



OPEN

Population dynamics and drug resistance mutations in *Plasmodium falciparum* on the Bijagós Archipelago, Guinea-Bissau

Sophie Moss¹✉, Emilia Mańko¹, Hristina Vasileva¹, Eunice Teixeira Da Silva^{1,2}, Adriana Goncalves¹, Ashley Osborne¹, Jody Phelan¹, Amabelia Rodrigues³, Paulo Djata⁴, Umberto D'Alessandro⁵, David Mabey¹, Sanjeev Krishna^{6,7,8}, Anna Last^{1,10}, Taane G. Clark^{1,9,10} & Susana Campino^{1,10}✉

Following integrated malaria control interventions, malaria burden on the Bijagós Archipelago has significantly decreased. Understanding the genomic diversity of circulating *Plasmodium falciparum* malaria parasites can assist infection control, through identifying drug resistance mutations and characterising the complexity of population structure. This study presents the first whole genome sequence data for *P. falciparum* isolates from the Bijagós Archipelago. Amplified DNA from *P. falciparum* isolates sourced from dried blood spot samples of 15 asymptomatic malaria cases were sequenced. Using 1.3 million SNPs characterised across 795 African *P. falciparum* isolates, population structure analyses revealed that isolates from the archipelago cluster with samples from mainland West Africa and appear closely related to mainland populations; without forming a separate phylogenetic cluster. This study characterises SNPs associated with antimalarial drug resistance on the archipelago. We observed fixation of the PfDHFR mutations N51I and S108N, associated with resistance to sulphadoxine-pyrimethamine, and the continued presence of PfCRT K76T, associated with chloroquine resistance. These data have relevance for infection control and drug resistance surveillance; particularly considering expected increases in antimalarial drug use following updated WHO recommendations, and the recent implementation of seasonal malaria chemoprevention and mass drug administration in the region.

Despite knowing the causes of malaria for over a century, this disease is still a major public health problem. The World Health Organization (WHO) Global Technical Strategy aims to reduce the global malaria burden by 90% by 2030¹. Considerable progress has been made. However, the decline in malaria mortality has slowed since 2014, and malaria deaths increased to an estimated 627,000 in 2020. This was a 12% rise from 2019, with most deaths caused by the parasite *Plasmodium falciparum*¹. To meet the WHO target by 2030, novel strategies and tools for malaria elimination are needed.

Recently updated WHO recommendations suggest that antimalarials will be increasingly deployed for the control and elimination of *P. falciparum*, on a mass scale. These recommendations included expanding the use of intermittent preventive treatment in pregnancy with sulphadoxine-pyrimethamine (IPTp-SP) to all pregnant

¹Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. ²Ministry of Public Health, Bissau, Guinea-Bissau. ³Bandim Health Project, Bissau, Guinea-Bissau. ⁴National Malaria Control Programme, Ministry of Public Health, Bissau, Guinea-Bissau. ⁵MRC the Gambia Unit, Fajara, The Gambia. ⁶Clinical Academic Group, Institute for Infection and Immunity, and St. George's University Hospitals NHS Foundation Trust, St. George's University of London, London, UK. ⁷Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon. ⁸Institut Für Tropenmedizin Universitätsklinikum Tübingen, Tübingen, Germany. ⁹Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. ¹⁰These authors contributed equally: Anna Last, Taane G. Clark and Susana Campino. ✉email: sophie.moss@lshtm.ac.uk; susana.campino@lshtm.ac.uk

women, broader use of perennial malaria chemoprevention (PMC) and seasonal malaria chemoprevention (SMC), and expanded guidance for the use of mass drug administration (MDA)². As a result of these recommendations, the mass use of antimalarials is likely to increase in a variety of transmission settings at the discretion of relevant control programmes. Molecular surveillance is a useful tool for evaluating the impact of mass antimalarial drug use through monitoring malaria prevalence, parasite population dynamics, and genomic markers of drug resistance.

Endemic island communities are a unique landscape within which to test malaria interventions, including MDA, due to their isolated geography. Understanding the population dynamics and genetic variation of *P. falciparum* populations on islands provides crucial information for the evaluation of these interventions, as well as routine malaria surveillance. Improved whole genome sequencing technology, including the development of Next Generation Sequencing (NGS), has dramatically reduced the time and cost of genomic sequencing through enabling the sequencing of millions of reads in parallel³. The subsequent increase in the generation and availability of genome sequence data allows for high resolution analysis of the genomic architecture of *P. falciparum* populations.

The Bijagós Archipelago is a group of 88 islands and islets situated off the coast of Guinea-Bissau, West Africa. The islands are isolated from mainland Guinea-Bissau; separated by 70 km of sea and accessible by ferry, which takes four hours and sails once each week. Currently, malaria interventions in the Bijagós include enhanced case finding, intermittent preventive treatment in pregnancy (IPTp) and long-lasting insecticidal nets (LLINs)⁴. Despite good coverage and adherence to these interventions, including maximal control efforts using LLINs, malaria case incidence and mortality both increased by between 5 and 25%, from 2015 to 2020¹. Although *P. falciparum* populations have been studied in mainland Guinea-Bissau, there has been no previous research conducted on the genomic architecture of *P. falciparum* parasites on the Bijagós Archipelago. This is vital for researchers to be able to understand the impact of malaria interventions within the region. Eighteen of the islands within the archipelago are inhabited year-round, with a total population of around 25,000 people⁵. The majority of the population live in forest villages and are largely hunter gatherers with some subsistence farming and fishing⁴. The islands have a rainy season between June and November and a dry season from December to May, with peak malaria transmission occurring at the end of the rainy season in October and November⁶. Malaria prevalence on the Bijagós is seasonal and stable, with low to moderate transmission intensity. Most malaria cases are asymptomatic. A malaria prevalence survey was conducted on the most populated island within the archipelago in August 2017, during the start of the malaria transmission season. This was a cross-sectional survey of the population, which identified the prevalence of *P. falciparum* infection to be 16.9% by qPCR⁵. Persistent circulation of subclinical parasitaemia makes reducing malaria transmission very difficult in this setting. This is because asymptomatic malaria cases are a major driver of malaria transmission, and asymptomatic individuals do not generally seek treatment⁵.

One of the main challenges of whole genome sequencing (WGS) clinical malaria parasite isolates from dried blood spots is obtaining sufficient high-quality parasite DNA from infected individuals. This is particularly challenging when there is low parasitaemia in asymptomatic infections, and when working with dried blood spot (DBS) clinical samples, as the majority of DNA extracted from DBS is human. To address this, selective whole genome amplification (SWGA) can be used to selectively amplify *P. falciparum* DNA directly from DBS samples. This can result in better sequencing coverage of low-density *P. falciparum* infections. This higher sequencing coverage is necessary for robust population genomic analyses⁷. Using the SWGA method, we generated 15 whole genome sequences of *P. falciparum* from 15 dried blood spot samples. These DBS samples were collected from asymptomatic individuals of all ages during a cross sectional malaria prevalence survey conducted on five islands on the archipelago in 2018. This is the first WGS data from *P. falciparum* parasites on these islands. These data provide insights into the genetic variation and population dynamics of the *P. falciparum* parasites, including genomic markers associated with antimalarial drug resistance. The genomic variation within the Bijagós Archipelago is contextualised with approximately 800 isolates from other regions in Africa, using publicly available WGS data⁸. Importantly, these data provide a baseline assessment of genomic variation in 2018, prior to the subsequent implementation of SMC and MDA programmes in the region.

Results

Whole genome sequence data. The filtering steps for *P. falciparum* isolates for inclusion in genomic analyses can be seen in Supplementary Fig. 4. DNA was extracted from 2500 dried blood spots. qPCR was then used to test for the presence of *P. falciparum* in the extracted DNA. This identified 145 samples which were positive for *P. falciparum* with a CT value < 30. *P. falciparum* DNA was successfully amplified using Selective Whole Genome Amplification to sufficient levels for whole genome sequencing in 48 samples. Downstream filtering for high-quality whole genome sequence data resulted in 15 samples being carried forwards for bioinformatic analysis (please refer to the Methods section for additional detail).

