Absolute quantification of gene copy number variation in Plasmodium falciparum novel putative genes associated to drug resistance (the actin-binding coronin protein, cysteine desulfurase and plasmepsin 2), using Eva green-based quantitative PCR

Houdou Diarra (houdou.diar@gmail.com)
Pan African University (PAUSTI)
Victor A. Mobegi
University of Nairobi
Edward E. Makhulu
International Centre of Insect Physiology and Ecology (ICIPE), Nairobi, Kenya.
Johnson Kinyua
Jomo Kenyatta University of Agriculture and Technology (JKUAT)
Jeremy K Herren
International Centre of Insect Physiology and Ecology (ICIPE), Nairobi, Kenya.

Research Article

Keywords: Plasmodium falciparum, antimalarials pressure, CNVs, drug resistance

Posted Date: March 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1451928/v1

Abstract

Background: Exposure of *Plasmodium falciparum* to host immune system and antimalarials pressure leads the parasite to exploit genetic factors such as genome sequence variation (structural variation and single nucleotide polymorphism) to withstand environmental conditions. Gene copy number variation (CNVs) is a type of genome sequence polymorphism and it has been shown that CNVs contribute to parasite adaptation to environmental pressure, acquisition of drug resistance and enable the parasite to escape from host immune system. In this study, we explored CNVs in *P.falciparum* three putative genes related to drug pressure.

Method: Blood samples were obtained from asymptomatic school children collected from December 2016 to October 2018 in Mbita sub-county, western Kenya. Genomic DNA was extracted using ISOLATE II Genomic DNA kit (Bioline, UK) following the manufacturer's instructions. Real-time quantitative PCR was performed and the target genes copy numbers were determined by absolute quantification method.

Results: The findings revealed differential level of copy number variation in target genes investigated. One isolate (KE024/18) was found to contain 5.29 copies of *Pfcoronin gene,* t(2) = 27.91, *P* = 0.0013, 95% CI [3.62 - 4.95]. In the *Pfcysteine desulfurase* gene, similar level of copy amplification was observed in the OS0149/18, (5.09 copies) t(2) = 7.85, *P* = 0.0148, 95% CI [1.85 – 6.33]. These two isolates were found to harbour the highest level of CNVs of the corresponding genes. *Pfplasmepsin 2* gene was found to be less polymorphic, however, three isolates showed high copy amplification, OS062/17 (9.63 copies), t(2) = 80.50, *P* = 0.0002, 95% CI [8.01 – 9.09]; AL106/16 (9.18 copies), t(2) = 28.61, *P* = 0.0012, 95% CI [6.95 – 9.41], and KM019/18 (9.41 copies), t(2) = 90.51, *P* = 0.0001, 95% CI [8.01 – 8.81].

Conclusion: The results show gene copy number variation in the clinical samples we explored. The findings suggest that *P.falciparum* may be under drug pressure in the study area, therefore, this calls for close surveillance of drug resistance.

Introduction

Plasmodium falciparum, the most lethal species of human malaria parasites can develop resistance to antimalarial drugs and escape the host immune system, becoming a challenge to drugs efficacy. This ability mostly results from the variation of *P.falciparum* genome that imparts to the parasite the capacity to survive under environmental conditions [1, 2]. Changes in the *P.falciparum* parasite genome include single nucleotide variation, insertion, deletion which may lead to alteration of chromosomes structure [1]. In addition, gene copy number variation (CNV) largely contribute to genome variability and has forthcome as one of the common causes of genetic diversity [3, 4]. CNV consists of the amplification or deletion of a single gene or a cluster of adjacent genes [5], resulting from the processes of rearrangements of the genome such as translocations, duplications, deletions and inversions leading to diverse types of CNVs, from tandem amplifications to the gains or losses of more complex sequences across the genome, [6]. CNVs are less frequent when compared to single nucleotide polymorphisms (SNPs), but they are more

