



## Research article

Single-nucleotide polymorphism characterization of gametocyte development 1 gene in *Plasmodium falciparum* isolates from Baringo, Uasin Gishu, and Nandi Counties, KenyaJosephat K. Bungei<sup>a,b</sup>, Victor A. Mobegi<sup>b,\*</sup>, Steven G. Nyanjom<sup>a</sup><sup>a</sup> Department of Biochemistry, JKUAT, Kenya<sup>b</sup> Department of Biochemistry, School of Medicine, University of Nairobi, Kenya

## ARTICLE INFO

## Keywords:

Bioinformatics  
 DNA sequencing  
 Gene mutation  
 Biochemistry  
 Molecular biology  
 Parasitology  
*Plasmodium falciparum*  
 Pfgdv1  
 Gametocytogenesis  
 Single-nucleotide polymorphism  
 Selection analysis

## ABSTRACT

**Introduction:** *Plasmodium falciparum* relies on gametocytogenesis to transmit from humans to mosquitoes. Gametocyte development 1 (*Pfgdv1*) is an upstream activator and epigenetic controller of gametocytogenesis. The emergence of drug resistance is a major public health concern and this requires the development of new strategies that target the transmission of malaria. As a putative drug target, *Pfgdv1* has not been characterized to identify its polymorphisms and alleles under selection and how such polymorphisms influence protein structure.

**Methods:** This study characterized single-nucleotide polymorphisms (SNPs) in primary sequences (n = 30) of *Pfgdv1* gene generated from thirty blood samples collected from patients infected with *P. falciparum* and secondary sequences (n = 216) retrieved from PlasmoDB. ChromasPro, MUSCLE, Tajima's D statistic, SLAC, and STRUM were used in editing raw sequences, performing multiple sequence alignment (MSA), identifying signatures of selection, detecting codon sites under selection pressure, and determining the effect of SNPs, respectively.

**Results:** MSA of primary and secondary sequences established the existence of five SNPs, consisting of four non-synonymous substitutions (nsSNPs) (p.P217H, p.R398Q, p.H417N, and p.D497E), and a synonymous substitution (p.S514S). The analysis of amino acid changes reveals that p.P217H, p.R398Q, and p.H417N comprise non-conservative changes. Tajima's D statistic showed that these SNPs were under balancing selection, while SLAC analysis identified p.P217H to be under the strongest positive selection. Further analysis based on thermodynamics indicated that p.P217H has a destabilizing effect, while p.R398Q and p.D497E have stabilizing effects on the protein structure.

**Conclusions:** The existence of four nsSNPs implies that *Pfgdv1* has a minimal diversity in the encoded protein. Selection analysis demonstrates that these nsSNPs are under balancing selection in both local and global populations. However, p.P217H exhibits positive directional selection consistent with previous reports where it showed differential selection of *P. falciparum* in low and high transmission regions. Therefore, *in-silico* prediction and experimental determination of protein structure are necessary to evaluate *Pfgdv1* as a target candidate for drug design and development.

## 1. Introduction

*Plasmodium falciparum*, the leading causative species of malaria, is a protozoan parasite transmitted by female *Anopheles* mosquitoes when they bite humans to obtain their blood meal. Globally, malaria transmission remains undeterred with epidemiological data showing that malaria affected about 228 million people with approximately 405,000 deaths in 2018 (WHO, 2019). Kenya is one of the malaria-endemic countries in sub-Saharan Africa with highland areas in the North Rift

region, such as Nandi, Uasin Gishu, and Baringo Counties, experiencing epidemic malaria (Kipruto et al., 2017; Noor et al., 2018). Despite great strides made in the use of insecticides, anti-malarial drugs, and effective healthcare services, prevention and control strategies of malaria are still challenging due to the emergence of drug resistance (Arama and Troye-Blomberg, 2014; Delves et al., 2018; Muduli et al., 2018; Sinden et al., 2012; WHO, 2018). Hence, there is a need to develop new strategies that target the transmission of malaria in both epidemic and endemic regions.

\* Corresponding author.

E-mail address: [vatunga@uonbi.ac.ke](mailto:vatunga@uonbi.ac.ke) (V.A. Mobegi).<https://doi.org/10.1016/j.heliyon.2020.e03453>

Received 24 September 2019; Received in revised form 18 January 2020; Accepted 17 February 2020

2405-8440/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Gametocytogenesis is a critical stage in the human-to-mosquito transmission of *P. falciparum*. In sexual development, about 10% of schizonts undergo gametocytogenesis, leading to the formation of gametocytes (Josling and Llinás, 2015). Transmission occurs when a mosquito bites and ingests gametocytes together with the blood meal. In the mosquito midgut, gametocytes grow into macro- and micro-gametes, which fuse to form a zygote (Campbell et al., 2010; Josling and Llinás, 2015). In this view, the mechanism of lifecycle demonstrates that gametocytogenesis plays a critical role in the host-vector transmission of malaria parasites.

Current strategies of malaria control focus on the development of transmission-blocking interventions targeting proteins associated with gametocytes and gametes in humans (Delves et al., 2018; Sinden et al., 2012). Vaccine development and drug design target different stages of parasites, which are gametocyte stage, pre-erythrocytic stage, and erythrocytic stage (Arama and Troye-Blomberg, 2014; Delves et al., 2018; Sinden et al., 2012). Therefore, understanding of the transmission of malaria parasites plays a significant role in the eradication of malaria across the world.

The elucidation of the mechanism of gametocytogenesis forms the basis of developing transmission-blocking vaccines and drugs. Recent studies have established that *P. falciparum* gametocyte development 1 (*Pfgdv1*) protein induces gametocyte development through the mechanism of antagonizing the silencing effect of heterochromatin protein 1 (Campino et al., 2016; Eksi et al., 2012; Filarsky et al., 2018; Josling and Llinás, 2015; Rea et al., 2018). *Pfgdv1* gene (PF3D7\_0935400) codes for a peri-nuclear protein that acts as an upstream activator of gametocytogenesis. Genetic characterizations of this gene and its respective proteins explain their functions and enhance understanding of transmission-blocking targets and gametocyte development in *P. falciparum* isolates.