A pairwise nucleotide matrix of 1,326,629 filtered SNPs was generated using whole genome sequences from 795 *P. falciparum* isolates, collected from 10 countries across Africa (Supplementary Table 1). This included isolates from Central Africa (Gabon, n = 57), West Africa (Ivory Coast n = 70, The Gambia n = 148, Bijagós Archipelago Guinea-Bissau (n = 15), East Africa (Kenya n = 138, Uganda n = 17), Southeast Africa (Malawi n = 149, Madagascar n = 24), South Central Africa (DRC n = 150) and Horn of Africa (Eritrea n = 27). The F_{WS} metric was used to investigate within-sample diversity. Most samples appeared monoclonal, with “high” F_{WS} estimates of ≥ 0.95 ⁹ (Supplementary Table 2). The samples from the Bijagós had a mean F_{WS} of 0.965, with 86.7% monoclonality ($F_{WS} \geq 0.95$), which was higher than isolates from other countries, including The Gambia and Ivory Coast, which exhibited 67.6% and 61.4% monoclonality respectively. This high F_{WS} score indicates low multiplicity of infection within samples from the Bijagós. This may be due to low transmission intensity overall, or be because

people generally live in small, more isolated communities, and supports the assumption that our population analyses are robust.

***P. falciparum* isolates from the Bijagós Archipelago cluster within West Africa.** Principal component analysis (PCA) was conducted using the previously described pairwise genetic distance matrices to investigate the population structure of the *P. falciparum* isolates. Isolates from the Bijagós Archipelago were found to cluster with samples from West Africa when compared to other regions of Africa (Fig. 1). A SNP based Neighbour-Joining tree was generated to further investigate where the Bijagós *P. falciparum* isolates were situated phylogeographically. The samples from the Bijagós (labelled as Guinea-Bissau in purple), clustered with other West African countries. The samples from the Bijagós, The Gambia and Ivory Coast appear to be mixed

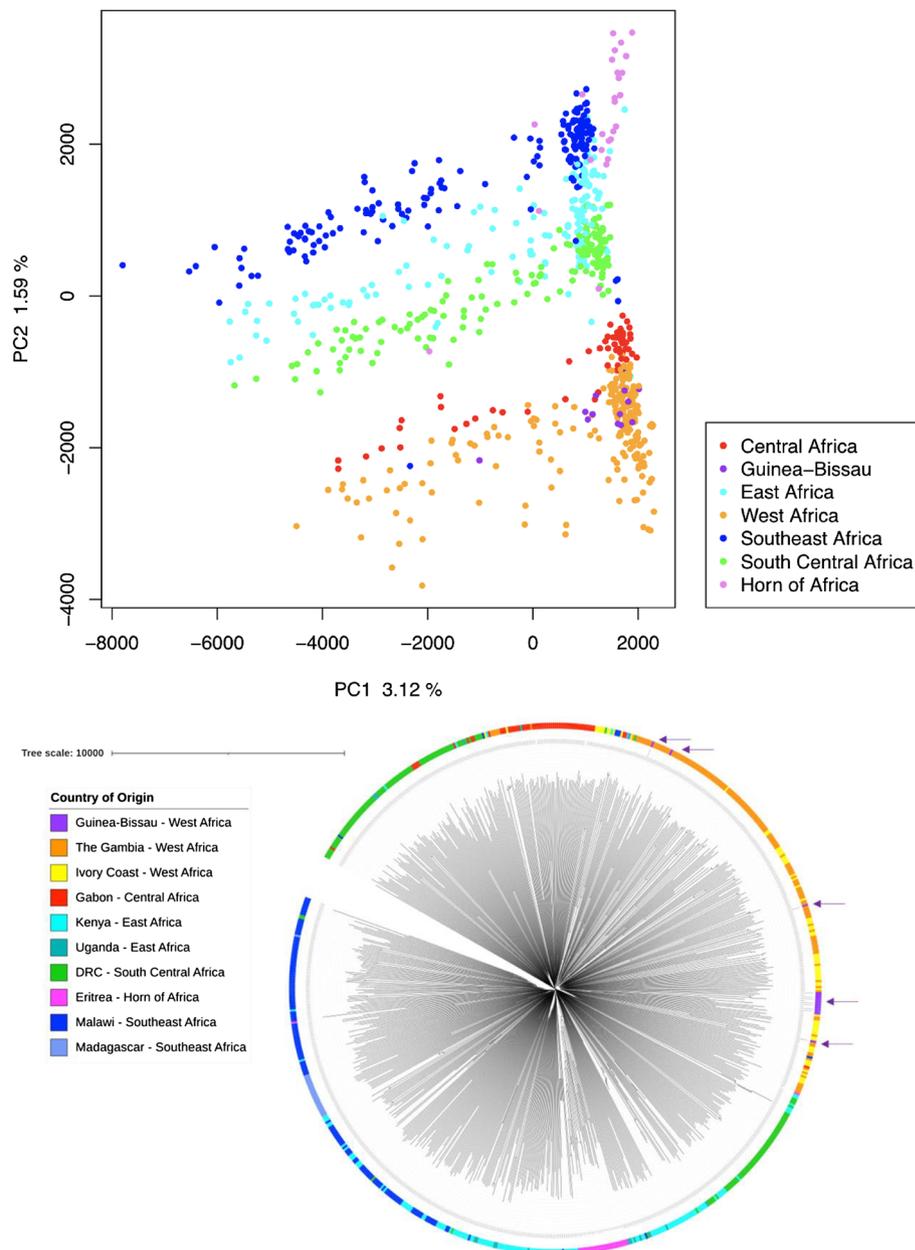


Figure 1. Population structure analysis: (top) Principal component analysis demonstrating that the isolates from the Bijagós (in purple) cluster with other isolates from West Africa (in orange and yellow). Generated using pairwise genetic distance matrices containing 1,326,629 SNPs from 795 isolates. (bottom) Neighbour-Joining tree for the samples and SNPs described above. Isolates were sampled from Central Africa (Gabon, $n = 57$), West Africa (Ivory Coast $n = 70$, The Gambia $n = 148$, Bijagós Archipelago Guinea-Bissau ($n = 15$), East Africa (Kenya $n = 138$, Uganda $n = 17$), Southeast Africa (Malawi $n = 149$, Madagascar $n = 24$), South Central Africa (DRC $n = 150$) and Horn of Africa (Eritrea $n = 27$). Bijagós isolates (indicated with purple arrows) are found interspersed with samples from West Africa (orange and yellow).

and closely related, without forming distinct clusters. This suggests a high degree of mixing between parasites from the Bijagós with those from neighbouring West African countries, reflecting a high degree of regional transmission of *P. falciparum* across West African borders, which has been previously reported¹⁰.

Population differentiation. Population differentiation due to genetic structure was measured using Fixation index (F_{ST}) statistics¹¹, where SNPs with values of 1 indicate perfect differentiation between populations. SNPs that differentiated the Bijagós Archipelago from other African countries or regions were identified. SNPs with a threshold of $F_{ST} > 0.75$, $F_{ST} > 0.85$, and $F_{ST} > 0.9$ were identified to investigate patterns of differentiation between isolates from the Bijagós and the rest of Africa.

The isolates from the Bijagós were compared separately to each of the other countries within the dataset. The number of SNPs with $F_{ST} > 0.75$ for each country comparison are described in Fig. 2. There was an expected trend towards greater SNP differences with increased geographical distance. The largest degree of population differentiation was found between the Bijagós and Eritrea, with 151 SNPs with $F_{ST} > 0.75$.

SNPs with an $F_{ST} > 0.85$ were identified by comparing isolates from the Bijagós with those from other countries in West Africa ($N = 233$), East Africa ($N = 155$), Southeast Africa ($N = 172$), South Central Africa ($N = 150$), and the Horn of Africa ($N = 27$), respectively (Table 1). When comparing the isolates from the Bijagós with samples from other African regions, there were 15 SNPs with $F_{ST} > 0.9$. This included 2 SNPs when comparing isolates from the Bijagós with East Africa, and 13 SNPs when comparing with the Horn of Africa (Table 1). SNPs with $F_{ST} > 0.9$ included those in the rhoptry-associated protein 1, RAP1 (PF3D7_1410400), and reticulocyte binding protein homologue 1 (Pfrh) (PF3D7_0402300). The RAP1 protein is part of the rhoptry-associated protein (RAP) complex, which is important for parasitophorous vacuole membrane structure and intraerythrocytic parasite growth¹². The Pfrh protein is part of the reticulocyte binding-like (RBL) family, binds to erythrocytes and plays a key role in merozoite invasion¹³.

Genetic relatedness. DNA segments that are identical between individuals and have been inherited from a common ancestor without recombination are termed identical by descent (IBD)¹⁴. IBD analysis was conducted to infer pairwise identity by descent between *Plasmodium* isolates within populations. Genomic relatedness was determined by analysing the proportion, or fraction, of the genome that was IBD between the *P. falciparum* isolates. Cumulative IBD across the genome was calculated using sliding window analysis for each population using windows of 10,000 bases (Supplementary Table 3). Isolates from the Bijagós exhibited median IBD fractions of 0.04, ranging between 0.001 and 0.217. This was slightly higher than the other countries in West Africa; Ivory Coast had a median IBD of 0.003, while The Gambia had a median IBD of 0.011. Higher IBD within a population indicates that a higher proportion of alleles at the same locus were inherited from a common ancestor. The higher IBD score for the Bijagós isolates suggests reduced outcrossing within the population, in comparison with other West African countries, which is likely to reflect their isolated island status.

