Plasmodium falciparum Malaria Parasites in Ghana Show Signatures of Balancing Selection at Artemisinin Resistance Predisposing Background Genes

Evolutionary Bioinformatics
Volume 17: 1–9

DOI: 10.1177/1176934321999640

Kwesi Z Tandoh[1], Lucas Amenga-Etego[1], Neils B Quashie[2,3], Gordon Awandare[1], Michael Wilson[4] and Nancy O Duah-Quashie[2]

[1]West African Centre for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, College of Basic and Applied Sciences, University of Ghana, Accra, Ghana. [2]Department of Epidemiology, Noguchi Memorial Institute for Medical Research, College of Health Sciences, University of Ghana, Accra, Ghana. [3]Centre for Tropical Clinical Pharmacology and Therapeutics, School of Medicine and Dentistry, College of Health Sciences, University of Ghana, Accra, Ghana. [4]Department of Parasitology, Noguchi Memorial Institute for Medical Research, College of Health sciences, University of Ghana, Accra, Ghana.

ABSTRACT: Sub-Saharan Africa is courting the risk of artemisinin resistance (ARTr) emerging in *Plasmodium falciparum* malaria parasites. Current molecular surveillance efforts for ARTr have been built on the utility of *P. falciparum* kelch13 (*pfk13*) validated molecular markers. However, whether these molecular markers will serve the purpose of early detection of artemisinin-resistant parasites in Ghana is hinged on a *pfk13* dependent evolution. Here, we tested the hypothesis that the background *pfk13* genome may be present before the *pfk13* ARTr-conferring variant(s) is selected and that signatures of balancing selection on these genomic loci may serve as an early warning signal of ARTr. We analyzed 12 198 single nucleotide polymorphisms (SNPs) in Ghanaian clinical isolates in the Pf3K MalariaGEN dataset that passed a stringent filtering regimen. We identified signatures of balancing selection in 2 genes (phosphatidylinositol 4-kinase and chloroquine resistance transporter) previously reported as background loci for ARTr. These genes showed statistically significant and high positive values for Tajima's D, Fu and Li's F, and Fu and Li's D. This indicates that the biodiversity required to establish a *pfk13* background genome may have been primed in clinical isolates of *P. falciparum* from Ghana as of 2010. Despite the absence of ARTr in Ghana to date, our finding supports the current use of *pfk13* for molecular surveillance of ARTr in Ghana and highlights the potential utility of monitoring malaria parasite populations for balancing selection in ARTr precursor background genes as early warning molecular signatures for the emergence of ARTr.

KEYWORDS: Population genomics, artemisinin resistance, molecular surveillance, malaria, *Plasmodium falciparum*, signatures of balancing selection

RECEIVED: August 29, 2020. **ACCEPTED:** February 5, 2021.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: KZT is supported by PhD fellowships from Building a New Generation of Academics (BANGA-Africa, University of Ghana, and Carnegie Corporation of New York); and WACCBIP-World Bank ACE. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (107755/Z/15/Z: Awandare) and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust, or the UK government.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Nancy O Duah-Quashie, Department of Epidemiology, Noguchi Memorial Institute for Medical Research, College of Health Sciences, University of Ghana, Accra, Ghana. Email: NDuah@noguchi.ug.edu.gh

Kwesi Z Tandoh, West African Centre for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, College of Basic and Applied Sciences, University of Ghana, Accra, Ghana. Email: kztandoh@st.ug.edu.gh

Introduction

The global *Plasmodium falciparum* malaria burden is highest in sub-Saharan Africa (SSA). In 2018, the World Health Organization (WHO) estimated that 213 million people in SSA suffered from malaria, with an estimated 380 000 deaths.[1] In Ghana, malaria transmission is hyperendemic, with a high malaria attributable morbidity and mortality, particularly among pregnant women and children under 5 years of age.[2] Until 2005, chloroquine was the main drug used for malaria treatment in Ghana, however, widespread *P. falciparum* resistance led to the replacement of chloroquine with artemisinin-based combination therapy (ACT) as the first-choice antimalarial chemotherapy for the management of uncomplicated malaria in line with the WHO recommendations.[3-5] Currently, ACTs recommended by the national malaria control program (NMCP) of Ghana include artesunate–amodiaquine (AS AQ), artemether-lumefantrine (AL), and dihydroartemisinin-piperaquine (DHAPQ).[6] Although artemisinin-resistant (ARTr) *P. falciparum* has not been confirmed in Ghana or SSA yet, there is a real risk of such parasites emerging and severely undermining malaria control efforts.[7-9] Since the discovery of ARTr parasites in South-East Asia (SEA),[10,11] there is global concern about the spread of these parasites to SSA. Currently, molecular surveillance for ARTr is achieved with validated *P. falciparum* kelch-13 (*pfk13*) molecular markers.[12-14] However, with the possibility of involvement of multiple mechanisms in ARTr development, there are concerns about the universal utility of *pfk13* as a marker of ARTr.[15-21] Additionally, SSA has been shown to harbor a biodiverse *P. falciparum* population with sufficient variation at the *pfk13* locus to allow a soft selection sweep under ART drug pressure.[9,18,22] Despite this, ARTr malaria parasites are yet to emerge in SSA.[8,9]