likely to influence phenotype diversity [7]. The role played copy number variants in addition to genetic diversity, is that CNVs may directly impact gene expression, impairing gene dosage and may indirectly modify chromatin environment thus leading to disrupted protein expression [8]. Therefore, CNVs may exert an influence to clinical isolate parasites' phenotypes as *in vitro* studies have shown their role in the gain of parasites' fitness notably erythrocyte invasion, cytoadherence, transmissibility and drug resistance [1]. Several *in vitro* studies have demonstrated the role of *P. falciparum* CNVs in evolution, adaptation and diseases in the host organism [9]. Laboratory adapted *P.falciparum* culture have shown many copy number variants in the *P.falciparum* parasite' genome [10, 11, 12], and the most observed CNVs in these studies was a deletion on chromosome 9 in a region containing genes involved in gametocytes formation, and in the transmission to hosts [1, 13], and a deletion on chromosome 2 spanning the knob-associated histidine-rich protein gene (KAHRP) known to mediate the cytoadherence, thus allowing merozoites to avoid moving through bloodstream toward the spleen where they would be demolished [14]. Another *in vitro* study of CNVs has detected duplication of reticulocyte-binding protein-1 gene (rh1) [15, 16], implicated in erythrocyte invasion [15]. Moreover, increased copy numbers of *P. falciparum* multiple-drug resistance gene 1 (*Pfmdr1*) has been shown to result in high levels of *Pfmdr1* protein expression and was associated with a decline of MQ and halofantrine susceptibility [17]. Furthermore, amplification of the GTP- cyclohydrolase I (*Pfgch1*) protein, has been reported to be associated with antifolate drugs resistance [18]. Amplification of *P. falciparum*' vacuole cysteine proteases falcipain 2 and falcipain 3 genes, both associated with artemisinin resistance [19], and involved in haemoglobin breakdown have also been reported [20]. Moreover, copy number variations have been observed in field parasite's isolates. Lumefantrine resistance has been observed in clinical isolates with increased copy number of multidrug resistance 1 (*Pfmdr1*) gene [17, 21]. Increased copy number of plasmepsin 2 (*Pfpm2*) gene located on chromosome 14, is associated with piperaquine resistance and has been found in field parasites, [22, 23], while high copy number of *Pfmdr1* associated with lumefantrine pressure has been also observed [24, 25]. Together, both *in vivo* and *in vitro* reported studies strongly support that CNVs are involved in parasite adaptation to antimalarial drugs. Indeed, about 0.3–6% of *P.falciparum* genome is subjected to copy number variation, a fraction greater than the prevalence of SNPs [1].

Determination of copy number variation to assess the role they play requires reliable methods of accurate quantification. Quantitative PCR (qPCR) is highly sensitive and specific and is commonly used as an accurate method where accurate quantification, for instance gene copy number or the level of gene expression, is required [26]. Absolute and relative quantification are the most routinely used approaches in gene copy number calculation. Both methods require the use of a calibrator sample that carries the gene of interest and serves as a normalizer, and the concentration of the calibrator sample must be determined, [27]. The test sample and the calibrator need to be run in parallel under the same PCR reactions conditions. Standards samples such as genomic DNA, recombinant plasmid DNA, or synthetic DNA fragment can be used as calibrator samples for copy number determination when they carry the target gene [27]. Moreover, relative quantification required a reference gene (housekeeping gene) used to calculate the fold gene in test samples relative to the reference gene. Conversely, absolute quantification

uses a standard curve constructed from a serial dilution of the calibrator sample to determine the exact concentration of the target gene in an unknown concentration sample relative to the CT values generated during the reaction, [28, 27, 29]. Compared to relative quantification, absolute quantification provides some advantages such as sensitivity, and high reproducibility of data [30]. In the presen_t study, we performed absolute quantification to assess copy number variation. This study aimed to determine copy number variation in three candidate genes: *P.falciparum* actin-binding coronin protein gene (PF3D7_1251200), the cysteine desulfurase gene (PF3D7_0727200), and plasmepsin 2 gene (PF3D7_1408000). Our interest in these genes relies upon their implication in drug resistance, [31, 32, 33, 23]. Artemisinin based-combination therapy (ACT) is in use in Kenya since 2004, [34], therefore, the assessment of copy number variation of putative genes will help the early detection of ACT drug pressure. The findings will be crucial in the management of malaria and inform the control strategies.