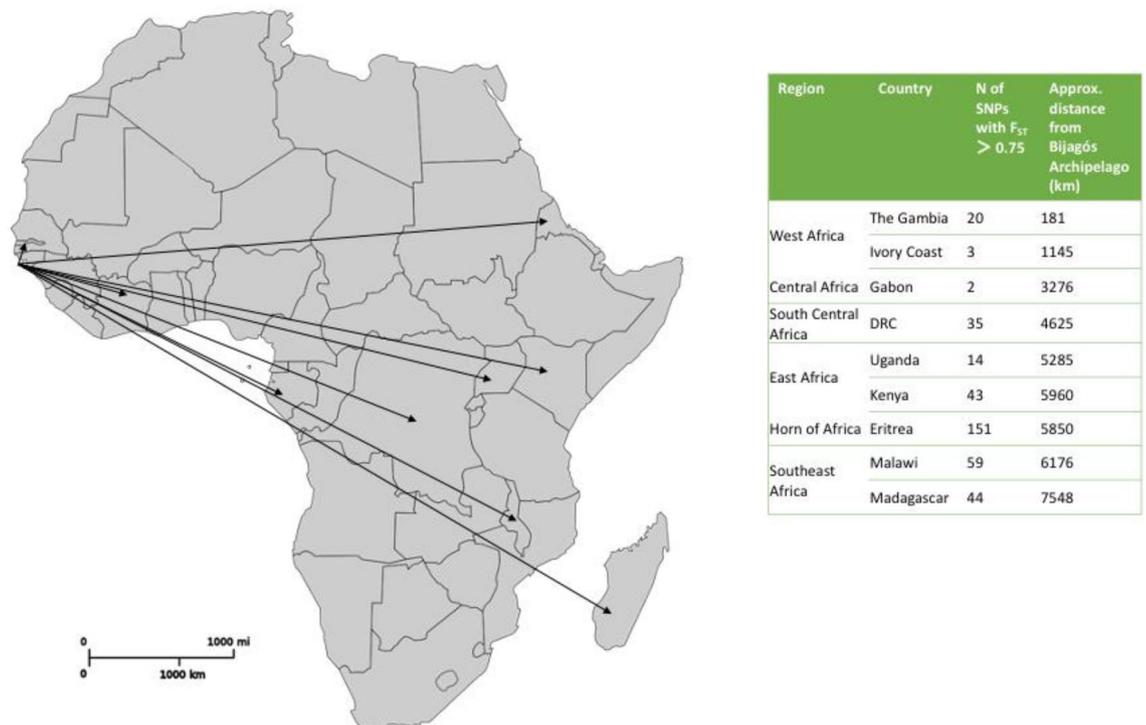


Figure 2. Map showing the proximity of the Bijagós Archipelago to the other countries within this dataset, and the number of SNPs with $F_{ST} > 0.75$ for each country comparison.

Chr	Position	Gene	F _{ST} value	% Alternate allele – Bijagós population	% Alternate allele – Population 2	Population 2
10	331,746	PF3D7_1008100: PHD finger protein PHD1	0.892	50.0	0.5	West Africa
12	1,265,899	PF3D7_1230800: pre-mRNA-splicing regulator WTAP, putative	0.864	40.0	0	West Africa
6	697,009	PF3D7_0616800 malate:quinone oxidoreductase	0.854	0	91.0	East Africa
6	1,110,905	PF3D7_0627700 transportin	0.855	50	0	East Africa
6	1,115,439	PF3D7_0627800 acetyl-CoA synthetase	0.888	64.3	0.6	East Africa
7	732,862	eukaryotic translation initiation factor 3 subunit I, putative	0.922	83.3	0.6	East Africa
9	1,178,903	high molecular weight rhostry protein 2	0.873	92.9	5.2	East Africa
9	1,402,365	PF3D7_0935500 <i>Plasmodium</i> exported protein, unknown function	0.869	73.3	1.6	East Africa
10	331,746	PF3D7_1008100: PHD finger protein PHD1	0.853	50.0	0.7	East Africa
10	658,448	serine/threonine protein kinase, FIKK family	0.865	81.8	3.9	East Africa
11	1,376,369	PF3D7_1135100: protein phosphatase PPM8, putative	0.870	13.3	94.2	East Africa
14	421,792	rhostry-associated protein 1	0.914	21.4	98.7	East Africa
6	1,115,439	PF3D7_0627800 acetyl-CoA synthetase	0.887	64.3	0.6	Southeast Africa
7	732,862	eukaryotic translation initiation factor 3 subunit I, putative	0.893	83.3	1.7	Southeast Africa
10	331,746	PF3D7_1008100: PHD finger protein PHD1	0.899	50.0	0	Southeast Africa
12	1,265,899	PF3D7_1230800: pre-mRNA-splicing regulator WTAP, putative	0.886	40.0	0	Southeast Africa
12	1,265,899	PF3D7_1230800: pre-mRNA-splicing regulator WTAP, putative	0.882	12.5	0	South Central Africa
2	318,919 and 319,674	PF3D7_0207900 serine repeat antigen 2	0.929	0	94.4	Horn of Africa
2	367,271	6-cysteine protein P230p	0.934	0	94.7	Horn of Africa
3	204,230	PF3D7_0304100 inner membrane complex protein 1e, putative	0.865	0	89.5	Horn of Africa
3	729,186	PF3D7_0317700: CPSF (cleavage and polyadenylation specific factor), subunit A, putative	0.859	0	87.5	Horn of Africa
4	138,554 and 138,623	reticulocyte binding protein homologue 1	0.942	0	95.8	Horn of Africa
5	608,869	PF3D7_0514600: ribose-5-phosphate isomerase, putative	0.888	0	91.3	Horn of Africa
5	1,042,252	PF3D7_0525100 acyl-CoA synthetase	0.865	78.6	0	Horn of Africa
7	732,865	eukaryotic translation initiation factor 3 subunit I, putative	0.882	100	8.7	Horn of Africa
7	894,384	PF3D7_0720700: phosphoinositide-binding protein PX1	0.923	100	3.7	Horn of Africa
8	585,854	PF3D7_0811600: conserved protein, unknown function	0.895	0	92.6	Horn of Africa
8	1,031,401	histone acetyltransferase GCN5	0.948	92.9	0	Horn of Africa
8	1,056,829	conserved <i>Plasmodium</i> protein, unknown function	0.948	100	3.7	Horn of Africa
8	1,239,302	PF3D7_0828800: GPI-anchored micronemal antigen	0.879	0	88.9	Horn of Africa
8	1,345,148	<i>Plasmodium</i> exported protein, unknown function	1.000	100	0	Horn of Africa
9	138,846	PF3D7_0801900: lysine-specific histone demethylase, putative	0.877	0	88.5	Horn of Africa
9	778,075	PF3D7_0918900: gamma-glutamylcysteine synthetase	0.870	0	90.5	Horn of Africa
9	1,178,903 and 1,179,069	high molecular weight rhostry protein 2	0.970	92.9	0	Horn of Africa