To achieve the objective of the study, molecular procedures of genomic isolation, amplification, and sequencing generated valid primary sequence data of *Pfgdv1* gene from field isolates of *P. falciparum*. Secondary sequences of *Pfgdv1* retrieved from PlasmoDB (<http://www.plasmodb.org>) provided a pool of global data for comparative analysis. Multiple sequence alignment (MSA) identified the presence of single-nucleotide polymorphisms (SNPs), which cause non-synonymous substitutions and non-conservative changes in amino acids. Tajima's D revealed that the SNPs identified are under balancing selection in the population of *P. falciparum*, while Single-Likelihood Ancestor Counting (SLAC) identified a codon site with the most substantial positive selection. Ultimately, thermodynamic analysis of these SNPs demonstrated that *Pfgdv1* gene codes for a relatively conserved protein, and thus a potential target of drug design to block the transmission of malaria parasites.

## 2. Methods

### 2.1. Study sites

The collection of malaria blood samples was done in Baringo County Referral Hospital (0.48N, 35.74E) in Baringo County, Uasin Gishu County Hospital (0.51N, 35.27E) in Uasin Gishu County, and Kapsabet County Referral Hospital (0.20N, 35.10E) in Nandi County in the high-land areas of North Rift region, Kenya, with altitudes of about 2000m. The laboratory work was carried out at Jomo Kenyatta University of Agriculture and Technology (JKUAT), Biochemistry Laboratory. Sequence data analysis was carried out at the University of Nairobi.

### 2.2. Ethical consideration

The study sought ethical clearance from Institutional Research and Ethics Committee (IREC) of Moi University, College of Health Sciences (MU/CHS) and Moi Teaching & Referral Hospital (MTRH) (Reference no. IREC/2017/41). To comply with ethical requirements, the study

informed patients regarding the objective of the study, sought written informed consent, allowed voluntary participation, and ensured confidentiality of information collected.

### 2.3. Blood sample collection

Thirty blood samples were collected from patients who presented in three hospitals: Uasin Gishu County Hospital (n = 10) and Baringo County Referral Hospital (n = 10) during the months of October and November 2017, and Kapsabet County Referral Hospital (n = 10) during January and February 2018. The study included malaria patients aged between 20 and 45 years but excluded pregnant women, people with mental illnesses, and prisoners.

Rapid diagnostic kits, Paracheck-Pf® (Orchid Biomedical Systems, Goa, India) were used to detect the presence of malaria parasites (Proux et al., 2001). Thin and thick Giemsa-stained blood films were used to confirm a positive diagnosis of malaria and estimate parasitaemia, respectively (Adu-Gyasi et al., 2015). Approximately 2–3 ml of blood was collected from each patient through venipuncture, stored in EDTA vacutainers (Becton Dickinson, Franklin Lakes, NJ), preserved at -4 °C during transport, and stored at -20 °C awaiting DNA extraction. Details of malaria blood samples recorded were age, date, blood volume, parasitaemia, and sample numbers.

### 2.4. Isolation and amplification of Plasmodium DNA

Genomic DNA was extracted from 30 whole blood samples using QIAamp DNA Mini Kit® (Qiagen, Hilden, Germany) following the manufacturer's protocol. The quantity and the quality of the extracted DNA were checked using NanoDrop spectrophotometer (PCRmax Limited, Staffordshire, UK), the nucleic acid ratio (260nm/280nm) (Desjardins and Conklin, 2010), and 1% agarose gel electrophoresis. The extracted DNA samples were stored at -20 °C to preserve them awaiting amplification.

Primers targeting the *Pfgdv1* gene were designed using Primer3Plus (Untergasser et al., 2007). The suitability of the designed primers was checked with the Sequence Manipulation Suite (Stothard, 2000). Parameters considered were primers without hairpin formation, self-annealing, single-base runs, and dinucleotide-base runs, but with substantial GC content, appropriate melting temperatures, and GC clamp. The control primers for the detection of *P. falciparum* DNA targeted a highly conserved 18S rRNA gene as previously described (Mangold et al., 2005) (see Supplementary Table 1).

The PCR set-up and reactions for the control primer and two sets of primers targeting *Pfgdv1* were done separately in the thermal cycler (Eppendorf Mastercycler Gradient 5331, Perkin-Elmer Corporation, Foster City, CA, USA). The control PCR was amplified using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) in a 20 µl reaction volume. The thermal cycling temperatures were an initial denaturation of 98 °C for 30 s, a denaturation of 98 °C for 10 s, an annealing of 54 °C for 30 s, an extension of 72 °C for 30 s, and a final extension of 72 °C for 10 min.

The first fragment of the gametocyte target gene (*Pfgdv1*, 1306 bp) was amplified with a forward primer (PF400\_50F [5'-GTAGGCGTC-GAAATAGTGCT-3']) and reverse primer (PF400\_1355R [5'-TCAG-GATGTGTTATGGTATC-3']) in a 50µl of reaction volume. The PCR conditions were initial denaturation at 98 °C for 30 s, then 35 thermal cycles of denaturation at 98 °C for 10 s, annealing at 54.5 °C for 30 s, and initial extension at 72 °C for 1 min, and eventually a final extension at 72 °C for 10 min. The second fragment of the gametocyte target gene (*Pfgdv1*, 650bp) was amplified with a forward primer (PF400\_1081F [5'-GGTATTCTGTTGTTATGAG-3']) and a reverse primer (PF400\_1731R [5'-AGAAGATGATGAATGCTGACG-3']). Similar thermal cycling conditions with the annealing temperature of 57.5 °C for 30 s were used in the amplification of the fragment in 50µl of reaction volume.

The quality and quantity of amplicons were assessed in a 2% agarose gel electrophoresis and purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. The presence and purity of the PCR products purified were checked using 2% agarose gel electrophoresis and packaged in 96-well plate in 30  $\mu$ l volumes. Sequencing by the Sanger method was outsourced at Macrogen (Macrogen Europe, Amsterdam, Netherlands). Sequence data was submitted to GenBank under accession numbers MN986482-MN986511.

### 2.5. Sequence data analysis

The data analysis entailed descriptive analysis of samples, retrieval of secondary sequences, processing of raw sequencing data, and detection of SNPs. Descriptive statistics were used to highlight the distribution of patients based on study sites, gender, age groups, blood volume, and parasitaemia.