A possible explanation for the delay of artemisinin resistance emergence in SSA is the prerequisite for a *pfk13* ARTr-conferring background genome. This model is framed on the hypothesis that the *pfk13* ARTr-conferring variants only express the phenotype in the presence of a particular *pfk13* genomic background that weathers any fitness cost of the *pfk13* ARTr conferring variant(s) and may augment any baseline partner drug resistance required for the parasites to be transmitted.[23,24] Whether this *pfk13* genome background evolves before, with, or after the *pfk13* ARTr variant is selected remains unclear.[18] This ARTr genomic model presents a high genetic barrier for the emergence of ARTr and may explain the delay in the manifestation of the phenotype in SSA. The *pfk13* background genome has been characterized by population association studies to involve single nucleotide polymorphisms (SNPs) in the following loci: ferredoxin (fd-D193Y), apicoplast ribosomal protein S10 (arps10-V127M), multidrug resistance protein2 (mdr2-T484I), and chloroquine resistance transporter (crt-N326S).[23] A retrospective longitudinal genomic study also identified background SNPs that were associated with the ARTr phenotype in malaria parasites from North-Western Thailand.[24] This study used genotype-phenotype association tests and genomic scans for signals of selection to determine genomic loci in *P. falciparum* that may contribute to the *pfk13* ARTr background genome. These variants followed a similar non-reference allele frequency temporal trajectory as the dominant *pfk13* ARTr conferring variant C580Y in SEA and included SNPs in the following genes: Phosphatidylinositol 4-kinase (S915G), Sec14 domain-containing protein (L498F and N615D), Ubiquitin-protein ligase (S57T), Ubiquitin carboxyl-terminal hydrolase (R3138H), and Sentrin-specific protease 2, putative protein (H423Y).[24]

To mitigate the risk of ARTr spreading and taking hold in SSA, molecular surveillance for parasites with ARTr is critical and a molecular-based early warning metric that predicts ARTr will be desirable. A tool that may serve this purpose is the detection of signatures of balancing selection on loci demonstrated to act as a background for ARTr emergence. To explore this, here we hypothesized that the background *pfk13* genome may be present before the *pfk13* ARTr-conferring variant is selected and that signatures of balancing selection on these genomic loci may serve as an early warning signal. Our analysis has identified loci previously reported as *pfk13* background mutations required for the emergence of ARTr to be under balancing selection in *P. falciparum* clinical samples from Ghana. Our findings show that 2 of these genomic loci required for ARTr to emerge have sufficient standing variation to allow a soft selective sweep and emergence of ART tolerant parasites under ART pressure. Our data also provides empirical evidence for malaria molecular surveillance in Ghana continue to use *pfk13* as the marker for ARTr despite the possibility of *pfk13* independent emergence of ARTr parasites in Ghana.

Results

Multiplicity of infection

Multiplicity of infection (MOI) can result in spurious signatures of selection metrics if it is not evaluated.[22] The proportion of samples with multiple infections was determined by calculating the genome-wide within sample F statistic (F_{ws}). We found 56% (343) of sample to have $F_{ws} < 0.95$, which is indicative of multi-genomic infection (Figure 1). These samples were excluded from tests of neutrality analysis to determine signatures of balancing selection.

P. falciparum population structure

Next, we interrogated the hypothesis that there is no population structure in the malaria parasite population in Ghana. Population structure analysis was undertaken using principal component analysis (PCA). We sought to uncover any effect of varying local transmission and geographical distribution on the parasite population structure. We adopted an approach that evaluates population structure by utilizing within-sample allele frequencies (WSAF) across all variable genomic loci.[25] Principal components were plotted using ggplot2 package in R version 4.0. No significant parasite population structure was observed (Figure 2).

PCA analysis for samples with single clonal infection (N = 274). Sample coordinates for PCA were obtained using *vcfdo*. A subsequent plot was generated in R.4.0 with the ggplot2 package. Principal components 1 and 2 show no genetic differentiation within the Ghana samples of *P. falciparum*

Summary of genomic diversity among single clone infections

We subsequently only analyzed sequence data for samples with single clone infections ($F_{ws} > 0.95$, N = 274) with 12 198 biallelic SNPs. Sixty percent of SNPs in the population had a non-reference allele frequency of less than 25% (Figure 3). Forty-five percent of SNPs were missense variants, 28% of SNPs were in intergenic regions, 22% of SNPs were synonymous variants, and 4% were in intronic regions (Figure 4). The SNPs analyzed were distributed across 2256 protein-coding genes.

The most abundant biological effect of the SNPs analyzed were missense variants, variants in intergenic regions, synonymous variants, and intron variants in decreasing order.

Pfk13 ARTr genome background loci under balancing selection

We used clonal samples ($F_{ws} > 0.95$) to screen for signatures of selection in the *P. falciparum* genome. We found 579 *P. falciparum* genes with Tajima's D greater than 1 (Figure 5, Supplemental Table S1).[26-28] Of these, genomic positions above the 90th percentile of the Tajima's D distribution showed