Continued

Chr	Position	Gene	F _{ST} value	% Alternate allele – Bijagós population	% Alternate allele – Population 2	Population 2
9	1,316,936	PF3D7_0933100: conserved <i>Plasmodium</i> protein, unknown function	0.925	0	92.3	Horn of Africa
9	1,406,576	PF3D7_0935600: gametocytogenesis-implicated protein	0.916	100	4.3	Horn of Africa
9	1,407,002	PF3D7_0935600: gametocytogenesis-implicated protein	0.884	0	90.5	Horn of Africa
9	1,437,289 and 1,438,135	PF3D7_0936300: ring-exported protein 3	0.884	92.3	3.8	Horn of Africa
10	658,448	serine/threonine protein kinase, FIKK family	0.908	81.8	0	Horn of Africa
11	354,078	PF3D7_1108000: IWS1-like protein, putative	0.890	0	91.3	Horn of Africa
11	812,065	PF3D7_1121400: WD repeat-containing protein, putative	0.894	0	92.0	Horn of Africa
11	1,050,233	PF3D7_1126800: alternative splicing factor SR-MG, putative	0.862	0	87.0	Horn of Africa
11	1,376,369	PF3D7_1135100: protein phosphatase PPM8, putative	0.894	13.3	100	Horn of Africa
11	1,507,877	guanylyl cyclase alpha	0.925	0	92.3	Horn of Africa
12	666,002	perforin-like protein 2	0.943	0	95.7	Horn of Africa
12	1,998,804	PF3D7_1248700: conserved protein, unknown function	0.876	0	88.5	Horn of Africa
13	2,107,756	PF3D7_1352700: intron-binding protein aquarius, putative	0.851	0	88.9	Horn of Africa
14	812,013	PF3D7_1419400: conserved <i>Plasmodium</i> membrane protein, unknown function	0.878	0	90.5	Horn of Africa
14	1,227,066	PF3D7_1431200: OST-HTH associated domain protein, putative	0.875	6.7	95.2	Horn of Africa
14	1,989,372	PF3D7_1448500: conserved <i>Plasmodium</i> protein, unknown function	0.888	7.7	96.3	Horn of Africa
14	2,004,493	PF3D7_1449000: gamete egress and sporozoite traversal protein, putative	0.899	13.3	100	Horn of Africa
14	2,340,921	PF3D7_1457100: conserved <i>Plasmodium</i> protein, unknown function	0.876	0	88.5	Horn of Africa
14	2,343,679	PF3D7_1457200: thioredoxin 1	0.879	0	88.9	Horn of Africa
14	2,357,358	PF3D7_1457400: conserved <i>Plasmodium</i> protein, unknown function	0.897	0	92.6	Horn of Africa
14	2,364,543	PF3D7_1457600: conserved <i>Plasmodium</i> protein, unknown function	0.869	0	92.6	Horn of Africa
14	2,364,847	PF3D7_1457600: conserved <i>Plasmodium</i> protein, unknown function	0.854	0	88.9	Horn of Africa
14	2,367,254	PF3D7_1457700: large ribosomal subunit nuclear export factor, putative	0.851	0	88.5	Horn of Africa
14	2,469,318	PF3D7_1460500: conserved <i>Plasmodium</i> protein, unknown function	0.894	0	91.7	Horn of Africa
14	2,535,998	PF3D7_1462400: conserved <i>Plasmodium</i> protein, unknown function	0.899	0	92.6	Horn of Africa
14	2,845,773	PF3D7_1469400: nucleoside transporter 3, putative	0.859	0	87.5	Horn of Africa

Table 1. Genetic markers of population differentiation between the isolates from the Bijagós Archipelago and isolates from countries in the other African regions, with $F_{ST} > 0.85$. SNPs with $F_{ST} > 0.9$ have been highlighted in bold. Only SNPs where all calls in the Bijagós population were either reference, alternate, or mixed, were included. Sample sizes: Bijagós Archipelago $N = 15$, West Africa $N = 233$, East Africa $N = 155$, Southeast Africa $N = 172$, South Central Africa $N = 150$, Horn of Africa $N = 27$.

Genomic regions under recent positive selection. WGS data was analysed for signatures of recent positive selection. Single population iHS and cross-population XP-EHH metrics were calculated, which are statistics derived from extended haplotype homozygosity. Within the isolates from the Bijagós, there were seven SNPs which exhibited significantly high iHS scores, where $(-\log_{10}[1 - 2 | \Phi_{iHS} - 0.5 |]) > 4.0$ (Supplementary Table 4). These SNPs were located within two surface-associated interspersed (SURFIN) pseudogenes, SURFIN 1.2 and SURFIN 13.1. Cross-population analysis was conducted between isolates from the Bijagós and other regional populations within Africa. This analysis identified seven candidate regions under directional selection based on XP-EHH values (Supplementary Table 5). When comparing isolates from the Bijagós with those from all other African regions, there were four common candidate regions under directional selection. These regions encoded cytoadherence linked asexual proteins 3.1 and 3.2, the ABC transporter B family member 4 (putative), and the CX3CL1-binding protein 2. When comparing isolates from the Bijagós to those from Central Africa, South Central Africa, Southeast Africa and West Africa, there were two additional common candidate regions under directional selection. These regions encode the 26S protease regulatory subunit 8 (putative) and a conserved *Plasmodium* protein of unknown function (PF3D7_1248700). When comparing isolates from the Bijagós with isolates from Southeast Africa and West Africa, two additional common candidate regions were identified, which encode merozoite surface protein 3 and methionine tRNA ligase. When comparing with East Africa, one additional candidate region was found, encoding Plasmepsin X. When comparing with South Central Africa, two additional candidate regions were identified, encoding the 26S protease regulatory subunit 6A (putative), and structural maintenance of chromosomes protein 1 (putative) (Supplementary Table 5).

Population admixture. Admixture analysis is a method of investigating the geographical origins of samples, based on their ancestry. Admixture analysis was conducted on 387,395 SNPs using ADMIXTURE software. Using a cross-validation approach, five ancestral populations were inferred (K value = 5). Figure 3 (top) shows these five ancestral populations ($K = 1$ to $K = 5$) as different colours. The figure includes each individual *P. falciparum* isolate along the x-axis, and the ancestral populations of each isolate along the y-axis, represented by colour. This analysis suggested that isolates from the Bijagós contained high proportions of ancestral genome fragments associated with West African *P. falciparum* populations (Fig. 3).

PCA was carried out using the previously described pairwise genetic distance matrix containing 1,326,629 SNPs from 795 isolates. Each *P. falciparum* isolate on the PCA was coloured according to its dominant ancestral population (of $K = 1$ to $K = 5$). This is shown in Fig. 3 (bottom). The overlapping of dominant ancestry reinforced high proportions of shared ancestral genome fragments for the samples across Africa (Fig. 3).

Identification of mutations associated with drug resistance. Mutations associated with resistance to antimalarial drugs were analysed using the WGS data from the Bijagós isolates ($n = 15$) and isolates from West Africa ($n = 218$). Comparison of the prevalence of drug resistance mutations between these populations was carried out to understand similarities and differences between these isolates (Table 2).

The K76T mutation, associated with resistance to chloroquine¹⁵, was found at a slightly lower frequency (28.6%) in isolates from the Bijagós compared to isolates from The Gambia and Ivory Coast in mainland West Africa (51.4%) (non-significant $P > 0.05$). Overall, PfDHFR mutations associated with resistance to pyrimethamine¹⁶ were found at higher prevalence in isolates from the Bijagós than those from mainland West Africa. N51I was found at significantly higher prevalence in the Bijagós (100%) than in mainland West Africa (76.6%) ($P < 0.05$). S108N was found at 100% prevalence in the Bijagós compared to 81.1% in mainland West Africa (non-significant $P > 0.05$). C59R was found at 92.9% prevalence compared to 77.5% in mainland West Africa (non-significant $P > 0.05$). Of the 15 Bijagós samples sequenced, one was a double mutant with N51I-S108N, and 13 were triple mutants with N51I-C59R-S108N. The PfDHPS A437G mutation, associated with resistance to sulphadoxine¹⁷, was found in 60% of isolates from the Bijagós. K540E was not found in isolates from the Bijagós or mainland West Africa. A581G was not found in isolates from the Bijagós but was found at a prevalence of 0.9% in isolates from mainland West Africa (non-significant difference $P > 0.05$). Of the 15 Bijagós samples sequenced, two were double mutant with S436A-A437G.

PfMDR1 mutations N86Y and Y184F, associated with resistance to chloroquine^{18,19}, were found in the Bijagós isolates at frequencies of 7.7% and 53.8%, respectively. The prevalence of N86Y and Y184F was slightly higher in isolates from mainland West Africa, at 15.6% and 60.1%, respectively (non-significant difference $P > 0.05$). The mutation D1246Y, also associated with chloroquine resistance^{18,19}, was not found in isolates from the Bijagós, but was found in 0.9% of isolates from mainland West Africa (non-significant difference $P > 0.05$). The N458Y mutation in PfKELCH13, associated with artemisinin partial resistance²⁰, was not found in isolates from the Bijagós or mainland West Africa. There was not sufficient genomic coverage within the Bijagós sample WGS data of other PfKELCH13 marker positions to report these reliably. Mutations which could not be reported on are listed in Supplementary Table 7. The PfEXO E415G mutation, associated with resistance to piperazine, and the PfAP2μ S160N/T mutation, which may be associated with artemisinin-resistance, were not found in either population²¹.

