonal infections (64%, 676/1057) were parasites carrying the chloroquine-sensitive haplotype, CVMIK, while the remaining 36% (381/1057) carried the chloroquine-resistant haplotype, CVIET (Figure S2A). The proportions of these haplotypes varied over time, with the resistant haplotype being dominant in the early years of the study, up to 1999, and the sensitive haplotype increasing in frequency to become the dominant haplotype from 2009 (Figure 2A). GAM-based analyses of these temporal trends showed a sustained decline in the proportions of parasites carrying the chloroquine resistant haplotype, CVIET (Figure 3A).

Mutations at amino acid positions 51, 59, 108 and 164 in the dihydrofolate reductase (dhfr) protein that confer resistance to pyrimethamine were analysed in 1140 successfully genotyped monoclonal parasites. Only 6.8% (78/1140) of the parasites contained the sensitive haplotype (NCSI) which confers 100% sensitivity to pyrimethamine. The frequency of NCSI declined over time, from a high of 45.9% in 1996 to 1.3% in 2018 (Figure 2B). The degree of resistance to pyrimethamine is often characterized by the number of mutations present in Pfdrfr (i.e. single, double, triple, or quadruple mutations), with higher numbers of mutations associated with greater resistance to pyrimethamine.
Parasites were collected across Kilifi county, Kenya, between 1994 and 2018 and frequencies of drug resistant haplotypes were computed in A) chloroquine resistance transporter (crt), associated with resistance to chloroquine, B) dihydrofolate reductase (dhfr), associated with resistance to pyrimethamine, C) dihydropteroate synthetase (dhps), associated with resistance to sulfadoxine, and D) multidrug resistance protein 1 (mdr1), associated with resistance to multiple antimalarial drugs, including chloroquine, amodiaquine and lumefantrine. Each line represents a specific resistant haplotype. Plots were generated for haplotypes identified in at least 10 parasites. The black line in each plot represents the drug-sensitive haplotype, which has been included for comparison.

to the drug. The triple mutant IRNI (N51I, C59R, S108N) was the most common in this parasite population (55.5%, 633/1140), with temporal analyses showing a rapid increase in its frequency, from 2.7% in 1996 to 81.6% in 2018 (Figure 2B). The double mutants ICNI and NRNI were present in 12.5% (142/1140) and 10.4% (119/1140) of the samples, respectively, and their frequencies remained low over the period studied (Figure 2B). Additional Pfdhfr mutants: IRNL, ICNL and NCNI, were each present at less than 1% frequencies (2/1140, 2/1140, and 6/1140, respectively). In total, 158 samples were mixed infections consisting of parasites carrying haplotypes associated with both sensitivity and resistance to pyrimethamine (Fig. S2B). Analyses of the temporal trends indicated a monotonic increase in the overall proportion of resistant haplotypes over time (Figure 3B).

Mutations at positions 436, 437, 540, 581 and 613 in the dihydropteroate synthetase protein confer varying levels of resistance to sulfadoxine, based on the number of mutations. The sensitive haplotype (SAKAA) was present in 22.1% (252/1142) of the parasites. Temporal analyses of changes in the frequency of this haplotype showed a general decrease over the study period, from a high of 91.7% in 1995, reaching 0% in 2015 before increasing slightly to 3.9% in 2018 (Figure 2C). The double mutant SGEAA (A437G, K540E) was the most prevalent in the population (67.9%, 775/1142), and its frequency increased from 4.3% in 1996 to 94.7% in 2018 (Figure 2C). Five additional mutants (CGEAA, HGEAA, SGEGA, SGKAA and AAKAA) were present at frequencies of less than 2% (Fig. S2C). Temporal analyses using GAM showed an increase in the overall proportion of resistant haplotypes over time (Figure 3C). The quintuple sulfadoxine/pyrimethamine (SP) mutant carrying mutations at positions 51, 59 and 108 in dhfr and positions 437 and 540 in dhps (IRNx + xGExx) is strongly predictive of SP treatment failure. The mutant was identified in 57.5% (539/937) of parasites that were successfully typed.
at all five amino acid positions, with its frequency increasing monotonically over time (Figure 3D).