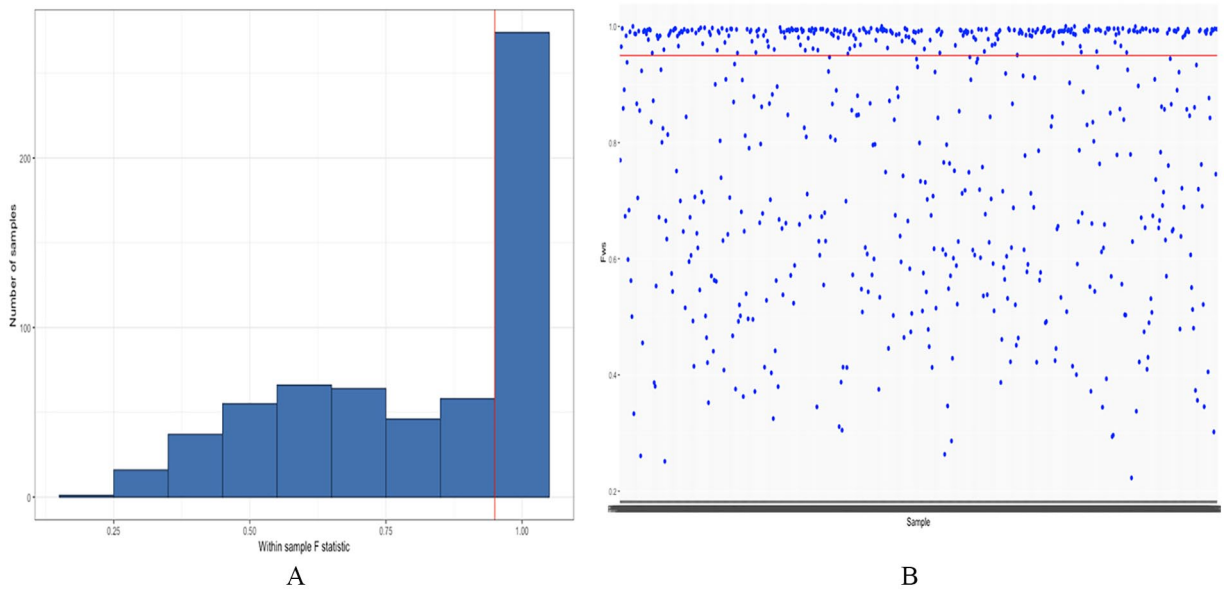

Figure 1. Distribution of multiplicity of infection in Ghana samples of *P. falciparum* analyzed (n = 617): (A) is a histogram showing the number of samples with within sample F statistic (F_{ws}) on the vertical axis within the range specified on the horizontal axis and (B) is a scatter plot that shows the distribution of the samples with F_{ws}.
The red line is at $F_{ws} = 0.95$, the cutoff point for MOI.

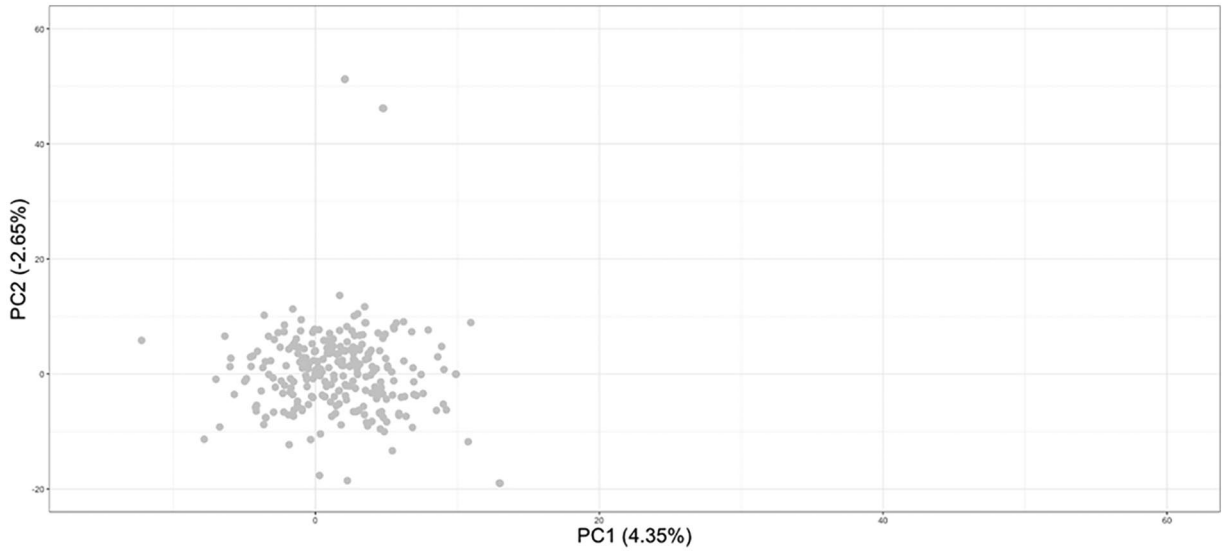

Figure 2. *P. falciparum* population structure analysis for Ghana samples.

gene ontology biochemical processes spanning multiple pathways (Supplemental Table S2). There were no SNPs in the *pfk13* gene and none of the reported SNPs that form the background genome for ARTr to emerge were identified in this analysis (Supplemental Table S3).

Five hundred seventy-nine *P. falciparum* genes had Tajima's $D > 1$. Chromosomes are identified by the alternately black and blue coloring, with SNPs plotted as individual points based on their position in the chromosome. The red dashed line marks Tajima's $D = 1$.

Among the genes under balancing selection, we looked for loci previously reported to be part of the required multi-locus genomic background on which ARTr-conferring *pfk13* variants emerged.[23,24] Among the 9 genomic loci implicated in this background phenomenon (Table 1), 2 were found to be under balancing selection in the Ghana sample set collected in 2010 (Table 2). One of these genes (phosphatidylinositol 4-kinase) stand out in their gene ontology terms (http://plasmodb.org) as directly connected to the molecular mechanisms of ARTr development.[15]

The phosphatidylinositol 4-kinase (PI4K) gene product shares similar gene ontology terms (phosphatidylinositol metabolic process, kinase activity, and cellular membrane component) with phosphatidylinositol 3-kinase (PI3K), which has