Secondary sequences of *Pfgdv1* were retrieved from PlasmoDB release 46 (<http://PlasmoDB.org>) with 216 *P. falciparum* isolates from 20 countries, predominantly Senegal (70), Gambia (65), Mali (23), French Guiana Region (22), Uganda (11), and Brazil (5) (Aurrecochea et al., 2008; PlasmoDB, 2019). The search strategy entailed the selection of SNPs based on gene identification and retrieval of those with over 80% base call, more than 80% read frequency, and greater than 2% minor allele frequency (MAF). Raw sequencing data in the form of chromatograms (ab1 files) were edited, and contiguous sequences were created from reverse and forward sequences, as well as from overlapping sequences, using ChromasPro® (Technelysium Pty Ltd, South Brisbane, Australia). Visual inspection of chromatograms was done to resolve ambiguities in contiguous sequences based on base-call quality scores. Contiguous sequences created were saved in the multifasta format for further bioinformatics analysis.

MSA was done to detect single nucleotide polymorphisms (SNPs). Nucleotide sequences of primary and secondary data were aligned using MUSCLE in MEGA X (Edgar, 2004; Kumar et al., 2018), and then visually inspected to identify SNPs. Furthermore, the aligned sequences were translated into amino acid sequences using the standard genetic code in MEGA X (Kumar et al., 2018) and visualized in Jalview (2018) to differentiate between synonymous and non-synonymous SNPs, along with conservative and non-conservative mutations. The findings of SNPs were tabulated to indicate coding sequence positions, minor alleles frequency, phenotypes, and effects on amino acids. In selection analysis, Tajima's D in MEGA X was used to determine if these SNPs are under selection (Tajima, 1989; Kumar et al., 2018), while SLAC (<http://datamonkey.org/slac>) (Pond and Frost, 2005) was used to calculate the rates of non-synonymous changes (dN) and synonymous changes (dS). Both the primary data (n = 30) and the secondary data (n = 184) were used in selection analysis.

The effect of SNPs on the predicted protein structure was done using STRUM (<https://zhanglab.cmb.med.umich.edu/STRUM/>), which predicts the stability change of a protein using Gibbs free energy (G). STRUM determines the difference between Gibbs free of folded (Gf) and unfolded (Gu) ( $\Delta G = Gu - Gf$ ), and then the difference between  $\Delta G$  of wild type and mutant protein structure ( $\Delta\Delta G = \Delta G_m - \Delta G_w$ ) (Quan et al., 2016). The thermodynamics rationale is that SNPs affect protein stability by reducing free energy difference between the fold and unfold conformations of a protein, and thus,  $\Delta\Delta G$  less than zero implies that the mutation destabilizes protein structure.

## 3. Results

### 3.1. Characteristic of samples

The study analyzed data from 30 samples of blood obtained from malaria patients (Females = 18, Males = 12), 10 samples from Baringo County Referral Hospital (B1-10), 10 samples from Uasin Gishu County Hospital (U1-10), and 10 samples from Kapsabet County Referral

Hospital (N1-10). The ages of patients ranged from 21 to 40 years (M = 30.47, SD = 5.11) with the highest proportion of patients (33.3%) in the age group of 31–35 years, followed by 26.7% in age group of 26–30 years and 20% in both age groups of 20–25 years and 36–40 years (Table 1). Blood volumes withdrawn from patients fluctuated from 1.5 ml to 4 ml (M = 2.35, SD = 0.59) and parasitaemia varied from 5040 to 80,000 parasites/ $\mu$ l (M = 29,060.67, SD = 22,920.31) (Table 1). Differences in the levels of parasitaemia were statistically insignificant by gender, age groups, and study sites ( $p > 0.05$ ).

Age group comparisons of parasitaemia exhibited positive skew in the distribution with the most variation occurring among patients in the group of 36–40 years followed by those in 20–25 years. In contrast, patients in the age group of 26–30 years had the lowest variation in the distribution of parasitaemia, while those in the age group of 32–35 years had a moderate variation with an outlier (Figure 1). Comparisons of parasitaemia based on gender and study sites depicted disparities in distributions (Figure 1). In samples from Baringo County, the distribution of parasitaemia exhibited positive skew with a similar range of variation in both gender, but males had a higher median than females. A considerable difference exists in the distribution of parasitaemia in samples from Uasin Gishu because females had a higher median and variation than males. Contrastingly, in samples from Nandi County, males had a lower median and variation than females. Overall, the distributions of parasitaemia displayed positive skew with the highest median in Baringo followed by Nandi County and Uasin Gishu.

### 3.2. Quantification of genomic DNA

The concentration of genomic DNA ranged from 8.206 ng/ $\mu$ l to 61.73 ng/ $\mu$ l (M = 30.256, SD = 13.084), while the analysis of the purity of DNA (260 nm/280 nm) indicated that the ratio ranged from 1.714 to 2.176 (M = 1.957, SD = 0.132). The resolution of genomic DNA extracted from the 30 samples obtained from Baringo, Uasin Gishu, and Nandi Counties revealed a band size of approximately 10 kbp (see Supplementary Figure 1). The amplification of the target regions of *Pfgdv1* and resolution of PCR products indicated the expected fragment sizes of 1306 bp (see Supplementary Figure 2) and 650 bp (see Supplementary Figure 3) for the coding regions of 50–1355 and 1081–1731 (see Supplementary Table 1).