Mutations at three amino acid positions (86, 184, and 1246) in multidrug resistance protein 1 (pfmdr1, PF3D7_0523000) are associated with resistance to different antimalarial drugs, including mefloquine and lumefantrine, chloroquine, and amodiaquine. The presence or absence of these mutations was determined in 868 parasites. 42.4% (368/868) of the parasites carried the 86Y variant, which enhances chloroquine resistance and mediates responses to amodiaquine, mefloquine and lumefantrine. 30.8% (268/868) of the parasites carried the 184F variant, which is associated with mefloquine and lumefantrine resistance, while 27.2% (236/868) of the parasites carried the 1246Y variant, which is associated with resistance to mefloquine and lumefantrine (Fig. S2D).

Analysis of haplotypes formed by concatenating amino acids at these three positions identified the sensitive haplotype, NYD, in 26.4% of the samples. The numbers of parasites carrying this haplotype varied widely from year to year, although the overall trend was that of increasing frequency over time (Figure 2D). The resistant haplotypes; NFD, YYY, and YYD, were present in 30.0%, 22.8%, and 18.7% of the samples, respectively. Analysis of their temporal trends showed increasing frequencies for NFD and decreasing frequencies for YYY and YYD (Figure 2D).

GAM-based regression tests were conducted for individual variants (86Y, 184F, and 1246Y), as the individual mutations are associated with resistance to different drugs. The results showed temporal declines in the proportions of 86Y (Figure 3E) and 1246Y alleles and a temporal increase in the proportion of the 184F allele (Figure 3F).

Additional mutations in Pfcrt, Pfdrfr, Pfdrps and Pfmdr1 that affect parasite responses to antimalarial drugs in vitro were successfully typed in 1169 parasites (Dataset 2). Most of these mutations were absent in this parasite population, or were present at very low frequencies (< 1%) (Table S1). However, three mutations, A220S, Q271E, and R371I, all in Pfcrt, were present at relatively high frequencies (30% - 35%) that remained constant over time, save for 2017, when the frequency of Q271E reduced to 20% while the frequencies of A220S and R371I increased to 40% (Fig. S3). The role of these mutations in causing drug resistance or enhancing the parasite’s tolerance to antimalarial drugs in vivo is yet to be determined.

To determine whether antimalarial drug-based selective pressures vary across the study site, parasites were grouped based on their sampling locations (Kilifi north vs Kilifi south). Temporal variations in the frequencies of antimalarial drug resistance markers were then computed in the two populations.

**Figure 3.** Generalized additive model-based analyses of temporal variations in frequencies of antimalarial drug resistant haplotypes. The analysis was conducted in Plasmodium falciparum parasites collected in Kilifi, Kenya, between 1994 and 2018. A) Pfcrt mutations at amino acid positions 72-76 (CVIET); B) Pfdrfr mutations identified at amino acid positions 51, 59, 108 and 164 (IRNL, ICNL, NCNI, NRNL, ICNI, and IRNI); C) Pfdrps mutations identified at amino acid positions 436, 437, 540, 581, and 613 (CGEAA, HGEAA, SGEGA, SGKAA, AAKAA, and SGEAA); D) The sulfadoxine-pyrimethamine quintuple mutant, IRNGE, obtained by concatenating amino acids at positions 51, 59, and 108 in Pfdrfr and amino acids at positions 437 and 540 in Pfdrps; E) Pfmdr1_86Y mutant (mutation at amino acid position 86 in the multidrug resistance protein 1); F) Pfmdr1_184F mutant (mutation at amino acid position 184 in the multidrug resistance protein 1); and G) Pfmdr1_1246Y mutant (mutation at amino acid position 1246 in the multidrug resistance protein 1).
Similar trends of increasing or decreasing frequencies in the variations of drug resistance markers were observed over time for most parasites collected in Kilifi north (Figure 4) and Kilifi south (Figure 5). However, different patterns in the distribution of parasites carrying the Pfmdr1 184 mutant (Figure 4F and Figure 5F) were observed.

We measured the prevalence of multidrug resistance in the Kilifi parasite population, focusing on four markers (Pfcrt [amino acids 72–76], PfΔhfr [amino acid positions 51, 59, 108 and 164], PfΔhps [amino acid positions 436, 437, 540, 581 and 613], and Pfmdr1 [amino acid positions 86, 184 and 1246]). Only one parasite in the dataset carried sensitive haplotypes across all four markers analysed, with the remaining parasites carrying at least one mutation associated with antimalarial drug resistance. The upset() function in the UpSetR package was used to compute and visualize the number of parasites carrying different combinations of the drug resistant markers. Majority of the parasites (285), carried drug resistant-associated mutations in the Pfmdr1, PfΔhps, and PfΔhfr genes, while 259 parasites carried drug resistant-associated mutations only in the PfΔhps and PfΔhfr genes. Overall, 138 parasites carried at least one antimalarial drug resistant-associated mutation in each of the four genes (Figure 6).