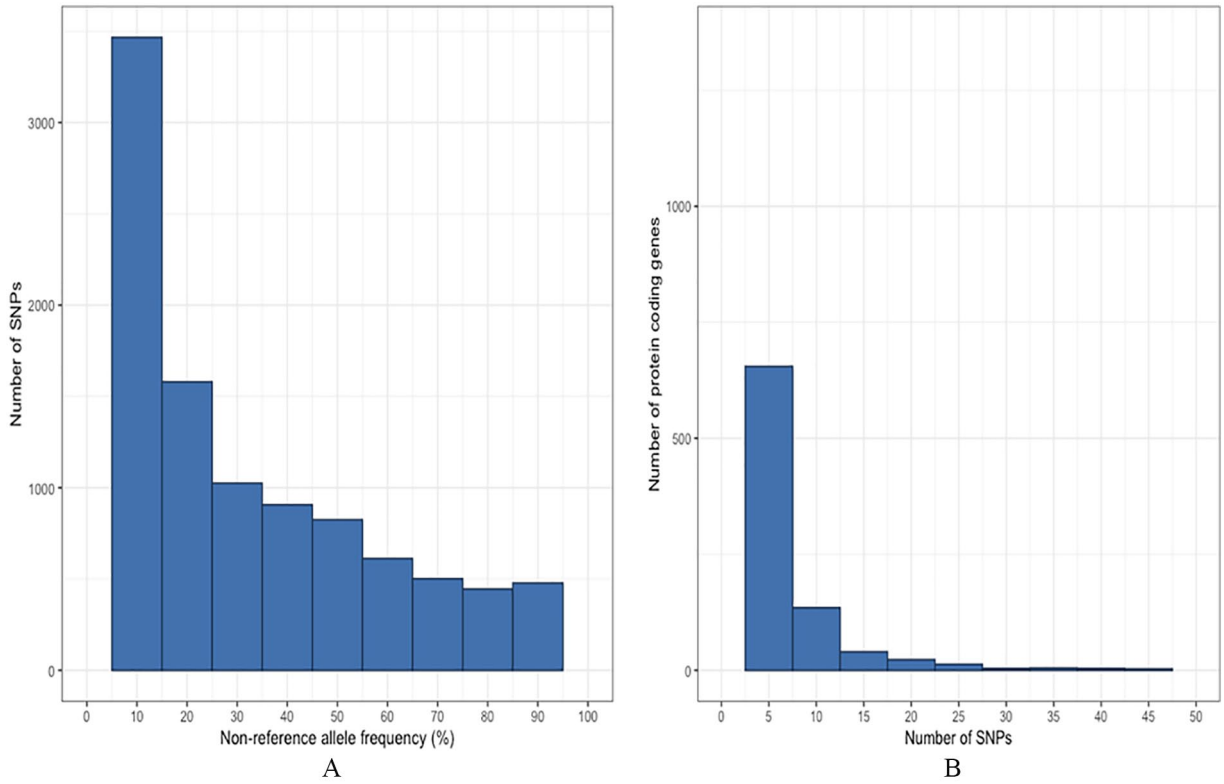

Figure 3. Summary of SNPs characteristics: (A) Frequency distribution of the non-reference allele for each of the biallelic SNPs in the sample of *P. falciparum* clinical isolates from Ghana (N = 274) and (B) distribution of numbers of protein-coding genes (N = 2256) with each given number of SNPs in the Ghana population sample of *P. falciparum* clinical isolates.

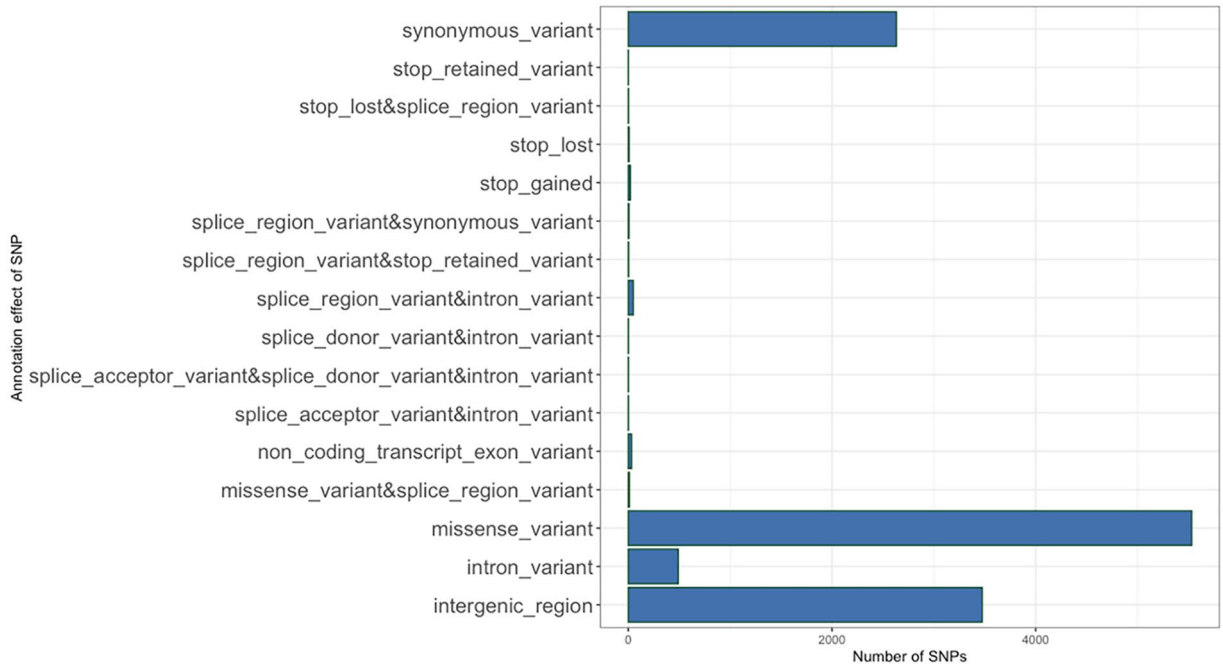

Figure 4. Distribution of SNPs effect across all genomic positions analyzed.

been associated with the primary ARTr-conferring variant *pfk13* C580Y. The gene is located on chromosome 4 and has 6 exons that encode for a protein with 5035 amino acid residues. We found 38 SNPs in this gene of which 16 were synonymous, 20 missense variants, and 2 intron variants. The ubiquitin carboxyl-terminal hydrolase 1 locus (UBP1) gene is located on chromosome 3 and has 3 exons that encode for a protein with 3499 amino acid residues. We found 6 SNPs in this gene of