### 3.3. Identified single nucleotide polymorphisms

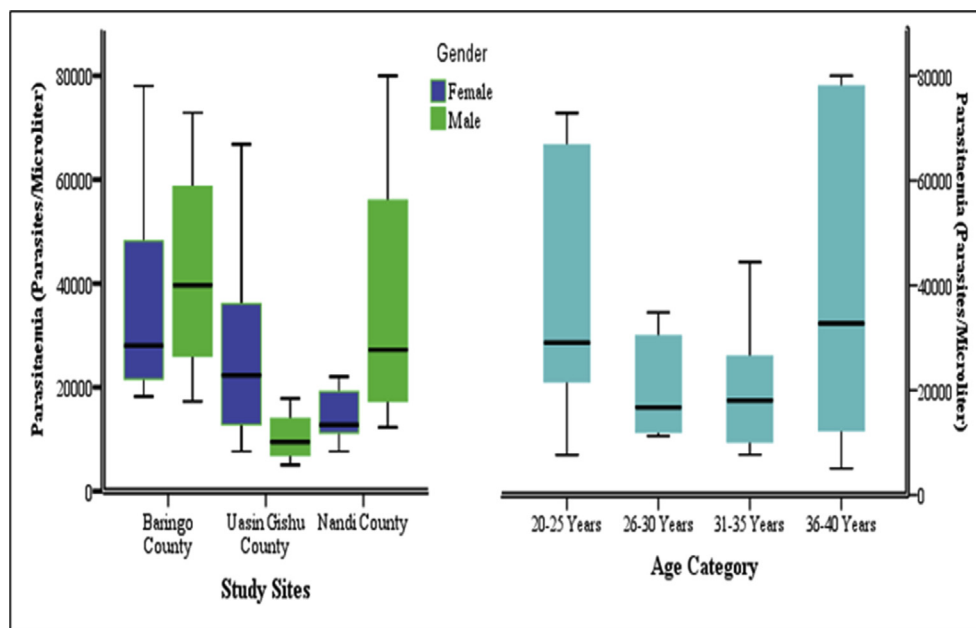
Multiple Sequence Alignment (MSA) of secondary sequences for *Pfgdv1* established the existence of five SNPs (Table 2). Additionally, MSA of primary sequences of *Pfgdv1* corroborated the existence of these five SNPs, namely, c.650C > A, c.1193G > A, c.1249C > A, c.1491T > A, and c.1542A > T (Table 3). These SNPs give rise to four non-synonymous mutations (c.650C > A, c.1193G > A, c.1249C > A, and c.1491T > A) (Figure 2) and one synonymous substitution (c.A1542T). The allelic distribution indicates that mutations of c.650C > A, c.1249C > A, and c.1542A > T form major alleles, whereas those of c.1193G > A and c.1491T > A constitute minor alleles. MSA shows that p.P217H and p.H417N dominated all isolates of *P. falciparum*, while p.R398Q and p.D497E were unique to those isolates obtained from Baringo and Uasin Gishu Counties, correspondingly (Figure 2).

### 3.4. Selection analysis

The analysis of the four-polymorphic sites of nsSNPs indicated that *Pfgdv1* has significant selection ( $p = 0.007$ ), leading to the rejection of the hypothesis of neutrality (Table 4). Results for the primary data (n = 30) showed a positive Tajima's D (0.209) since observed variation ( $\pi = 0.0019$ ) is more than expected variation ( $\Theta = 0.0018$ ), which is statistically significant ( $p = 0.007$ ). Additionally, results obtained from the secondary data (n = 184) revealed a similar trend due to a positive Tajima's D (0.929) with observed variation ( $\pi = 0.0018$ ) that is greater

**Table 1.** Characteristics of malaria patients and samples collected.

Statistics	Age	Blood volume drawn (ml)	Parasitaemia (p/μl)	
Mean	30.47	2.35	29060.67	
Standard Deviation	5.11	0.59	22920.31	
Range	19.00	2.50	74960.00	
Minimum	21.00	1.50	5040.00	
Maximum	40.00	4.00	80000.00	
Count (N)	30.00	30.00	30.00	
Groups	Frequency	Percent (%)	M±SD (p/μl)	Significance
Females	18	60	26434 ± 19020.28	F (1,28) = 0.582, p = 0.452
Males	12	40	33000 ± 27317.04	
20–25 Years	6	20	37846 ± 26433.69	F (3,26) = 1.339, p = 0.283
26–30 Years	8	26.7	20542 ± 9792.89	
31–35 Years	10	33.3	23960 ± 19465.42	
36–40 Years	6	20	40133 ± 33527.06	
Baringo County	10	33.3	39164 ± 21733.83	F (2,27) = 1.712, p = 0.200
Uasin Gishu County	10	33.3	20984 ± 19020.28	
Nandi County	10	33.3	27034 ± 25848.98	



**Figure 1.** The distribution of parasitaemia levels of 30 patients according to age groups, gender, and study sites.

than expected variation ( $\Theta = 0.0012$ ). SLAC shows that p.P217H is under the strongest positive selection in both the primary ( $dN/dS = 361.68$ ) and secondary data ( $dN/dS = 158.98$ ) based on statistically significant model of nucleotide substitution (The General Time Reversible) and stringent significance level ( $p < 0.05$ ) (Tables 5 and 6). Moreover, SLAC indicates that p.R398Q and p.D497E are other nsSNPs that exhibit positive selection but not statistically significant ( $p > 0.3$ ).

### 3.5. Effect of non-synonymous substitutions

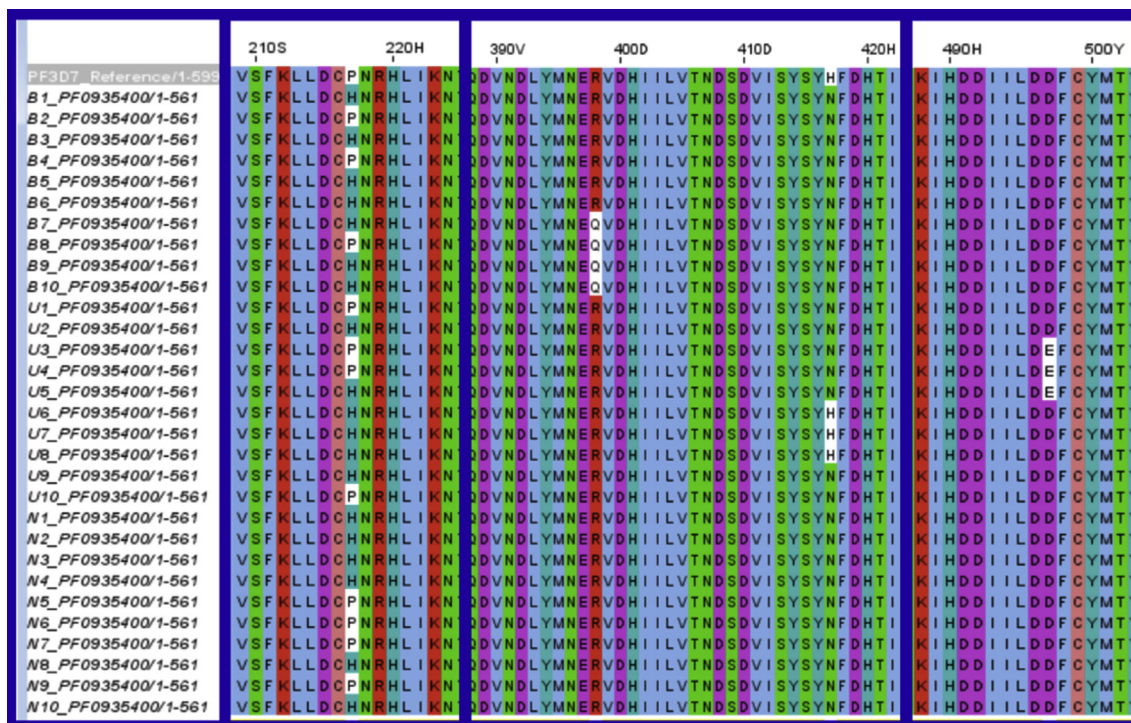
The prediction of protein stability changes revealed that three out of four nsSNPs influence the stability of the *Pfgdv1* protein structure (Table 7). While p.P217H ( $ddG = -2.22$ ) has a destabilizing effect, p.D497E ( $ddG = 1.62$ ) and p.H417N ( $ddG = 0.92$ ) have a stabilizing effect on the protein structure. Since its Gibbs free-energy gap is close to zero, p.