Discussion

This study presents *Plasmodium falciparum* sequence data from the Bijagós Archipelago, Guinea-Bissau. Investigating the genomic architecture of malaria parasites in endemic countries increases the understanding of parasite genetic variation and population dynamics. This includes the prevalence of genomic mutations associated with drug resistance. Knowledge of drug resistance markers and population dynamics can assist in malaria control and surveillance, which are key to reducing the global malaria burden in line with the World Health Organization (WHO) Global Technical Strategy¹. Genomic surveillance of drug resistance markers is particularly

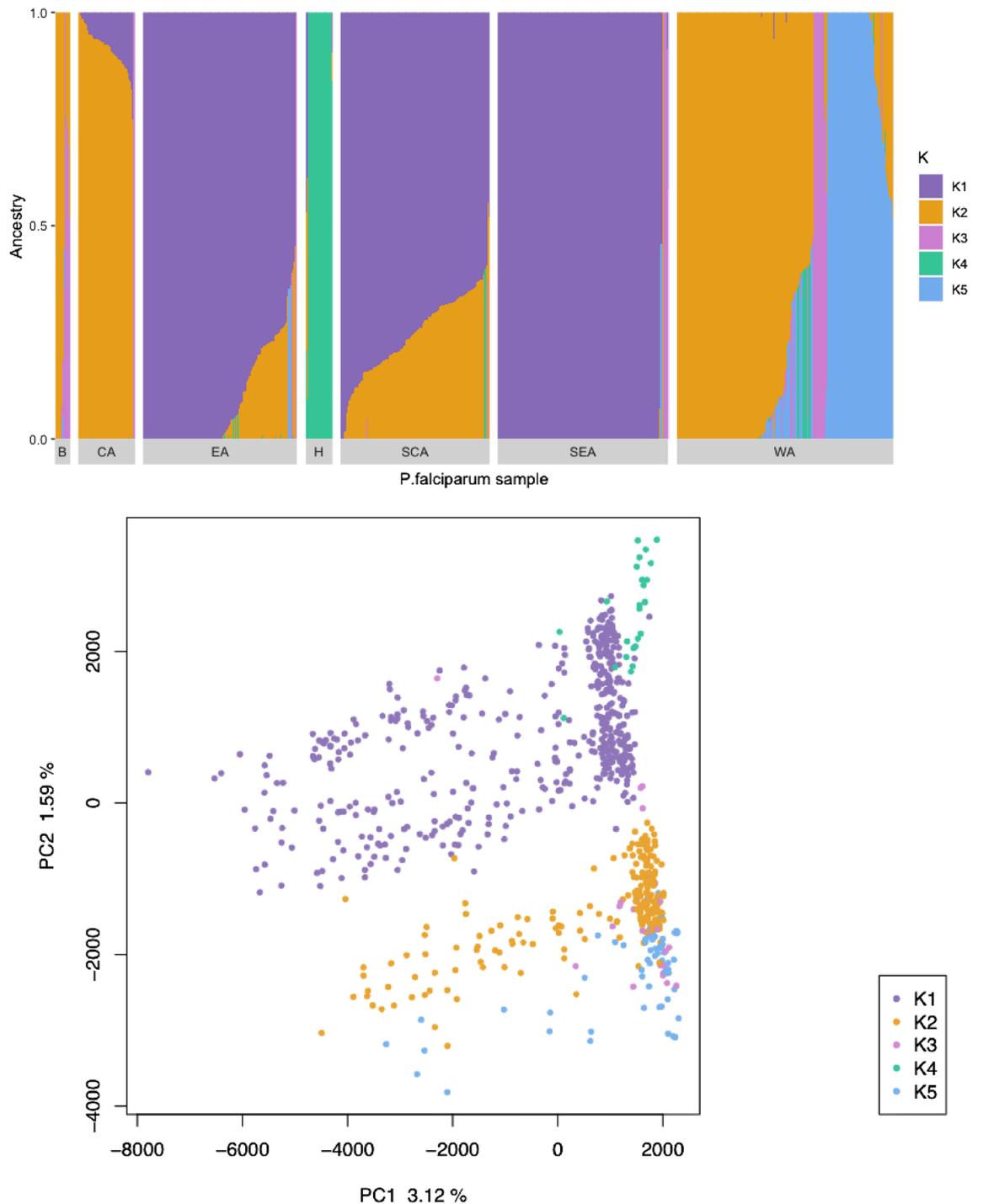


Figure 3. (top) Ancestry analysis of *P. falciparum* isolates from (B) the Bijagós Archipelago, (CA) Central Africa, (EA) East Africa, (H) Horn of Africa, (SCA) South Central Africa, (SEA) Southeast Africa and (WA) West Africa. Individual isolates are plotted along the x-axis, with ancestry coefficients on the y-axis. The number of ancestral populations, K , used in this analysis was $K = 5$. (bottom) Principal component analysis generated using pairwise genetic distance matrices. This is coloured by the maximum K value for each isolate.

important considering the likely increase in the mass use of antimalarials following recently expanded WHO recommendations².

The Bijagós Archipelago is a region with low-moderate malaria transmission intensity, exhibiting stable and seasonal malaria cases. Many malaria cases in the region are asymptomatic, which presents a difficult challenge to control programmes, as asymptomatic cases sustain a reservoir of infections²². Consequently, this reservoir is capable of continually driving malaria transmission, which makes eliminating malaria in this area extremely difficult^{23,24}. Furthermore, drug resistance is hypothesised to spread faster in low transmission settings than

Gene	Position	Mut	Ref allele	Alt allele	n/N mutation Bijagós	n/N mutation W. Africa	% Mut in Bijagós	% Mut in W. Africa	P value (X ²)
<i>pfprt</i>	403,625	K76T	A	C	4/14	112/218	28.6	51.4	0.098
<i>pfalhfr</i>	748,239	N51I	A	T	14/14	167/218	100	76.6	0.040
<i>pfalhfr</i>	748,262	C59R	T	C	13/14	169/218	92.9	77.5	0.158
<i>pfalhfr</i>	748,410	S108N	G	A	14/14	176/218	100	81.1	0.070
<i>pfalpps</i>	549,993	K540E	A	G	0/15	0/218	0	0	-
<i>pfalpps</i>	549,685	A437G	C	G	9/15	162/218	60.0	74.3	0.225
<i>pfalpps</i>	550,117	A581G	C	G	0/15	2/218	0	0.9	0.709
<i>pfalpps</i>	549,681	S436A	T	G/C	3/15	64/218	20.0	29.4	0.439
<i>pfmdr1</i>	958,145	N86Y	A	T	1/13	34/218	7.7	15.6	0.440
<i>pfmdr1</i>	958,440	Y184F	A	T	7/13	131/218	53.8	60.1	0.656
<i>pfmdr1</i>	961,625	D1246Y	G	T	0/13	2/218	0	0.9	0.729
<i>pfkelch13</i>	1,725,626	N458Y	T	A/C	0/14	0/218	0	0	-
<i>pfexo</i>	2,504,560	E415G	A	G	0/13	0/218	0	0	-
<i>pfap2μ</i>	718,433	S160N/T	G	A/C	0/14	12/218	0	5.5	0.367

Table 2. Frequency of mutations associated with resistance to different anti-malarial drugs, from Bijagós *P. falciparum* isolates (n = 15, sample collection in 2018) and *P. falciparum* isolates from West Africa (Total n = 218, Ivory Coast n = 70—sample collection in 2013, The Gambia n = 148—sample collection in 2008 and 2014).

in high transmission settings due to a higher proportion of infections being exposed to antimalarial drugs. This increases the probability of treatment failure and the subsequent transmission of resistant parasites in the population²⁵.

The data analyses presented in this study characterise asymptomatic malaria infections. These infections are more difficult to study due to the low quantities of parasite DNA in clinical samples, particularly in dried blood spots (DBS). This means that despite the importance of understanding asymptomatic cases, most research into genomic variation and drug resistance is conducted on parasites from symptomatic malaria infections. This study made use of Selective Whole Genome Amplification (SWGA) to amplify parasite DNA from isolates collected using DBS, from asymptomatic infections within the Bijagós⁷. These advances in molecular methods, along with improved Next Generation Sequencing technology, allow for feasible routine surveillance of parasite genetic variation and drug resistance markers from dried blood spot samples, which can be collected during routine cross-sectional surveys.

Following successful amplification of low-density asymptomatic *P. falciparum* infections obtained in the Bijagós, population analyses were conducted. The Bijagós *P. falciparum* genomic variation was then contextualised with the wider African *P. falciparum* population. Principal component analysis (PCA) indicated that the malaria samples from the Bijagós cluster with samples from other countries in West Africa, including The Gambia and Ivory Coast. A relatively low percentage of the variance between malaria samples from different countries was explained within the PCA. This, along with the absence of SNPs with $F_{ST} = 1$, reflects the high degree of mixing between *P. falciparum* isolates in the Bijagós Archipelago and those on the mainland. Despite this relatively small sample size of 15 from the Bijagós, these results corroborate the high degree of mixing between *P. falciparum* isolates demonstrated across the African continent¹⁰. The samples from the Bijagós had a mean F_{WS} of 0.965, with 86.7% monoclonality (with $F_{ws} \geq 0.95$). This was the highest percentage of monoclonality when compared with samples from other countries across the African continent. The high percentage of monoclonal infections may be due to people in the Bijagós living in small, relatively isolated, communities, which could reduce the likelihood of simultaneous infection with parasites of different genotypes. This monoclonality may also reflect the low to moderate transmission intensity of *P. falciparum* in the Bijagós. Despite the high degree of population mixing, identity by descent analysis revealed that isolates from the Bijagós had slightly higher IBD fractions than isolates from the other West African countries, which may reflect their more isolated island status. Admixture analysis gave further insight into the ancestral origins of the Bijagós Archipelago isolates amongst African populations, suggesting Bijagós isolates obtained large proportions of their ancestral genome fragments from West African *P. falciparum* lineages.