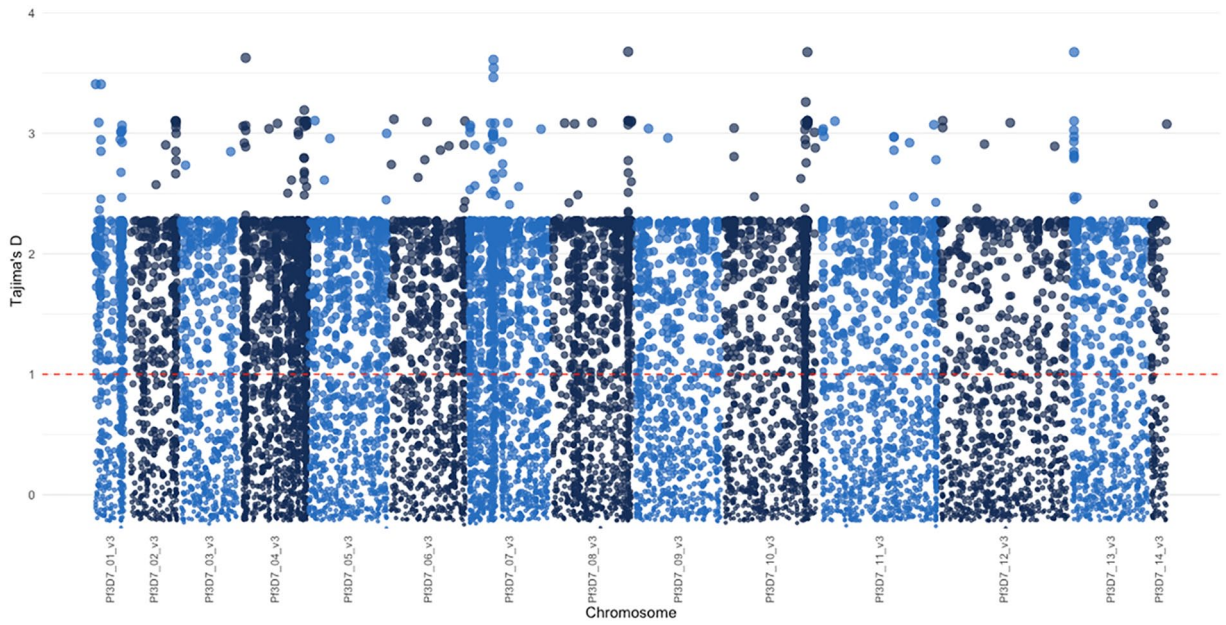

Figure 5. Genome-wide distribution of Tajima's D values summarizing the site frequency spectra for *P. falciparum* genes.

Table 1. List of genes of interest reported to be associated with the genomic background for artemisinin resistance in *P. falciparum*.

GENE NAME	GENE ID	SNP
Ferredoxin	PF3D7_1318100	D193Y
Apicoplast ribosomal protein S10	PF3D7_1460900	V127M
Multidrug resistance protein 2	PF3D7_1447900	T484I
Chloroquine resistance transporter	PF3D7_0709000	N326S
Phosphatidylinositol 4-kinase	PF3D7_0419900	S915G
Sec14 domain-containing protein	PF3D7_0626400	L498F, N615D
Ubiquitin-protein ligase	PF3D7_1448400	S57T
Ubiquitin carboxyl-terminal hydrolase	PF3D7_0104300	R3138H
Sentrin-specific protease 2, putative	PF3D7_0801700	H423Y

which 2 were synonymous and 4 missense variants. The chloroquine resistance transporter (CRT) gene is located on chromosome 7 and has 13 exons that encode for a protein with 424 amino acid residues. We found 6 SNPs in this gene of which 3 were missense variants and 3 were intron variants. One missense variant was the chloroquine-resistance K76T mutation with an allele frequency of 0.57%.[29]

The Sentrin-specific protease 2, putative (SENP2) gene is located on chromosome 8 and has 3 exons that encode for a protein with 1784 amino acid residues. We found 2 SNPs in this gene of which one was a missense variant and the other a synonymous variant (Table 3).

Discussion

It is important to improve malaria molecular surveillance, particularly detection and monitoring of *P. falciparum* parasites with ARTr genotypes. Given the absence of confirmed ARTr in SSA and the complex evolution of ARTr, which includes precursor genomic background, we explored the plausibility of using molecular early warning signatures of balancing selection in 9 genes associated with the genomic background on which ARTr conferring variants may emerge.[18,23,24] We found signatures of balancing selection in 2 loci reported to contribute to the genomic background on which definitive ARTr-conferring variants arise (Table 2). Genomic loci under balancing selection provide a standing variation on which a soft selective sweep can occur, especially in the presence of high transmission rates, which is common in Ghana.[30-32] Although we did not find any of the reported multi-loci ARTr predisposing background SNPs (Supplemental Table S3) and none of the variants identified at these loci have been reported to be associated with ARTr,[33] the increased genetic diversity observed at these

Table 2. Tests of neutrality on 4 genes of interest under balancing selection in Ghana population of *Plasmodium falciparum*.

GENE NAME	GENOMIC LENGTH (BP)	SITES[1]	π ($\times 10^{-5}$)	K ($\times 10^{-3}$)	T' D	*F*	*D*	HKA (π/K)	MK *P* VALUE
Chloroquine resistance transporter	3096	6	64	559.71	2.7*	1.96*	1.04	0.11	.68
Phosphatidylinositol 4-kinase	15845	38	67	645.02	2.2 *	2.6*	2.18*	0.10	.32
Ubiquitin carboxyl-terminal hydrolase	10962	6	17	615.51	1.8	1.5	1.04	0.03	.64
Sentrin-specific protease 2, putative	5579	2	0.07	591.25	1.1	0.9	0.62	11×10^{-3}	1

[1]Number of polymorphic sites; π is nucleotide diversity index; *K* is nucleotide divergence; T'D is Tajima's D; *F* is Fu and Li's F; *D* is Fu and Li's D.
*P value $< .05$.

Table 3. Summary of number of SNPs and effect at loci of interest.

GENE NAME	ANNOTATION EFFECT		
	MISSENSE VARIANT	SYNONYMOUS VARIANT	INTRON VARIANT
Chloroquine resistance transporter (crt)	3	0	3
Phosphatidylinositol 4-kinase	20	16	2
Ubiquitin carboxyl-terminal hydrolase (ubp1)	4	2	0
Sentrin-specific protease 2, putative (senp2)	1	1	0

loci as of 2010 (Table 2) suggests that the biodiversity required to establish a *pfk13* background genome may be primed in Ghana. Since balanced alleles may be randomly lost over long periods, whether these variants will be maintained, increase in frequency or be lost over time remains unclear. Against the backdrop of the absence of ARTr in Ghana despite this background, the delay may be attributed to the use of ACTs or breakdown of standing variation by rampant recombination due to the high transmission intensity in Ghana.