In total, 101 SNPs unrelated to drug resistance and distributed across the P. falciparum genome were genotyped in 1245 samples, including the 1162 samples in which drug resistance markers were genotyped (Dataset 3). Variants at these positions were concatenated to form a genetic SNP barcode. Details of these SNPs, including their chromosomal locations, reference and alternative alleles, and gene IDs can be found in Supplementary File 1 under the ‘Content of the data release’ at https://www.malariagen.net/resource/29. Three SNPs (P3D7_05_v3_350933, P3D7_04_v3_1102392 and P3D7_08_v3_701557) and 103 samples with more than 10% genotyping failure rates were excluded from analysis, leaving 98 SNPs and 1142 samples which were included in subsequent analyses. Analysis of the overall frequencies of the 98 SNPs showed that most were common in this parasite population, ranging in frequency from 2.3%–92.9% (Figure 7).

The statistical significance of the temporal variations in the frequencies of these SNPs were computed using simple linear regression. Most of the temporal trends were random (p> 0.05) and only 19 SNPs showed statistically significant patterns in their temporal distributions (Fig. S4). Histograms of p-values generated from the regression analyses revealed an ‘anti-conservative’ distribution, with a skew towards low p values (Fig. S5).

Figure 4. Temporal variations in the proportions of antimalarial drug resistant haplotypes in Kilifi north, Kenya. The analysis was conducted in Plasmodium falciparum parasites collected between 1994 and 2018. Plots show the distributions of A) Pfcrt mutations identified at amino acid positions 72-76 (CVIET); B) PfΔhfr mutations identified at amino acid positions 51, 59,108 and 164 (IRNL, ICNL, NCNI, NRNI, ICNI, and IRNI); C) PfΔhps mutations identified at amino acid positions 436, 437, 540, 581, and 613 (CGEAA, HGEAA, SGEGA, SGKAA, AAKAA, and SGEAA); D) The sulfadoxine-pyrimethamine quintuple mutant, IRNGE, obtained by concatenating amino acids at positions 51, 59, and 108 in PfΔhfr and amino acids at positions 437 and 540 in PfΔhps; E) Pfmdr1 86Y mutant (mutation at position 86 in the multidrug resistance protein 1); F) Pfmdr1 184F mutant (mutation at position 184 in the multidrug resistance protein 1) and G) Pfmdr1 1246Y mutant (mutation at position 1246 in the multidrug resistance protein 1).
Figure 5. Temporal variations in the proportions of antimalarial drug resistant haplotypes in Kilifi south, Kenya. The analysis was conducted in Plasmodium falciparum parasites collected between 1994 and 2018. Plots show the distributions of A) Pfcrt mutations identified at amino acid positions 72-76 (CVIET); B) Pfdrfr mutations identified at amino acid positions 51, 59, 108, and 164 (IRNL, ICNL, NCNI, NRNI, ICNI, and IRNI); C) Pfdrfr mutations identified at amino acid positions 436, 437, 540, 581, and 613 (CGEAA, HGEAA, SGEAA, SGKAA, AAKAA, and SGEAA); D) The sulfadoxine-pyrimethamine quintuple mutant, IRNGE, obtained by concatenating amino acids at positions 51, 59, and 108 in Pfdrfr and amino acids at positions 437 and 540 in Pfdrfr; E) Pfmdr1_86Y mutant (mutation at position 86 in the multidrug resistance protein 1); F) Pfmdr1_184F mutant (mutation at position 184 in the multidrug resistance protein 1); G) Pfmdr1_1246Y mutant (mutation at position 1246 in the multidrug resistance protein 1).

However, only 1 SNP (Pf3D7_07_V3_635985, Fig. S4f) remained statistically significant after Bonferroni correction for multiple testing, indicating a non-random distribution of this variant over time. This SNP is found in a gene (PF3D7_0713900) encoding a conserved Plasmodium protein with unknown function and is located on chromosome 7, approximately 200 Mb downstream of Pfcrt.