**Table 2.** SNPs identified in *Pfgdv1* among secondary sequences from PlasmoDB.

SNPs	Alleles		Amino Acid Change	Phenotypes		% Calls	Isolate Count
	Reference (3D7)	Minor Allele Frequency (MAF)		Genetic Code	Amino Acid		
c.650C > A	C	C (39%)	p.P217H	non- synonymous	non-conservative	87.6	207
c.1193G > A	G	A (25%)	p.R398Q	non- synonymous	non- conservative	83.9	205
c.1249C > A	C	C (18%)	p.H417N	non- synonymous	non-conservative	83.5	202
c.1491T > A	T	A (10%)	p.D497E	non- synonymous	conservative	89	208
c.1542A > T	A	A (26%)	p.S514S	Synonymous	conservative	90.4	200

**Table 3.** SNPs identified in *Pfgdv1* among primary sequences from Kenyan Isolates.

SNPs	Alleles		Amino Acid Change	Phenotypes		% Calls	Isolate Count
	Reference (3D7)	Minor Allele Frequency (MAF)		Genetic Code	Amino Acid		
c.650C > A	C	C (37%)	p.P217H	non-synonymous	non-conservative	94	30
c.1193G > A	G	A (13%)	p.R398Q	non-synonymous	non-conservative	94	30
c.1249C > A	C	C (10%)	p.H417N	non-synonymous	non-conservative	94	30
c.1491T > A	T	A (10%)	p.D497E	non-synonymous	conservative	94	30
c.1542A > T	A	A (20%)	p.S514S	Synonymous	conservative	94	30



**Figure 2.** MSA outcome visualized using Jalview to highlight four nsSNPs (p.P217H, p.R398Q, p.H417N, and p.D497E) in *Pfgdv1* sequences generated from 30 primary isolates of *P. falciparum* (B1–B10, U1–U10, N1–N10) aligned to the reference sequence (PF3D7).

**Table 4.** Results from Tajima's neutrality test.

M	S	$p_s$	$\Theta$	$\pi$	D
30	4	0.007	0.0018	0.0019	0.209
184	4	0.007	0.0012	0.0018	0.929

Abbreviations:  $m$  = number of sequences,  $S$  = number of polymorphic sites,  $p_s$  = p-value ( $S/n$ ),  $\Theta$  = expected variation ( $p_s/a_1$ ),  $\pi$  = observed variation (nucleotide diversity), and  $D$  = the Tajima test statistic.

R398Q (ddG = 0.07) has no considerable effect on the stability of protein structure.

#### 4. Discussion

##### 4.1. Parasitaemia range in the study sites

This study collected blood samples with high levels of parasitaemia as determined by microscopy using the assumed white blood cell (WBC) count of 8000/ul (Adu-Gyasi et al., 2015). Comparative analysis revealed that parasitaemia levels did not vary statistically significantly based on gender, age group, and study sites. However, descriptive statistics indicated that males, patients

aged 36–40 years, and study sites of Baringo County and Nandi County exhibited high levels of parasitaemia. In a comparative study, researchers established children aged 5–14 years and in lowland areas had higher levels of parasitaemia than those in highland areas but similar variations by gender (Lo et al., 2015). The extent of malaria transmission explains differences in the level of parasitaemia in various regions.

##### 4.2. Gametocyte development 1 gene

Gametocyte development 1 (*Pfgdv1*) plays a major role in the epigenetic regulation of gametocytogenesis in *P. falciparum*. Genetic characterization indicates that *Pfgdv1* (PF3D7\_0935400) is a 1.8kbp gene

**Table 5.** Selection analysis outcomes of primary data.

Local Isolates (n = 30)								
Partition	Sites	Branches			Branch Length		Selected at p ≤ 0.05	
		Tested	Total	Tested	% of total	Total	Positive	Negative
1	561	15	15	0.00233	50.0%	0.00466	1	0
Model fits								
Model	AICC	log L		Parameters	Rate distributions			
Nucleotide GTR	4373.58	-2163.75		23				
Global MG94xREV	4202.49	-2071.06		30	non-synonymous/synonymous rate ratio for *test*			
					100% @ 0.946			
DN/DS Analysis								
Codon Site	ES	EN	S	N	dS	dN	dN-dS	
217	0.63	2.38	0.00	4.00	0.00	1.68	361.68	
398	1.65	1.35	0.00	1.00	0.00	0.74	158.98	
497	0.02	2.98	0.00	1.00	0.00	0.34	72.17	
417	0.00	3.00	0.00	1.00	0.00	0.33	0.00	
514	1.00	1.69	1.00	0.00	1.00	0.00	-214.70	