PI4K may have an impact on the levels of phosphatidylinositol-3-phosphate (PI3P) which has been associated with ARTr *in vitro*.[16] PI4K may also contribute to ARTr development by modulating the activity of members of the phosphoinositol pathways in triggering the unfolded protein response to ART stress via the inositol requiring enzyme 1α/β (IRE1).[24,34] UBP1 is a putative protein predicted to function in a ubiquitin-dependent protein catabolic process (https://plasmodb.org/plasmo/app/record/gene/PF3D7_0104300#category:function-analysis). The ubiquitination/proteasome pathway has been associated with ARTr[16,20,35] via the cellular stress response to ART assault. Thus our finding of signatures of balancing selection in UBP1 gene suggests an association with these pathways related to the molecular mechanisms of ARTr. Our data support the point conveyed by Cerqueira et al[24] that the complex multi-loci nature of ARTr makes it less likely to occur *de novo* in SSA. This is because of the high rate of outcrossing and recombination that may derail the stability of the required complex genomic background. However, the presence of signatures of balancing selection in loci, that predispose to the emergence of *pfk13*-dependent ARTr, highlights the need for molecular surveillance of parasite populations in SSA to generate data for interrogating the risk of ARTr emerging in SSA whether *de novo* or imported by gene flow from SEA. Regardless of how ARTr arises in SSA, our analysis suggests that the requisite adaptation for ARTr in Ghana has occurred and conditions are ripe for a soft selective sweep either via importation or independent emergence occur sooner than later. It is in the light of this that we recommend prospective genomic surveillance for signatures of balancing selection in ARTr precursor background genes as they accumulate more variation and could serve as a molecular early warning system for ARTr emergence.

However, it is worth noting some limitations of this study. A key weakness of this study is the absence of statistical power evaluations to ascertain whether the hard filtering pipeline and the resulting number of SNPs were sufficient to interrogate the central hypothesis for this analysis. It may be argued that the size of the relatively small variant set was not powered enough to capture the veracity or otherwise of our hypothesis. In addition, the evidence from Africa on *pfk13* ARTr suggests that different variants from those seen in SEA may emerge.[36] However, we are confident, the absence of any known ARTr–conferring variants in the samples analyzed is a strong indication that the background genome may be established before the emergence of the definitive ARTr-conferring variant(s). Although tests of departures from neutrality are best suited for evaluating deviations of nucleotide diversity from expectations under a neutral

model of population evolution, they may also suggest a demographically driven change such as population bottleneck (eg, drug pressure).[37,38] What remains unknown and merits further investigation is the time scale between these events. Given that the effective population size is greater in SSA compared to SEA, the high recombination rates in SSA may prolong the evolutionary time scale to ARTr emergence. Therefore, the departures from neutrality from different complementary approaches observed in this study, though indicative of balancing selection in some of these genes, are restricted by this fact. However, the evidence of balancing selection supported by Tajima's D, Fu and Li's D, Fu and Li's F is quite strong for 2 of our genes of interest (Table 2). The lack of concordance between Tajima's D and HKA in our analysis is consistent with previous findings for genes widely reported to be under balancing selection in African parasite populations (Supplemental Table S5).[26,27] This may be attributed to high recombination rates, changing effective population size, and the presence of mutations that may not be selectively neutral.[39,40]

In general, we have detected loci under balancing selection, in Ghanaian *P. falciparum* isolates, which intersect with the genome background required for ARTr to emerge in SEA. Although Ghana and SSA are yet to report solid evidence of ARTr, our analysis shows that molecular surveillance efforts may be improved by scanning the background of such loci associated with *pfk13* ARTr. This approach can signal the emergence of the ARTr phenotype but may require optimization to serve as the missing tool needed in the current landscape of ARTr surveillance of *P. falciparum* malaria parasites in Africa. Thus, we have suggested the potential utility of surveying the genome of parasites for balancing selection in ARTr precursor background genes as a potential molecular early warning system for ARTr surveillance efforts.

Methods

Sequencing and genotyping

DNA sequences were retrieved from the Malaria Genomic Epidemiology Network (MalariaGEN) Pf3k data repository (https://www.malariagen.net/projects/parasite/pf3k). Paired-end sequencing was done using the Illumina HiSeq platform[41] by Sanger sequencing core. Read lengths ranged between 200 and 300 bp and approximately 1 Gbp of data per sample was produced as described.[18] Quality control, variant discovery, and sample genotyping were done using custom pipelines, details of which may be reviewed in Manske et al.[22] Briefly, short sequence reads from *P. falciparum* isolates that passed stringent quality filters were aligned against the *P. falciparum* 3D7 reference sequence V3 (ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium/falciparum/3D7/3D7.latest_version/version3/), using the bwa program (Li and Durbin, 2009) (http://bio-bwa.sourceforge.net/) as previously described.[23]

Genotyping was done with the Genome Analysis Tool Kit (GATK 3.3-46) best practices pipeline for *de novo* variant discovery. After mapping each sample to the 3D7 reference genome, PCR, and sequencing duplicates were marked using Picard SortSam and the MarkDuplicates function of GATKs pipeline. Realignment around InDels (insertion and deletion sequencevariants)wasdoneusingGATK'sRealignerTargetCreator/IndelRealigner. Then base quality score recalibration was done with GATK's BaseRecalibrator. The following *P. falciparum* genetic crosses 1.0 databases were used for the base quality recalibration: ftp://ngs.sanger.ac.uk/production/malaria/pf-crosses/1.0/7g8_gb4.combined.final.vcf.gz; ftp://ngs.sanger.ac.uk/production/malaria/pf-crosses/1.0/hb3_dd2.combined.final.vcf.gz; and ftp://ngs.sanger.ac.uk/production/malaria/pf-crosses/1.0/3d7_hb3.combined.final.vcf.gz.