Fixation index (F_{ST}) analyses were used to measure population differentiation due to genetic structure. F_{ST} values range from 0 to 1, where a value of 0 for a particular SNP indicates that two populations have no difference in allele frequency, whereas a value of 1 indicates complete differentiation in allele frequency between the two populations¹¹. Therefore, a value of 0 would indicate a high degree of population mixing. Whereas, higher F_{ST} values at a particular SNP indicate greater population differentiation. Fifteen SNPs with a high $F_{ST} > 0.9$ were identified when comparing isolates from the Bijagós with those from other African regions. The greatest amount of population differentiation was found between samples from the Bijagós and Eritrea. Interestingly, SNPs in the Bijagós samples were not fixed at either allele for many of the loci with high F_{ST} values, indicating that the population is in transition for many of these alleles. Overall, there were no SNPs with $F_{ST} = 1$ in any of the regional comparisons. This indicates that there were no SNPs specific to the Bijagós parasite population. It

is likely that a larger sample size would be needed to reveal these SNPs with $F_{ST} = 1$, which could then be used to generate a molecular barcode for the *P. falciparum* samples from the Bijagós able to predict geographic origin²⁶.

Regions under selective pressure included seven loci within the surface-associated interspersed (SURFIN) pseudogenes, SURFIN 1.2 and SURFIN 13.1. SURFINS are variable erythrocyte surface antigens and are targets of naturally acquired immunity against malaria²⁷. SURFIN 1.2 has been associated with chloroquine sensitivity²⁷, and it has been suggested that both SURFIN 1.2 and SURFIN 13.1 are involved in cell surface adhesion^{27,28}. However, additional phenotypic experimentation is required to confirm the role of these two pseudogenes. Cross-population analysis identified seven candidate regions under directional selection based on XP-EHH values. These candidate regions included genes encoding Plasmeprin X (PMX), the ABC transporter B family member 4 (ABCB4), and *P. falciparum* merozoite surface protein 3 (PfMSP3). Plasmeprin X is essential for asexual parasite development and is considered a multi-stage antimalarial target²⁹. ABCB4 is a member of the ABC (ATP-binding cassette) B family. This is a group of transport proteins which includes the PfMDR1 protein, and polymorphisms in the ABC transporters have been involved in anti-malarial drug resistance³⁰. PfMSP3 is a merozoite surface protein and is a potential vaccine candidate³¹.

This study reported on genomic markers associated with drug resistance in the Bijagós Archipelago. Anti-malarial treatments administered on the islands include artemether-lumefantrine and IPTp with sulphadoxine-pyrimethamine (SP). Intravenous artesunate or quinine are only used in severe, or complicated, cases. The PfCRT K76T mutation, associated with resistance to chloroquine, had a frequency of 28.6% on the islands, despite chloroquine not being included in malaria treatment policy since 2006 in Guinea-Bissau. This was a slightly lower prevalence than observed in isolates from mainland West Africa, but this difference was not significant ($P > 0.05$). Persistence of the K76T mutation may indicate that it has not yet been lost due to fitness-cost. The PfDHFR mutations associated with resistance to pyrimethamine, N51I and S108N, were found to be fixed (100% prevalence) in the Bijagós isolates, and C59R was found close to fixation at 92.9% prevalence. These mutations were found at a higher prevalence in the Bijagós isolates than in the isolates from mainland West Africa, with the N51I mutation at significantly higher prevalence in the Bijagós ($P = 0.04$). This difference may be partially explained by the collection years of these isolates, as isolates from the Bijagós were collected in 2018, whereas isolates from mainland West Africa were collected earlier, between 2008 and 2014. Therefore, the Bijagós isolates were collected following additional years of selection pressure from the use of SP in the region (Supplementary Table 6). Thirteen of the Bijagós isolates sequenced exhibited the triple mutant N51I-C59R-S108N haplotype. In addition, the PfDHPS A437G and S436A mutations, associated with resistance to sulphadoxine, were found at 60% and 20% prevalence, respectively. Although this is a small sample set of 15 isolates, the presence of mutations associated with resistance to SP warrants further investigation with a larger sample size. Encouragingly, the PfEXO E415G mutation associated with piperazine resistance, and the PfAP2 μ S160N/T mutation, which may be associated with artemisinin-resistance, were not found in the Bijagós isolates. Limited sequencing coverage of the PfKELCH13 gene made investigating the majority of validated or candidate resistance mutations in this gene unreliable. However, there was sufficient coverage of the N458Y genome position to identify reliably that this mutation was absent in the Bijagós isolates. The artemisinin-combination therapy dihydroartemisinin-piperazine (DHA-PPQ) was not used on the islands prior to this study. However, there has since been MDA with DHA-PPQ across the archipelago.

Despite the limitations in generating sequence data from asymptomatic isolates collected using DBS, this study reported high quality whole genome sequence data for 15 Bijagós isolates. This is a small sample size which has limitations for generalisation. However, the population genomic analyses results are consistent with those from neighbouring African countries, indicating that these results are robust. Furthermore, this study further demonstrates the capacity of SWGA methods in generating WGS data from low-density parasite concentrations. This dataset forms the only dataset of WGS data for the Bijagós. Therefore, it forms an important baseline in the context of evaluating subsequent MDA and SMC interventions that have been implemented in the region. The Bijagós WGS data was contextualised with publicly available WGS sequence data from isolates across Africa. Information about the symptomatic or asymptomatic status of these infections is not publicly available, so comparisons concerning patient symptom status cannot be made using this dataset. Additional sequencing of *P. falciparum* isolates is required to reveal a more comprehensive picture of regional dynamics, particularly concerning the prevalence of drug resistance markers. This should ideally include the analysis of samples collected from both asymptomatic and symptomatic infections, and across different years, to understand temporal changes in resistance marker frequency. High throughput, low-cost, methods, such as amplicon sequencing, have the potential to achieve this. In turn, this information can be used to inform programmatic decisions in malaria control, working towards malaria elimination.

Materials and methods

Dried blood spot collection. Dried blood spots were collected onto filter paper from residents of the Bijagós Archipelago during a cross-sectional survey in 2018. Dried blood spots were sampled as part of this mapping survey, regardless of whether the participant had symptoms of malaria.

This study was conducted in accordance with the Declaration of Helsinki and granted ethical approval by the Comité Nacional de Ética de Saúde (Guinea-Bissau) and the London School of Hygiene and Tropical Medicine Research Ethics Committee (UK). Written or thumbprint informed consent was obtained from all study participants.

DNA extraction and qPCR for malaria positivity. Following collection, dried blood spots were punched for DNA extraction using 3 mm diameter punches. Three punches were taken per sample for DNA

extraction. DNA was extracted using the PureLink™ Pro 96 Genomic DNA Purification Kit (ThermoFisher), (cat No: K182104A). DNA extract was quantified for *Plasmodium falciparum* 18S gene using qPCR.

Plasmodium genome amplification and sequencing. DNA extracts which had a qPCR CT value < 30 were tested using the Qubit for DNA concentration. Using this CT threshold, n = 145 samples were selected for further analysis. DNA concentration in the extracted dried blood spot samples was very low, of the range 0.011 ng/μl to 0.268 ng/μl. Selective whole genome amplification (SWGA) was used to amplify the *Plasmodium* DNA prior to sequencing, using a protocol adapted from Oyola et al.⁷. Each SWGA reaction used 30 μl of extracted DNA, which was combined with 5 μl of *phi29* DNA Polymerase Reaction Buffer (New England Biolabs), 0.5 μl of Recombinant Albumin (New England Biolabs), 0.5 μl of SWGA primers (250 mM) (Roche), 5 μl dNTPs (New England Biolabs) and 6 μl of water, to a total volume of 50 μl per reaction. SWGA reactions were then run in a thermocycler using the following programme steps: 35 °C for 5 min, 34 °C for 10 min, 33 °C for 15 min, 32 °C for 20 min, 31 °C for 25 min, 30 °C for 16 h, 65 °C for 10 min, and held at 4 °C. SWGA reactions increased the DNA concentration of samples to > 300 ng DNA in of the 48 of the 145 (33.1%) samples amplified.