Genetic structure of the parasite population was interrogated using principal component analysis. Since conventional PCA cannot be computed on a dataset containing missing data, the 10% SNPs that were unsuccessfully genotyped were replaced with the corresponding alleles in the Pf3D7 reference genome. PCA plots for the first three principal components showed no discernible structure in the parasite population, and the parasites could not be resolved in coordinate space (Figure 8).

Mapping parasite genotypes against their geographical sampling locations showed no distinct spatial clusters of parasites based on genotypes (Fig. S6), and temporal analyses of genotypes showed no clear changes in the parasite population over the study period (Fig. S7).

Because PCA did not identify distinct parasite sub-populations, Moran’s I spatial autocorrelation analysis was used to determine whether population structure could be identified in the presence of spatial data. PC scores were used to represent parasite genotypes, while latitude and longitude coordinates were used to specify sampling locations. Autocorrelation coefficients were then computed for parasite pairs separated by different distances. Positive correlations were observed between parasite pairs collected in close proximity, with little or no autocorrelation between parasite pairs collected more than 5 km apart (Fig. S8).

Finally, to determine whether temporal changes similar to those observed in the parasite population using antimalarial drug resistant markers can be detected using the unrelated set of 98 genome-wide distributed barcode SNPs, correlations between yearly median PC scores (i.e. the ‘median’ genotype representing all parasites collected from a single year) computed using these barcode SNPs and time were computed and their temporal trends compared with results obtained from the temporal analysis of drug resistant markers. Unlike the results from the analysis of drug resistant markers, no clear temporal trends representing shifts in the parasite population were observed using the genetic barcode. Instead, median PC scores oscillated around zero, indicating that parasite population structure did not change significantly from year to year (Figure 9).
Figure 6. Prevalence of markers of resistance to multiple antimalarial drugs in a Kilifi *Plasmodium falciparum* population. The intersection size bar plots represent the number of parasites carrying different combinations of antimalarial drug resistance-associated mutations in the four genes (*crt*, *mdr1*, *dhps* and *dhfr*). The lines and circles below the bars represent the different combinations of resistant haplotypes in individual parasites. The set size represents the number of samples in which individual drug resistance-conferring markers were typed. CRT, chloroquine resistance transporter; MDR1, multidrug resistance protein 1; DHPS, dihydropteroate synthetase; DHFR, dihydrofolate reductase.

Figure 7. Distribution of non-reference allele frequencies in a *Plasmodium falciparum* population from Kilifi, Kenya. Frequencies were computed at 98 variable sites across the genomes of 1142 parasites collected between 1994 and 2018. Single nucleotide polymorphisms (SNPs) are labelled based on their chromosomal locations in the Pf3D7 reference genome.
Figure 8. Principal component analysis of 98 genome-wide distributed single nucleotide polymorphisms. The analysis was conducted in 1142 *Plasmodium falciparum* parasites sampled from Kilifi, Kenya, between 1994 and 2018. Each dot represents an individual parasite. Plots were generated to identify population sub-structure based on the first 3 principal components (PCs), with parasites coloured based on their geographical location of origin (kilifi north vs kilifi south vs unknown). The first principal component (PC1) vs the second principal component (PC2), PC1 vs the third principal component (PC3), and PC2 vs PC3.

Figure 9. Changes in the population structure of *Plasmodium falciparum* parasites over time. The samples were collected in Kilifi, Kenya, between 1994 and 2018. Principal component analysis was used to generate principal component (PC) scores representing the genotypes of individual parasites. A median PC score representing the ‘median genotype’ in each year was then computed and the score plotted against the year of parasite sampling. 10,000 resampling steps were used to compute confidence intervals around the median. PC scores varied slightly based on whether they were computed using information from a) the first principal component, b) the second principal component, or c) the third principal component.
Discussion

Molecular surveillance allows rapid and convenient identification and tracking of mutations conferring resistance to antimalarial drugs, enabling timely implementation of control measures. In this study, we analysed the temporal and spatial distribution of mutations associated with antimalarial drug resistance in 1162 Plasmodium falciparum parasites collected over a 25-year period across Kilifi county, Kenya. This represents the largest longitudinal dataset of antimalarial drug resistance markers from a single geographical region in Africa, and provides an invaluable resource for studying the evolution of P. falciparum parasites in response to changing national malaria treatment guidelines, as well as systematic analysis of multiple loci associated with drug resistance.