Variants calling was done using GATK's CombineGVCFs and GenotypeGVCFs. This was followed by variant quality score recalibration using GATKs VariantRecalibrator/ApplyRecalibration functions. The same set of *P. falciparum* crosses 1.0 databases above were used. Finally, variants annotation was done using SnpEff (v4.1) with a database created from ftp://ftp.sanger.ac.uk/pub/project/pathogens/gff3/2015-08/*Pfalciparum*.gff3.gz; annotate variants by regions identified in Pf crosses 1.0 data (ftp://ngs.sanger.ac.uk/production/malaria/pf-crosses/1.0/regions-20130225.onebased.txt). Variants were filtered out where VQSLOD ≤ 0 or RegionType !="Core" to generate a final VCF file available at the MalariaGEN Pf3k data repository.

Hard filtering pipeline of Ghana 2009/2010 clinical isolates variants

Bcftools was used in a custom bash script to extract the Ghana 2009/2010 samples and concatenate all the variant information into a single VCF without structural variants such as InDels. Quality hard filtering of single nucleotide polymorphisms (SNPs) was done to filter out SNPs with the following features: SNPs with very poor sequencing depth (<10 reads in 1 sample) and coverage across samples; SNPs that had more than 2 alleles; and SNPs located in the highly polymorphic *var*, *rifin*, and *stevor* regions.[22] Further snpEFF annotation was done using the global barcode SNPs from a genome-wide study of *P. falciparum* population divergence by natural selection.[42] The final VCF file was filtered again using the default parameters of the GATK SelectVariants function. Samples with missing genotype calls of >5% were filtered out. VCFtools (0.1.16) was used to determine the genome-wide average nucleotide diversity.

Evaluation of multiplicity of infection and population structure

We used the within-sample F statistic (F_{ws}) to determine MOI in the Ghana cohort of clinical isolates of *P. falciparum*.[23,43] The moimix package (https://github.com/bahlolab/moimix) was used to determine the F_{ws} in R version 4.0.

Principal component analysis (PCA) was done using *vcfdo* (https://github.com/IDEELResearch/vcfdo) function "*pca*."

Tests for departures from neutrality

We used the R package PopGenome for genome scans and determination of the allele frequency distribution selection metric Tajima's D at each SNP site.[44] Subsequently, we used custom bash scripts to extract the DNA sequences of the 9 genes of interest (Supplemental Table S1). DnaSP6.0 was used to determine departures from neutrality, for the 9 genes of interest, using allele frequency distribution measures Tajima's D, Fu and Li's F, and Fu and Li's D, and comparisons of variations within and between species with Hudson–Kreitman–Aguade (HKA) and McDonald–Kreitman (MK) ratios.[45]

Tajima's D measures the difference between the average pairwise nucleotide diversity between sequences (π) and Watterson's population nucleotide diversity (θ).[27,46] This will make the expected value of Tajima's D for populations under neutrality to be zero. When conditions of neutrality are violated, Tajima's D will capture a skew in the allele frequency distribution. These significant deviations from zero can either be positive or negative. Positive values of Tajima's D indicate an excess of intermediate frequency alleles. This can be caused by population bottlenecks, population structure, and/or balancing selection.[47]

Fu and Li's F statistic is determined by the difference between the number of single nucleotide variants observed only once in a sample and the total number expected under neutrality given the number of segregating sites and Watterson's estimate of nucleotide diversity (θ); Fu and Li's D is the difference between the number of derived nucleotide variants observed only once in a sample with the total number of derived nucleotide variants.[47,48] HKA ratio was used to determine *P. falciparum* genes with high ratios of polymorphism(p) over divergence(K) from the closely related chimpanzee parasite *Plasmodium reichenowi*.[40,49] The MK test is based on the ratios of non-synonymous (NS) and synonymous (S) nucleotide divergence and polymorphism within species and a closely related species, using Fisher's exact test on the 2×2 contingency table.[50]

Finally, we used the PlasmoDB Gene Ontology (GO) analysis tool (http://plasmodb.org) to identify the biological processes of genes significantly deviating from neutrality.

Acknowledgements

This publication uses data generated by the Pf3k project (www.malariagen.net/pf3k). We thank the MalariaGEN Consortium for allowing the use of this data. KZT is also indebted to C.M. Morang'a and V. Appiah of the bioinformatics laboratory at WACCBIP, Department of Biochemistry, Cell and Molecular Biology for their invaluable help in troubleshooting some parts of the computational analysis pipeline. We acknowledge the University of Ghana for providing the high-performance computing resources (the ZUPUTO) used for this work.

Author Contributions

KZT conceptualized the research question, did the computational data analysis, and wrote the manuscript. LA, MW, NBQ, GA, and NOD reviewed and supervised the work. All authors read and approved the final manuscript.