Whole genome sequencing. The SWGA product from these samples was purified using KAPA pure beads (Roche). Beads were equilibrated to room temperature and mixed with SWGA product at a 1:1 volume ratio. The KAPA pure beads Genomic DNA Purification protocol steps were then followed (<https://rochesequencing.store.com/wp-content/uploads/2022/07/KAPA-Pure-Beads-Technical-Data-Sheet.pdf>). DNA was then quantified using the Qubit dsDNA HS Kit (Thermo Fisher). Samples contained > 300 ng DNA and were sequenced at Eurofins Genomics using the Illumina Novaseq 6000, producing a minimum of 3.75 M paired reads (250 bp reads) per sample.

Bioinformatics. FastQ files from WGS were used to produce VCF files for each sample. To do this, FastQ files were trimmed using *trimmomatic* software (version 0.39, using the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36) to remove poor quality sequences³². The trimmed data was then aligned to the reference *P. falciparum* 3D7 genome (version 3) using *bwa-mem* software (version 0.7.17-r1188, default parameters), to produce a BAM file for each sample. The *samtools* (version 1.12) functions *fixmate* and *markdup* were applied to the resulting BAM files³³. Following this, the *GATK*'s BaseRecalibrator and ApplyBQSR functions were applied for calibration and correction of the BAM files using the *P. falciparum* genetic crosses 1.0 dataset (<http://www.malariagen.net/data/pf-crosses-1.0>). SNPs and indels were called with *GATK*'s HaplotypeCaller (version 4.1.4.1) using the option -ERC GVCF³⁴. Validated VCFs were imported into GenomicsDB using *GATK*'s function GenomicsDBImport, and a combined VCF was created using *GATK*'s GenotypeGVCFs function. This VCF file contains all the genotype calls for all the variant positions across all samples.

For further analysis, only variants in core regions of the genome were included and assigned a quality score using *GATK*'s Variant Quality Score Recalibration (VQSR). VariantRecalibrator was run using the previously mentioned genetic crosses dataset as a training set with the following parameters used for SNPs: -an QD -an FS -an SOR -an DP -maxGaussians 8 and -mq-cap-for-logit-jitter-transform 70. Following this, *GATK*'s ApplyVQSR was run using the parameter -truth-sensitivity-filter-level 99.0 to obtain a Variant Quality Score Log-Odds (VQSLOD). Variants with a VQSLOD score < 0, representing variants more likely to be false than true, were filtered out.

Additionally, isolates with more than 40% of SNPs missing genotype data (177/2129 isolates) were excluded from downstream analysis to ensure that samples which passed quality control had uniform sequencing coverage across the whole genome. SNPs were then processed to replace the genotype call in the VCF for a mixed call whenever the secondary MAF was at least 20% in a given SNP on each individual sample. A subset of only biallelic SNPs was also obtained from this VCF. The filtered VCF was then converted into a binary matrix consisting of genotypes (0 = reference allele; 1 = mutant allele; 0.5 = mixed allele) across all SNPs and isolates. The final matrix consisted of 1952 isolates and 1,326,629 high-quality SNPs. This bioinformatics pipeline is described in detail by Benavente et al.³⁵. An annotated VCF of all extracted variants was made using *snpEff* software (version 5.1), which annotates the variants³⁶. Fifteen of the Bijagós samples had sufficient quality data for downstream bioinformatic analysis, with over ~ 40% of the genome covered by > 5 reads.

Population genomic analysis. Whole Genome Sequence data from the samples from the Bijagós were compared with publicly available data from mainland Guinea-Bissau and other African countries, through the MalariaGEN Pf3k project and ongoing LSHTM studies (<https://www.malariagen.net/parasite/pf3k>). This included whole genome sequences from 795 isolates, from 10 different countries across Africa (Supplementary Table 1). This included isolates from Central Africa (Gabon, n = 57), West Africa (Ivory Coast n = 70, The Gambia n = 148, Bijagós Archipelago Guinea-Bissau (n = 15), East Africa (Kenya n = 138, Uganda n = 17), Southeast Africa (Malawi n = 149, Madagascar n = 24), South Central Africa (DRC n = 150) and Horn of Africa (Eritrea n = 27). A binary matrix of pairwise genetic distances was constructed from this WGS data. To investigate population structure, Principal Component Analysis was conducted using the filtered bi-allelic nucleotide matrix, using the R package *amap* (version 0.8–18), and a neighbour-joining (N-J) tree was constructed using the R package *ape* (version 5.5). Neighbour-joining trees were visualised in iTOL³⁷. Fixation index statistics (F_{ST}) were calculated to infer nucleotide diversity and population differentiation using VCFtools (version 0.1.16, <https://vcftools.github.io/documentation.html#fst>). F_{ST} statistic outputs were only included in downstream analysis if a particular SNP had coverage in at least 2/3 of both of the total sample populations that were being compared.

The F_{WS} metric was used to quantify the extent of multiplicity of infection (MOI), by assessing within host diversity of *Plasmodium* in comparison to local population diversity³⁸. This was calculated using the moimix R package (version 0.0.2.9001, <https://bahlolab.github.io/moimix/>)³⁹. This indicated the frequency of inbreeding

and outcrossing in the population. High F_{WS} (> 0.95) is highly indicative of monoclonality, whereas low F_{WS} indicates high levels of mixing between the parasites in the population, and subsequently a poorly defined population sub-structure.

IBD analysis was conducted using *hmmIBD* software (version 2.0.4) (default parameters). This software implements a hidden Markov model for detecting genomic regions that are identical by descent (IBD) for pairs of haploid samples¹⁴. Additional filtering steps were performed on the binary SNP matrix for IBD analysis: SNPs were selected with minor allele frequency > 0.01 , samples were filtered for $F_{WS} > 0.95$ and mixed calls were filtered to alternate calls. IBD analysis enabled inference of pairwise identity by descent between *Plasmodium* isolates within populations. Genomic relatedness was determined by analysing the proportion, or fraction, of the genome that was IBD between *Plasmodium* isolates. Cumulative IBD across the genome was calculated using sliding window analysis (using windows of 10,000 bases) for each population. The median and range of IBD values were calculated. Thirteen of the Bijagós samples passed F_{WS} filtering (≥ 0.95) to be included in IBD analysis, and despite this small sample size, the raised IBD results appear robust.

Whole genome sequence data was interrogated for signatures of selection, using the *rehh* package to calculate two different statistics derived from extended haplotype homozygosity: *iHS* and *XP-EHH*⁴⁰. These statistics are based on the decay of haplotype homozygosity, due to recombination, in the absence of selection on a particular allele⁴¹. Integrated haplotype score (*iHS*) identifies signatures of recent selection in a particular population. A negative *iHS* suggests that selection has favoured a derived allele, whereas a positive *iHS* suggests selection for the ancestral allele⁴¹. Cross-population extended haplotype homozygosity (*XP-EHH*) compares selection between two populations, through comparing the lengths of haplotypes associated with the same allele in different populations⁴². Positive *XP-EHH* indicates selection occurring in population A, whereas negative *XP-EHH* indicates selection occurring in population B. Highly variable genes (e.g., *PfEMP1*) were removed from this analysis due to the increased likelihood of them indicating false signals of selection pressure, due to an increased risk of alignment errors.

Admixture analysis was conducted using the *ADMIXTURE* software (version 1.3)⁴³. Analysis was conducted on 387,395 SNPs with minor allele frequencies of at least 0.001. This is a program used to estimate ancestry in unrelated individuals, from autosomal SNP genotype sets. *PLINK* (version 1.9) was used to convert the filtered VCF file to a bed file. *ADMIXTURE* was then used to estimate ancestry using this bed file and a specified *K* value. *K* is the estimated number of ancestral populations. The optimum *K* value for ancestral admixture coefficients was estimated through cross-validation of 1–10 dimensions of eigenvalue decay and found to be *K* = 5. Cross-validation was repeated using 10 randomised seed inputs and results were averaged across seeds to find the optimum *K* value. The output was visualised in *R* (v3.5.1). SNPs associated with resistance to antimalarial drugs were analysed in whole genome sequence data using the *dplyr* package (v1.0.10) in *R*.

Data availability

The processed datasets are available at NCBI: <http://www.ncbi.nlm.nih.gov/bioproject/905557>. Bioproject: PRJNA905557. The following sample SRR accession numbers can be used to find individual sample sequence data using the SRA database search in NCBI, <https://www.ncbi.nlm.nih.gov/sra/>: SRR22412636, SRR22412629, SRR22412635, SRR22412622, SRR22412623, SRR22412624, SRR22412625, SRR22412626, SRR22412627, SRR22412628, SRR22412630, SRR22412631, SRR22412632, SRR22412633, and SRR22412634.