**Table 6.** Selection analysis outcomes of secondary data.

Global Isolates (n = 184)								
Partition	Sites	Branches			Branch Length		Selected at p ≤ 0.05	
		Tested	Total	Tested	% of total	Total	Positive	Negative
1	599	49	49	0.00333	35.30%	0.00945	1	1
Model fits								
Model	AICC	log L		Parameters	Rate distributions			
Nucleotide GTR	4890.08	-2387.97		57				
Global MG94xREV	4706	-2288.73		64	non-synonymous/synonymous rate ratio for *test*			
					100% @ 0.328			
Codon Site	ES	EN	S	N	dS	dN	dN-dS	
217	0.62	2.38	0.00	4.00	0.00	1.68	177.70	
497	0.02	2.98	0.00	2.00	0.00	0.67	70.94	
398	0.87	2.13	0.00	1.00	0.00	0.47	49.72	
417	0.00	3.00	0.00	5.00	0.00	1.67	0.00	
514	1.00	1.62	5.00	0.00	5.00	0.00	-529.25	

**Table 7.** Gibbs free-energy gaps (ddG) of nsSNPs in *Pfgdv1*.

Amino Acid Position	Wild Type	Mutant Type	ddG
217	P	H	-2.22
398	R	Q	0.07
417	H	N	0.92
497	D	E	1.62

located in the right arm of chromosome 9 in the sub-telomeric region (Campino et al., 2016). As one of the 5,777 genes of *P. falciparum*, *Pfgdv1* is an AT-rich gene with a GC content of 14.4%, which is lower than 19.4% of the overall genome of 3D7 strain (Gardner et al., 2002; Lu et al., 2015; PlasmoDB, 2019). The two set of primers designed (PF\_0935400\_50F [5'-GTAGGCGTCGAAATAGTGCT-3']/PF\_0935400\_1355R [5'-TCAGGATGTGTTATGGTATC-3'], and PF\_0935400\_1081F [5'-GGTATTCCTGTTGTTATGAG-3']/PF\_0935400\_1731R [5'-AGAA-GATGATGAATGCTGACG-3']) targeted overlapping sections of *Pfgdv1* gene with amplified fragment sizes of 1306 bp and 650 bp, respectively, in 30 isolates obtained from the North Rift region of Kenya, namely, Nandi, Uasin Gishu, and Baringo Counties.

### 4.3. Single-nucleotide polymorphism analysis

As a variation at the genetic level, SNP analysis provides a basic way of characterizing genes and improving their understanding. Based on both primary and secondary data, MSA established the existence of SNPs in *Pfgdv1*, which exhibited the proportion that is greater than 1% as per the conventional requirement of a SNP (Karki et al., 2015). This threshold eliminates the consideration of random mutations as SNPs because they rarely occur in various genes. Results showed that *Pfgdv1* has five SNPs comprising of four non-synonymous mutations and a synonymous substitution with variation in the proportion of minor alleles from 10% to 39% in secondary sequences and 10%–37% in primary isolates. The presence of SNPs in high proportions supports the earlier findings that *Pfgdv1* is the molecular marker of gametocytogenesis and transmission level that exhibits geographical divergence (Duffy et al., 2018; Mobegi et al., 2014). Overall, the analysis of SNPs indicates that mutations of c.650C > A, c.1249C > A, and c.1542A > T formed major alleles, whereas those of c.1193G > A and c.1491T > A constituted minor alleles.

The four non-synonymous substitutions identified in *Pfgdv1* had proportions of more than 10% in both global and local populations. The first SNP is a transversion (c.650C > A) that forms a major allele of about 62% and causes the occurrence of non-synonymous substitution

(p.P217H). In a genome-wide analysis study, this SNP is one of the five SNPs that accumulate within the 15kb region of chromosome 9 with the highest value of the fixation index (Mobegi et al., 2014). Moreover, this form of mutation causes a non-conservative change in amino acid from polar proline (P) to positively charged histidine (H). In their study of determining the rate of gametocyte conversion, Usui et al. (2019) established that this mutation increases the effectiveness of gametocyte differentiation in *P. falciparum*.

A subsequent SNP is a transition (c.1193G > A) that causes non-synonymous substitution (p.R398Q) in approximately 19% of minor alleles. This form of substitution results in a non-conservative change in amino acid from positively charged arginine (R) to polar glutamine (Q). There is a transversion (c.1249C > A), leading to the formation of a major allele of around 84% and non-synonymous substitution (p.H417N). Another transverse substitution (c.1491T > A) forms a minor non-synonymous substitution of 10% (p.D497E). The change in amino acid from aspartate (D) to glutamate (E) is conservative because both amino acids have the same properties due to their negatively charged side chains. Conservative amino acids reduce the genetic diversity of a protein, making it a potential target for vaccine development (Srisutham et al., 2018).

Moreover, the synonymous substitution reported in *Pfgdv1*, which is a transversion (c.1542A > T) forming a major allele of approximately 77%, does not affect amino acid change. The degeneracy of the genetic code allows more than one form of codon to code for the same amino acid (Chan et al., 2017). Consequently, synonymous substitutions have no biological significance because their occurrence neither change amino acids nor alter protein functions.

#### 4.4. Implication of selection analysis

The positive values of Tajima's D obtained from both primary and secondary data imply that SNPs on *Pfgdv1* exhibit balancing selection. Tajima's D values were not only positive for both primary isolates (0.209) and secondary isolates (0.929) but also statistically significant ( $p < 0.05$ ). Since the observed variations were higher than the expected variations, it shows that *Pfgdv1* evolve through the process of balancing selection (Tajima, 1989). Genes that exhibit balancing selection in their evolution are target candidates for elucidating immune mechanisms and vaccine development. SLAC analysis pointed out that p.P217H is the only SNP under the strongest positive selection in *Pfgdv1*. In the genome-wide analysis, Mobegi et al. (2014) established that p.P217H is under differential selection between Gambian and Guinean isolates with a fixation index of 0.3. These findings suggest that p.P217H SNP may play a major role in this protein.

#### 4.5. Effect of SNPs on stability of protein

Out of four nsSNP, three of them have a considerable effect on the stability of the predicted protein structure of *Pfgdv1*. A nsSNP causes energy changes in the fold and unfolded conformations of proteins, which have destabilizing or stabilizing effects on protein structure (Quan et al., 2016). Since Gibbs free-energy gap is close to zero, p.D497E has no considerable effect on the thermal stability of the protein structure. Two nsSNPs, p.R398Q and p.H417N, have a stabilizing effect on the protein structure of *Pfgdv1*. Nevertheless, p.P217H is a nsSNP that destabilizes the protein structure because it has a substantial negative free-energy gap ( $\Delta\Delta G = -2.22$ ). In addition, the SNP is a non-synonymous mutation that causes a non-conservative change in amino acid from polar proline (P) to positively charged histidine (H). The predicted destabilization of the protein structure supports earlier findings that this mutation increases the induction of sexual differentiation and determines transmission rates (Mobegi et al., 2014; Usui et al., 2019).