Method

Ethics

Study subjects or parents/legal guardians provided written informed assent prior to participant enrollment. This study was conducted under the ethical grant of the original study (KEMRI/RES/7/3/1), obtained from the Kenya Medical Research Institute (KEMRI). Assent to reuse the samples was part of the original grant form. The experiments were carried out in compliance with experimental guidelines, and laboratory relevant regulations.

Study area.

This study concerned Mbita sub-county located in western Kenya, on the edge of Lake Victoria [35]. Mbita sub-county is part of Homa-Bay County, lay from west, north, and south by the Lake Victoria and situated between latitudes (0° 21' and 0° 32' S), and longitudes (34° 04' to 34° 24' E) [36]. The district has superficies around 163.28 km^2, with a population of approximately 115,896, and is about 400 km west of Nairobi [35]. The area annually experiences an approximated rainfall of 1,300mm, and daily temperatures are estimated from 26°C to 34°C [37]. Mbita sub-county is characterized as one of the high transmission settings of malaria in Kenya, with *P. falciparum* infection estimated at 40% among the residents [38, 39].

Blood sampling

Samples used in the present study were obtained from asymptomatic schools children (5 to 15 years), collected from December 2016 to October 2018. The inclusion criteria in the study were the range of age, no malaria symptoms, and written consent. Participants' parents or guardians provided written consent before participant enrollment. Blood sample collection was performed at the Thomas Odhiambo Campus of the International Centre for Insect Physiology & Ecology (ICIPE) in Mbita. Venous blood (4 mL) was collected in EDTA from each participant. The rapid diagnostic tests (RDTs) (SD Bioline malaria Ag Pf/Pan

(HRP-II/pLDH), and microscopy observation were used to screen samples for the presence of *Plasmodium* parasite. Drops of blood, stained using 10% Giemsa for 10 minutes were prepared for microscopy analysis. The blood smears were analyzed by the microscopists. Subsequently, *Plasmodium*-positive samples confirmed using RDT and microscopy examination were selected, and conserved at −80°C.

DNA preparation

Genomic DNA was isolated from 85 blood samples (200 µl) using ISOLATE II Genomic DNA kit (Bioline, UK) as per the manufacturer's guideline. The DNA samples were stored at − 20°C until used.