The predominant PfKelch13 haplotype in the Kilifi parasite population was the wildtype, artemisinin-sensitive haplotype. Although 10 mutations were identified in this population, they are not among those currently validated or associated with artemisinin resistance\(^1\). The most frequent mutation, A578S, is also the most predominant Kelch13 mutation in P. falciparum parasites of African origin\(^2\). Previous genetic studies of other Kenyan P. falciparum parasite populations have also not identified resistance-conferring mutations in Kelch13\(^3,4\). However, the mutation R561H, which is associated with artemisinin resistance, has been identified in Tanzania\(^5\) and Rwanda\(^6\). More recently, C469Y and A675V mutations were not only associated with delayed parasite clearance in a Ugandan population, but were also observed to have increased in prevalence over the five-year study period, from 2015 to 2019\(^7\). These findings underscore the urgent need for ongoing surveillance and monitoring for artemisinin drug resistance in Kenya, which shares a porous border with Uganda. We found no evidence of piperquine resistance in the Kilifi parasite population, which is unsurprising, as although dihydroartemisinin-piperaquine is the second line antimalarial drug in Kenya, it is not widely used and there is therefore little selective pressure associated with it.

We observed significant changes in the prevalence of chloroquine and sulfadoxine-pyrimethamine (SP) drug resistance markers. CVIET, the marker of chloroquine drug resistance, was present in 75% of the parasites sampled in 1994, the earliest year when parasites in this study were collected. In 1999, there was a change in the national malaria treatment guidelines, and chloroquine was replaced with SP. Following the withdrawal of chloroquine, the frequency of the CVIET haplotype declined, while the frequency of the chloroquine-sensitive haplotype, CVMNK, increased. This pattern of reduction in chloroquine-resistance conferring mutations following the withdrawal of the drug is replicated in other regions of eastern Africa\(^8\), although specific alleles associated with chloroquine resistance, including Pfcrt K76T and Pfmdr1 N86Y, are still relatively widespread in parts of western Africa and Southeast Asia\(^9\).

As resistance to chloroquine declined, resistance to sulfadoxine-pyrimethamine increased. Detailed temporal analyses showed high frequencies of resistance-conferring mutations in dhps and dhfr throughout the study period. In the case of dhfr mutations associated with drug resistance were present at high frequencies years prior to the introduction of pyrimethamine (62.2% in 1994) and remained high years after SP was replaced with artemether-lumefantrine as the first line treatment (98.7% in 2018). In comparison, dhps mutations associated with sulfadoxine resistance were present in 29.5% of the parasites in 1994, and increased steadily up to and following the withdrawal of SP as the first line treatment for uncomplicated malaria (96.1% in 2018). Similar results have been reported in other parasite populations\(^10,11\). The presence of drug resistant mutations at high frequencies prior to the introduction of SP may explain why this drug became ineffective so soon after its introduction, necessitating its replacement with ACTs. Notably, the quintuple SP mutant, IRNGE, which is strongly predictive of SP treatment failure\(^12\), was present in more than half of the parasites in this population. Temporal analysis showed an overall increase in the prevalence of this mutant haplotype, from 16.7% in 1994 to 92.8% in 2017. These trends illustrate not only the clear patterns of shifts in the resistant parasite population in response to changing treatment policies, but also show the rate at which these resistant parasites increase and become dominant, rendering available treatments ineffective.

We further identified mutations in multidrug resistance protein 1 (mdr1) gene and analysed the prevalence and temporal distribution of these mutations. Mutations in Pfmdr1 are associated with resistance to different antimalarial drugs, including mefloquine and lumefantrine\(^13,14\), chloroquine\(^15\), and amodiaquine\(^16,17\). We observed declining frequencies of parasites carrying the Pfmdr1 86Y and 1246Y mutations, and increasing frequencies of parasites carrying the Pfmdr1 184F mutation in the Kilifi P. falciparum population. Our findings concur with those of Okell et al\(^18\), who conducted a meta-analysis of studies reporting the frequencies of the different Pfmdr1 alleles in 30 African countries and found that the 86Y and 1246Y alleles declined over time following the introduction of ACT, with a more rapid decline in countries where artemether-lumefantrine (AL) was the first line antimalarial. Additional analyses showed increasing frequencies of parasites carrying the NFD haplotype and decreasing frequencies of parasites carrying the YYY and YD haplotypes. NFD is associated with reduced sensitivity to AL\(^19\). Although a recent clinical trial by Hamaluba et al\(^20\) showed that P. falciparum parasites in Kilifi are still sensitive to AL, temporal increases in the number of parasites carrying the NFD haplotype is worrying. Therefore, close surveillance and monitoring of this parasite population is required to ensure that reduction in AL sensitivity that is potentially attributed to this mutation is detected early.