#### 4.6. Mechanism of gametocytogenesis

Gametocytogenesis is an important process in the transmission of *Plasmodium* parasites from humans to mosquitoes. The commitment to gamete production is under a stringent epigenetic control mechanism that allows conversion of 10% of schizonts into transmissible gametes (Josling and Llinás, 2015). In the elucidation of the mechanism of gametocytogenesis, researchers have established critical findings pointing out *Pfgdv1* as a key inducer of transcription (Filarsky et al., 2018; Josling and Llinás, 2015). Preliminary studies elucidated that gametocytogenesis is under the epigenetic control of AP2-G by heterochromatin protein 1 (HP1), which is a transcriptional repressor (Bechti and Waters, 2017; Brancucci et al., 2014). Recent studies exposed that *Pfgdv1* induces gametocytogenesis by removing and lessening the repressive effect of HP1 on AP2-G, the master regulator of transcription in gametocytogenesis (Filarsky et al., 2018; Rea et al., 2018; Usui et al., 2019). Based on the elucidated mechanism of gametocytogenesis, it is apparent that repression or disruption of *Pfgdv1* would cause a constitutive epigenetic silencing due to the stability of HP1. Therefore, SNP characterization highlighted that *Pfgdv1* a relatively conserved protein, making it a potential target in blocking the transmission of *Plasmodium* parasites.

#### 5. Conclusions

In this study, the genetic characterization of the *Pfgdv1* gene reported five SNPs comprising of four non-synonymous substitutions and a synonymous substitution. The presence of four non-synonymous SNPs implies that the *Pfgdv1* gene exhibits some level of diversity among *Plasmodium* species. Furthermore, three non-synonymous SNPs have non-conservative changes on amino acids, which have marked influence on protein structure and function, and thus, suggesting a possible effect on the process of gametocytogenesis. Tajima's D indicated that SNPs identified in *Pfgdv1* exhibit balancing selection in both local and global isolates of *P. falciparum*. Additionally, SLAC analysis indicated that p.P217H is under the strongest positive selection and may account for high transmission rates. Thermodynamic analysis of the changes in Gibbs free energy demonstrated that nsSNPs have both stabilizing and destabilizing effects on *Pfgdv1* protein, indicating a net conservative effect. Thus, these findings could form the basis of *in-silico* prediction and experimental determination of protein structure to target *Pfgdv1* in drug design studies.

#### Declarations

##### Author contribution statement

J. Bungei: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

V. Mobegi: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

S. Nyanjom: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

##### Funding statement

This work was supported by AFRICA-ai-JAPAN PROJECT under research grants of 2017/2018 (JKU/ADM/10B).

##### Competing interest statement

The authors declare no conflict of interest.

## Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2020.e03453>.

## Acknowledgements

We are grateful to the malaria patients who contributed samples to this study and the technical staff of Baringo County Referral Hospital, Uasin Gishu County Hospital, and Kapsabet County Referral Hospital for support in sample collection. The authors thank Jomo Kenyatta University of Agriculture & Technology for providing required laboratory resources.