ORCID iD

Kwesi Z Tandoh https://orcid.org/0000-0002-1628-2845

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Organization WH. World Malaria Report 2019. 2019.
2. Ministry of Health G. NMCP: Annual Report of the National Malaria Control Programme of Ghana, 2017. 2017.
3. Koram KA, Abuaku B, Duah N, Quashie N. Comparative efficacy of antimalarial drugs including ACTs in the treatment of uncomplicated malaria among children under 5 years in Ghana. *Acta Trop*. 2005;95:194-203.
4. Quashie NB, Duah NO, Abuaku B, Koram KA. The in-vitro susceptibilities of Ghanaian Plasmodium falciparum to antimalarial drugs. *Ann Trop Med Parasitol*. 2007;101:391-398.
5. Organization WH. Guidelines for the Treatment of Malaria. 2006.
6. Quashie NB, Ranford-Cartwright LC, de Koning HP. Uptake of purines in Plasmodium falciparum-infected human erythrocytes is mostly mediated by the human equilibrative nucleoside transporter and the human facilitative nucleobase transporter. *Malar J*. 2010;9:36.
7. Amambua-Ngwa A, Amenga-Etego L, Kamau E, et al. Major subpopulations of Plasmodium falciparum in sub-Saharan Africa. *Science*. 2019;365:813-816.
8. Conrad MD, Rosenthal PJ. Antimalarial drug resistance in Africa: the calm before the storm? *Lancet Infect Dis*. 2019;19:e338-e351.
9. Taylor SM, Parobek CM, DeConti DK, et al. Absence of putative artemisinin resistance mutations among Plasmodium falciparum in Sub-Saharan Africa: a molecular epidemiologic study. *J Infect Dis*. 2015;211:680-688.
10. Dondorp AM, Nosten F, Yi P, et al. Artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med*. 2009;361:455-467.
11. Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM. Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med*. 2008;359:2619-2620.
12. Ariey F, Witkowski B, Amaratunga C, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*. 2014;505:50.
13. Nyunt MH, Wang B, Aye KM, et al. Molecular surveillance of artemisinin resistance falciparum malaria among migrant goldmine workers in Myanmar. *Malar J*. 2017;16:97.
14. Mvumbi DM, Bobanga TL, Kayembe JN, et al. Molecular surveillance of Plasmodium falciparum resistance to artemisinin-based combination therapies in the Democratic Republic of Congo. *PLoS One*. 2017;12:e0179142.
15. Suresh N, Haldar K. Mechanisms of artemisinin resistance in Plasmodium falciparum malaria. *Curr Opin Pharmacol*. 2018;42:46-54.
16. Mbengue A, Bhattacharjee S, Pandharkar T, et al. A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. *Nature*. 2015;520:683-687.
17. Bhattacharjee S, Coppens I, Mbengue A, et al. Remodeling of the malaria parasite and host human red cell by vesicle amplification that induces artemisinin resistance. *Blood*. 2018;131:1234-1247.
18. Project MPfC. Genomic epidemiology of artemisinin resistant malaria. *Elife*. 2016;5:e08714.
19. Mok S, Ashley EA, Ferreira PE, et al. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*. 2015;347:431-435.
20. Rocamora F, Zhu L, Liong KY, et al. Oxidative stress and protein damage responses mediate artemisinin resistance in malaria parasites. *PLoS Pathogens*. 2018;14:e1006930.
21. Birnbaum J, Scharf S, Schmidt S, et al. A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. *Science*. 2020;367:51-59.
22. Manske M, Miotto O, Campino S, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature*. 2012;487:375-379.
23. Miotto O, Amato R, Ashley EA, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat Genet*. 2015;47:226.

24. Cerqueira GC, Cheeseman IH, Schaffner SF, et al. Longitudinal genomic surveillance of Plasmodium falciparum malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. *Genome Biol*. 2017;18:78.
25. Verity R, Aydemir O, Brazeau NF, et al. The impact of antimalarial resistance on the genetic structure of Plasmodium falciparum in the DRC. *Nat Commun*. 2020;11:2107.
26. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in Plasmodium falciparum. *Mol Biol Evol*. 2010;27:2344-2351.
27. Ochola-Oyier LI, Wamae K, Omedo I, et al. Few Plasmodium falciparum merozoite ligand and erythrocyte receptor pairs show evidence of balancing selection. *Infect Genet Evol*. 2019;69:235-245.
28. Amambua-Ngwa A, Tetteh KK, Manske M, et al. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genet*. 2012;8:e1002992.
29. Wellems TE, Plowe CV. Chloroquine-resistant malaria. *J Infect Dis*. 2001;184:770-776.
30. Barrett RD, Schluter D. Adaptation from standing genetic variation. *Trends Ecol Evol*. 2008;23:38-44.
31. Botwe AK, Asante KP, Adjei G, et al. Dynamics in multiplicity of Plasmodium falciparum infection among children with asymptomatic malaria in central Ghana. *BMC Genet*. 2017;18:67.
32. Lamptey H, Ofori MF, Kusi KA, et al. The prevalence of submicroscopic Plasmodium falciparum gametocyte carriage and multiplicity of infection in children, pregnant women and adults in a low malaria transmission area in Southern Ghana. *Malar J*. 2018;17:331.
33. WWARN K13 Genotype-Phenotype Study Group. Association of mutations in the Plasmodium falciparum Kelch13 gene (Pf3D7_1343700) with parasite clearance rates after artemisinin-based treatments – a WWARN individual patient data meta-analysis. *BMC Med*. 2019;17:1.
34. Adams CJ, Kopp MC, Larburu N, Nowak PR, Ali MMU. Structure and molecular mechanism of ER stress signaling by the unfolded protein response signal activator IRE1. *Front Mol Biosci*. 2019;6:11.
35. Dogovski C, Xie SC, Burgio G, et al. Targeting the cell stress response of Plasmodium falciparum to overcome artemisinin resistance. *PLoS Biol*. 2015;13:e1002132.
36. Uwimana A, Legrand E, Stokes BH, et al. Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. *Nat Med*. 2020;26:1602-1608.
37. Kloch A, Wenzel MA, Laetsch DR, et al. Signatures of balancing selection in toll-like receptor (TLRs) genes – novel insights from a free-living rodent. *Sci Rep*. 2018;8:8361.
38. Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*. 2008;179:555-567.
39. Hudson RR. Gene genealogies and the coalescent process. *Oxford Surv Evol Biol*. 1990;7:44.
40. Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. *Genetics*. 1987;116:153-159.
41. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53-59.
42. Neafsey DE, Schaffner SF, Volkman SK, et al. Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. *Genome Biol*. 2008;9:R171.
43. Auburn S, Campino S, Miotto O, et al. Characterization of within-host Plasmodium falciparum diversity using next-generation sequence data. *PLoS One*. 2012;7:e32891.
44. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 2014;31:1929-1936.
45. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299-3302.
46. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585-595.
47. Casillas S, Barbadilla A. Molecular population genetics. *Genetics*. 2017;205:1003-1035.
48. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133:693-709.
49. Innan H. Modified Hudson-Kreitman-Aguade test and two-dimensional evaluation of neutrality tests. *Genetics*. 2006;173:1725-1733.
50. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature*. 1991;351:652-654.