To analyse the effects of geographic space on the temporal distribution of drug resistant markers, we separated the parasites...
based on their location of collection; in either the north or the south of Kilifi, and analysed the data subsets separately to identify changes in drug resistant markers over time in the two locations. Kilifi north and Kilifi south are separated by the Kilifi creek, which is a potential geographical barrier to human movement. Our analyses showed no differences in the temporal distribution of most of the antimalarial drug resistant markers between parasite populations sampled from the north and the south of the creek, indicating that there is free gene flow across the study area, with little or no spatial barriers to movement between the north and the south. Movement between the two regions is likely facilitated by a bridge connecting the two locations, which allows humans, and hence malaria parasites, to move freely. Some differences were observed in the temporal distribution of the Pﬂmdr1 184F mutant. The reason for these differences is currently unclear, and further investigation, including in a larger number of samples, is required.

Our findings have implications both for the spread of drug resistance and the effectiveness of control measures implemented in the region, because on one hand, drug resistance-conferring mutations arising in one region are likely to spread unimpeded throughout the county, but on the other, control interventions targeted to one region are likely to affect parasites beyond the boundaries of the targeted area, hence benefitting the wider community.

We used a set of 98 genome-wide distributed genetic barcode SNPs to study parasite population structure and to analyse spatial autocorrelation between parasites. We found no discernible structure in the parasite population based on principal component analysis of these SNPs, and parasites could not be separated in co-ordinate space. However, using the same SNPs, we identified clusters of genetically related parasites across the study site based on Moran’s I statistic. Using this metric, we showed that parasites separated by short distances (≤ 5 km) were more similar to each other at the genetic level, corroborating findings that we previously observed using a different group of parasites sampled from the same population25. Despite this evidence that this genetic barcode can detect spatial autocorrelation over short distances, it did not reveal any major differences in parasite population structure associated with the striking changes in drug resistance profiles over the period studied. This is in contrast to the very strong fluctuations in parasite population structure that were associated with the early emergence of artemisinin resistance in Southeast Asia22. However, this observation should be interpreted with caution, as a SNP barcode is a relatively insensitive tool for detecting population structure, particularly at higher levels of transmission intensity that act to increase the population recombination rate, and an important area for further work is to examine this question using whole genome sequence data.

In conclusion, we have presented data from the largest longitudinal study of malaria parasites from a single geographical location to date. Using this dataset, we have confirmed the marked decline in chloroquine resistance and shown the predominance of the Pfdhfr-Pfdhps quintuple marker associated with resistance to sulfadoxine-pyrimethamine. We have also identified several mutations in Kelch13, although these remained at very low frequencies throughout the study period and are not currently associated with artemisinin resistance. However, we observed an increase in the frequency of the Pﬂmdr1 NFD haplotype that is associated with increased tolerance to lumefantrine. Furthermore, we noted similar trends in temporal variations in drug resistance markers in parasites collected in the north and the south of the study area, pointing to uninterrupted gene flow across the region.

Finally, concomitant analysis of a genetic barcode composed of 98 genome-wide distributed SNPs not associated with drug resistance showed no clear population structure, weak spatial autocorrelation and no temporal changes in the parasite population. Although retrospective datasets such as the one analysed here allow a robust analysis of the evolution of drug resistance, impact-based molecular surveillance requires the ability to detect novel changes in the parasite population, including drug resistance markers emerging in response to new control interventions. This requires prospective studies and scalable technologies such as selective whole genome amplification of parasite DNA extracted from dried blood spots. Using whole genome sequence data instead of SNP subsets may also provide the resolution needed to detect haplotypic signatures of selection or population structure due to founder effects. These will form the basis of future research in this field.

Data availability
Underlying data


Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).
Spatial location data has been excluded for confidentiality reasons. The data can be made available upon submission of a formal request through the KWTRP Data Governance Committee by emailing dgc@kemri-wellcome.org. Data access guidelines can be found on the KEMRI Wellcome Trust data repository (https://dataverse.harvard.edu/dataverse/kwtrp).

Extended data

This collection contains the following extended data:
- Figs S1 to S8
- Table S1

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements
The paper is published with the permission of the director of KEMRI.

References