## References

- Adu-Gyasi, D., Asante, K.P., Newton, S., Amoako, S., Dosoo, D., Ankrah, L., Adjei, G., Amenga-Etego, S., Owusu-Agyei, S., 2015. Malaria parasite density estimated with white blood cells count reference value agrees with density estimated with absolute in children less than 5 years in central Ghana. *Malar. Res. Treat* 2015, 1–8.
- Arama, C., Troye-Blomberg, M., 2014. The path of malaria vaccine development: challenges and perspectives. *J. Intern. Med.* 275, 456–466.
- Aurrecochea, C., Ross, C., Pennington, C., Nayak, V., Heiges, M., Gao, X., Kraemer, E., Li, W., Gingle, A., Iodice, J., Kissinger, J.C., Treatman, C., Brunk, B.P., Stoeckert, C.J., Grant, G., Miller, J.A., Gajria, B., Pinney, D.F., Innamorato, F., Roos, D.S., Fischer, S., Wang, H., Brestelli, J., Harb, O.S., Dommer, J., 2008. A functional genomic database for malaria parasites. *PLoS Pathog.* 4, e312.
- Bechtsi, D.P., Waters, A.P., 2017. Genomics and epigenetics of sexual commitment in *Plasmodium*. *Int. J. Parasitol.* 47, 425–434.
- Brancucci, N.M.B., Bertschi, N.L., Zhu, L., Niederwieser, I., Chin, W.H., Wampfler, R., Freymond, C., Rottmann, M., Felger, I., Bozdech, Z., Voss, T.S., 2014. Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe* 16, 165–176.
- Campbell, T.L., de Silva, E.K., Olszewski, K.L., Elemento, O., Llinás, M., 2010. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* 6, 1–15.
- Campino, S., Benavente, E.D., Assefa, S., Thompson, E., Drought, L.G., Taylor, C.J., Gorvett, Z., Carret, C.K., Flueck, C., Ivens, A.C., Kwiatkowski, D.P., Alano, P., Baker, D.A., Clark, T.G., 2016. Genomic variation in two gametocyte non-producing *Plasmodium falciparum* clonal lines. *Malar. J.* 15, 1–10.
- Chan, S., Ch'ng, J.H., Wahlgren, M., Thutkawkorapin, J., 2017. Frequent GU wobble pairings reduce translation efficiency in *Plasmodium falciparum*. *Sci. Rep.* 7, 1–14.
- Delves, M.J., Angrisano, F., Blagborough, A.M., 2018. Antimalarial transmission-blocking interventions: past, present, and future. *Trends Parasitol.* 34, 735–746.
- Desjardins, P., Conklin, D., 2010. Nanodrop microvolume quantification of nucleic acids. *J. Vis. Exp.* 45, 1–4.
- Duffy, C.W., Amambua-Ngwa, A., Ahouidi, A.D., Diakite, M., Awandare, G.A., Ba, H., Tarr, S.J., Murray, L., Stewart, L.B., D'Alessandro, U., Otto, T.D., Kwiatkowski, D.P., Conway, D.J., 2018. Multi-population genomic analysis of malaria parasites indicates local selection and differentiation at the *gdfv1* locus regulating sexual development. *Sci. Rep.* 8.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.* 5, 1–19.
- Eksi, S., Morahan, B.J., Haile, Y., Furuya, T., Jiang, H., Ali, O., Xu, H., Kiattitubtr, K., Suri, A., Czesny, B., Adeyemo, A., Myers, T.G., Sattabongkot, J., Su, X., zhuan, Williamson, K.C., 2012. *Plasmodium falciparum* gametocyte development 1 (*Pf*gdfv1) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog.* 8, 1–14.
- Filarsky, M., Frschka, S.A., Niederwieser, I., Brancucci, N.M.B., Carrington, E., Carrió, E., Moes, S., Jenoe, P., Bártfai, R., Voss, T.S., 2018. GDFV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* 359, 1259–1263 (80-).
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M.A., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Davis, R.W., Fraser, C.M., Barrell, B., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511.
- Jalview, 2018. Jalview 2.10.5 [WWW document]. URL: <http://www.jalview.org/development/release-history/Jalview-2105>.
- Josling, G.A., Llinás, M., 2015. Sexual development in plasmodium parasites: knowing when it's time to commit, nature reviews microbiology. Nature Publishing Group.
- Karki, R., Pandya, D., Elston, R.C., Ferlini, C., 2015. Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC Med. Genomics* 8, 1–7.
- Kipruto, E.K., Ochieng, A.O., Anyona, D.N., Mbalanya, M., Mutua, E.N., Onguru, D., Nyamongo, I.K., Estambale, B.B.A., 2017. Effect of climatic variability on malaria trends in Baringo County, Kenya. *Malar. J.* 16 (220).
- Kumar, S., Stecher, G., Li, M., Nnyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549.
- Lo, E., Zhou, G., Oo, W., Afrane, Y., Githeko, A., Yan, G., 2015. Low parasitemia in submicroscopic infections significantly impacts malaria diagnostic sensitivity in the highlands of Western Kenya. *PLoS One* 10, 1–15.
- Lu, X.M., Bunnik, E.M., Pokhriyal, N., Nasser, S., Lonardi, S., Le Roch, K.G., 2015. Analysis of nucleosome positioning landscapes enables gene discovery in the human malaria parasite *Plasmodium falciparum*. *BMC Genomics*.
- Mangold, K.A., Manson, R.U., Koay, E.S.C.C., Stephens, L., Regner, M.A., Thomson, R.B., Peterson, L.R., Kaul, K.L., 2005. Real-time PCR for detection and identification of *Plasmodium* spp. *J. Clin. Microbiol.* 43, 2435–2440.
- Mobegi, V.A., Duffy, C.W., Amambua-Ngwa, A., Loua, K.M., Laman, E., Nwakanma, D.C., MacInnis, B., Aspelting-Jones, H., Murray, L., Clark, T.G., Kwiatkowski, D.P., Conway, D.J., 2014. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol. Biol. Evol.* 31, 1490–1499.
- Muduli, S., Pandey, P., Devatha, G., Babar, R., Thirupanthaka, M., Kothari, D.C., Kabir, M., Pillai, P.P., Ogale, S., 2018. Photoluminescence quenching in self-assembled CsPbBr<sub>3</sub> quantum dots on few-layer black phosphorus sheets [WWW document]. In: *Chemie - Int. Angew.* (Ed.).
- Noor, A.M., Kiptui, R., Waqo, E., Snow, R.W., Macharia, P.M., Giorgi, E., Okiro, E.A., 2018. Spatio-temporal analysis of *Plasmodium falciparum* prevalence to understand the past and chart the future of malaria control in Kenya. *Malar. J.* 17, 1–13.
- PlasmoDB, 2019. *Plasmodium* genomic resource [WWW Document]. URL: <http://plasmodb.org/plasmo/>.
- Pond, S.L.K., Frost, S.D.W., 2005. Frost, not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222.
- Proux, S., Hkiriareon, L., Ngamngonkiri, C., McConnell, S., Nosten, F., 2001. Short communication: Paracheck-Pf®: a new, inexpensive and reliable rapid test for *P. falciparum* malaria. *Trop. Med. Int. Health* 6, 99–101.
- Quan, L., Lv, Q., Zhang, Y., 2016. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* 32, 2936–2946.
- Rea, E., Le Roch, K.G., Tewari, R., 2018. Sex in *Plasmodium falciparum*: silence play between GDFV1 and HP1. *Trends Parasitol.* 34, 450–452.
- Sinden, R.E., Carter, R., Drakeley, C., Leroy, D., 2012. The biology of sexual development of *Plasmodium*: the design and implementation of transmission-blocking strategies. *Malar. J.* 11, 1–11.
- Srisutham, S., Saralamba, N., Sriprawat, K., Mayxay, M., Smithuis, F., Nosten, F., Pukrittayakamee, S., Day, N.P.J., Dondorp, A.M., Imwong, M., 2018. Genetic diversity of three surface protein genes in *Plasmodium malariae* from three Asian countries. *Malar. J.* 17, 1–10.
- Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102–1104.
- Tajima, F., 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, 71–74.
- Usui, M., Prajapati, S.K., Ayanful-Torgby, R., Acquah, F.K., Cudjoe, E., Kakaney, C., Amponsah, J.A., Obboh, E.K., Reddy, D.K., Barbeau, M.C., Simons, L.M., Czesny, B., Raiciulescu, S., Olsen, C., Abuaku, B.K., Amoah, L.E., Williamson, K.C., 2019. *Plasmodium falciparum* sexual differentiation in malaria patients is associated with host factors and GDFV1-dependent genes. *Nat. Commun.* 10, 1–15.
- WHO, 2019. World malaria report 2019 [WWW Document]. URL: <https://www.who.int/malaria/publications/world-malaria-report-2019/en/>.
- WHO, 2018. Responding to antimalarial drug resistance [WWW Document]. URL: [https://www.who.int/malaria/areas/drug\\_resistance/overview/en/](https://www.who.int/malaria/areas/drug_resistance/overview/en/).