Received: 23 November 2022; Accepted: 8 April 2023

Published online: 18 April 2023

References

- World Malaria Report 2021. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.
- Geneva: World Health Organization; 2022 (WHO/UCN/GMP/2022.01 Rev.2). License: CC BY-NC-SA 3.0 IGO. *WHO Guidelines for malaria - 3 June 2022*. <http://apps.who.int/bookorders>. (2022).
- Nsanjabana, C., Djalle, D., Guérin, P. J., Ménard, D. & González, I. J. Tools for surveillance of anti-malarial drug resistance: An assessment of the current landscape. *Malar J* **17**, 1–16 (2018).
- Hutchins, H. *et al.* A survey of knowledge, attitudes and practices regarding malaria and bed nets on Bubaque Island Guinea-Bissau. *Malar. J.* <https://doi.org/10.1186/s12936-020-03469-1> (2020).
- McGregor, D. *et al.* The epidemiology of plasmodium falciparum malaria in the Bijagós islands of Guinea-Bissau. *Am. J. Trop. Med. Hyg.* **104**, 2117–2122 (2021).
- Ant, T. *et al.* A survey of Anopheles species composition and insecticide resistance on the island of Bubaque, Bijagós Archipelago, Guinea-Bissau. *Malar. J.* **19**, 1–10 (2020).
- Oyola, S. O. *et al.* Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malar. J.* <https://doi.org/10.1186/s12936-016-1641-7> (2016).
- MalariaGEN Pf3k Project. <https://www.malariagen.net/parasite/pf3k>.
- Auburn, S. *et al.* Characterization of within-host plasmodium falciparum diversity using next-generation sequence data. *PLoS ONE* **7**, 32891 (2012).
- Amambua-Ngwa, A. *et al.* Major subpopulations of Plasmodium falciparum in sub-Saharan Africa. <https://www.science.org>.
- Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution (N Y)* **38**, 1358 (1984).
- Ghosh, S. *et al.* The Plasmodium rhoptry associated protein complex is important for parasitophorous vacuole membrane structure and intraerythrocytic parasite growth. *Cell Microbiol.* **19**, e12733 (2017).
- Triglia, T., Tham, W. H., Hodder, A. & Cowman, A. F. Reticulocyte binding protein homologues are key adhesins during erythrocyte invasion by Plasmodium falciparum. *Cell Microbiol.* **11**, 1671–1687 (2009).
- Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. *hmmIBD*: Software to infer pairwise identity by descent between haploid genotypes. *Malar. J.* **17**, 1–4 (2018).
- Roepe, P. D. Molecular and physiologic basis of quinoline drug resistance in Plasmodium falciparum malaria. *Future Microbiol.* **4**, 441–455. Preprint at <https://doi.org/10.2217/fmb.09.15> (2009).

16. Basco, L. K. & Ringwald, P. Molecular epidemiology of malaria in Yaounde, Cameroon. vi. Sequence variations in the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene and in vitro resistance to pyrimethamine and cycloguanil. *Am. J. Trop. Med. Hyg.* **62** (2000).
17. Hyde, J. E. Drug-resistant malaria - An insight. *FEBS J.* **274**, 4688–4698. Preprint at <https://doi.org/10.1111/j.1742-4658.2007.05999.x> (2007).
18. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, A. F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* **403**, 906–909 (2000).
19. Pickard, A. L. *et al.* Resistance to antimalarials in Southeast Asia and genetic polymorphisms in pfmdr1. *Antimicrob. Agents Chemother.* **47**, 2418–2423 (2003).
20. Ménard, D. *et al.* A Worldwide Map of *Plasmodium falciparum* K13-propeller polymorphisms. *N Engl J Med* **374**, 2453–2464 (2016).
21. Amato, R. *et al.* Genetic markers associated with dihydroartemisinin–piperaquine failure in *Plasmodium falciparum* malaria in Cambodia: A genotype–phenotype association study. *Lancet Infect. Dis.* **17**, 164–173 (2017).
22. Bousema, J. T. *et al.* *Plasmodium falciparum* gametocyte carriage in asymptomatic children in western Kenya. *Malar. J.* **3**, (2004).
23. Zoghi, S. *et al.* Survey for asymptomatic malaria cases in low transmission settings of Iran under elimination programme. *Malar. J.* **11**, 1–10 (2012).
24. Andolina, C. *et al.* Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: A longitudinal, observational cohort study. *Lancet Infect. Dis.* **21**, 1568–1578 (2021).
25. Klein, E. Y., Smith, D. L., Boni, M. F. & Laxminarayan, R. Clinically immune hosts as a refuge for drug-sensitive malaria parasites. *Malar. J.* **7**, (2008).
26. Diez Benavente, E. *et al.* A molecular barcode to inform the geographical origin and transmission dynamics of *Plasmodium vivax* malaria. *PLoS Genet* **16**, e1008576 (2020).
27. Kanoi, B. N. *et al.* Global repertoire of human antibodies against *Plasmodium falciparum* RIFINs, SURFINs, and STEVORs in a malaria exposed population. *Front. Immunol.* **11**, 893 (2020).
28. Prajapati, S. K. *et al.* The transcriptome of circulating sexually committed *Plasmodium falciparum* ring stage parasites forecasts malaria transmission potential. *Nat. Commun.* **11**, (2020).
29. Richardson, L. W. *et al.* Substrate peptidomimetic inhibitors of *P. falciparum* Plasmepepsin X with potent antimalarial activity. *ChemMedChem* **17**, e202200306 (2022).
30. Koenderink, J. B., Kavishe, R. A., Rijpma, S. R. & Russel, F. G. M. The ABCs of multidrug resistance in malaria. *Trends Parasitol.* **26**, 440–446 (2010).
31. Alves, K. C. S., Guimarães, J. M., de Almeida, M. E. M. & Mariúba, L. A. M. *Plasmodium falciparum* merozoite surface protein 3 as a vaccine candidate: A brief review. *Rev. Inst. Med. Trop. Sao Paulo* **64**, (2022).
32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
33. Danecsek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
34. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
35. Benavente, E. D. *et al.* Distinctive genetic structure and selection patterns in *Plasmodium vivax* from South Asia and East Africa. *Nat. Commun.* **12**, 1–11 (2021).
36. Cingolani, P., Platts, A., Wang le L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., Ruden, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**(2), 80–92 (2012). PMID: 22728672
37. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
38. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375 (2012).
39. Lee, S., Harrison, A., Tessier, N., Tavul, L., Miotto, O., Siba, P., Kwiatkowski, D., Müller, I., Barry, A. E., Bahlo, M. Assessing clonality in malaria parasites using massively parallel sequencing data, 2016, in preparation.
40. Gautier, M., Klassmann, A. & Vitalis, R. rehh 2.0: A reimplement of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17**, 78–90 (2017).
41. Saetre, G.-P., Ravinet, M. *Evolutionary Genetic - Concepts, Analysis and Practice* (OUP Oxford, 2019).
42. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**, e171 (2007).
43. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

Acknowledgements

We wish to thank the communities and study participants of the Bijagós Archipelago, Guinea-Bissau.

Author contributions

This study was designed by SM. AL was Principal Investigator of the DTNMaPa mapping survey, during which dried blood spots were collected by the mapping team, led by AG and ETS. HV conducted DNA extraction of dried blood spots and qPCR for malaria positivity for dried blood spot samples, under the supervision of AL. SM conducted all *Plasmodium* DNA amplification under the supervision of SC. SM performed all bioinformatic analysis with the support of EM, AO, and JP, and under the supervision of SC and TGC. SM wrote the first draft of the manuscript, which was reviewed and contributed to by all authors. The final manuscript was read and approved by all authors.

Funding

SM is funded by Medical Research Council UK (Grant No. MR/N013638/1). EM is funded by MRC MR/X005895/1. TGC and SC are funded by Medical Research Council UK grants (Ref. no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, MR/R020973/1, and MR/X005895/1). AL, HV and ETS are funded by Joint Global Health Trials Scheme (MRC, Wellcome Trust, UKRI, NIHR, Grant no. MR/S005013/1). SK is co-chair of the Guidelines Development Group for Antimalarials of the World Health Organization; views here are personal and do not represent those of the committee. AO is funded by a LSHTM PhD studentship funded by the WISE programme of MEXT. The DTNMaPa mapping survey was funded by MR/P023843/1 MRC Global Challenges Research Foundation.

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/s41598-023-33176-1>.

Correspondence and requests for materials should be addressed to S.M. or S.C.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023