Quantitative real-time PCR amplification

Three primer pairs (Table 1.) specific to the target genes were designed in this study. Quantitative PCR provides high sensitive amplification when the length of the amplicons are shorter. To this end, primers were designed such that the size of the amplicons ranged from 233 to 277bp. *Plasmodium falciparum* 3D7 genome (ID: GCA_000002765), was used as reference for primer designing. For sensitive amplification, prior primers designing, the templates that were amplified were firstly assessed for GC% using Beacons Designer tool (http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1). The evaluation of the GC content and the sequence quality of primers and the target template are important as both primer and template can affect the amplification efficiency. Primers were designed using the online Primer3Plus tool (https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The selected primers sequences were checked for their quality using the OligoAnalyzer tool, available on the Integrated DNA Technology (IDT) website (https://www.idtdna.com/pages/tools/oligoanalyzer), where primer GC content, hairpin formation, secondary structures and heterodimers formation were assessed. The Oligosequences were subsequently checked for specificity using the National Center for Biotechnology Information (NCBI) primer BLAST. *Plasmodium falciparum* strain 3D7 genomic DNA (ATCC PRA-405D) obtained from (https://www.atcc.org), was used as calibrator sample. The 3D7 strain harbors single copy of *Pfcoronin*, *Pfpm2* and *Pfcysteine desulfurase IscS* genes [31, 22, 40]. The concentration of the calibrator sample (13.34ng/µL) was determined using NanoDrop 2000C (Thermo Fisher Scientific, USA), this was considered as the stock solution. By using molarity calculation (C_1 xV_1 = C_2 xV_2), a working solution was prepared from the stock which concentration was 1ng/µL in a total volume of 70 µL. Six-fold serial dilutions of the calibrator sample ranging from $4x10^3$ to 0.04 copies/µL of the target genes were prepared using the working solution (Fig. 1.), (Table 2.). A standard curve was constructed for each target gene (Fig. 4), by running the six-fold dilutions in triplicate under the same PCR conditions in parallel with the test samples. A total of 36 samples (14 for *Pfcysteine desulfurase*, 13 for *Pfpm2* and 9 for *Pfcoronin*) were assessed for copy number variation. These samples were selected based on their sequencing results that were found to harbour significant mutations among which some have been reported to be associated with ACT drug reduced susceptibility [41]. The amplification was carried out using Mic qPCR Cycler, v2.8.1 (Biomolecular systems, Australia). The reaction was run in 20 µL

containing 4 µL of 5X HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne, Estonia), 0.3 µM of forward and reverse primers (Macrogen, South Korea), 2 µL of DNA sample and 13.4 µL of nuclease-free water. The run profile was set up as follow: hold at 95°C for 15mn, followed by 40 cycles of 95°C for 30s, 56.7°C for 30s, 72°C for 30s. The same run profile was maintained for the three target genes, and all test samples were run in triplicates in parallel with the calibrator serially diluted samples. PCR negative control (a non-template sample) and positive control (PC) were included in each run.

Table 1
Primer used for amplification of target genes

Gene	Primer sequences (5'——> 3')	Tm	Ta	Product size
Pfcoronin	F: AGGTAAAGGTGATGGTAATTGTCG	61.8°C	56.7°C	277
	R: GTTTGTACTACGTTCAGGATCTCT	61.8°C		
Pfcysteine	F: CTGCTGGAAAAGTCCCTATCG	61.3°C	56.7°C	233
	R: AAGAACATACCTTAGCAGCTTCG	61.1°C		
Pfpm2	F: TATGTGTCAGGAACTGTTAGTGG	61.1°C	56.7°C	249
	R: AGGTAAGTAAAAGGTGAAAAGAGCA	60.9°C		

Tm and **Ta** represent primers melting and annealing temperatures respectively. Forward and reverse primers are indicated by **F** and **R.**

Serial dilution of the calibrator sample

The mass DNA of the calibrator sample needed to construct the standard curves was determined based on the copy number of interest of the target genes. Single copy of the target genes is present in the calibrator sample (*P.falciparum* 3D7 strain) which genome size is 23000000bp [42]. The 23 Mb represent the haploid genome of *P.falciaparum* [42], therefore, the target genes have single copy in the haploid genome of the 3D7 strain. To determine the copy of interest of the target genes needed to construct the standard curves, the mass of the haploid genome was calculated as described by Tajebe et al., [43] using the following equation.

$$m = [\mathrm{n(bp)}]\left[\frac{1}{6.023\mathrm{x}10^{23}\mathrm{molecules/mole}}\right]\left[\frac{660\mathrm{g}}{\mathrm{mole}}\right] = [\mathrm{n}]\left[1.096\mathrm{x}10^{-21}\frac{\mathrm{g}}{\mathrm{bp}}\right]$$

Where m represent the mass, n is the genome size (bp), Avogadro's number equals $6.023\mathrm{x}10^{23}$ molecules/mole, and 660g/mole is the average molecular weight of double-stranded DNA.

To calculate the mass of the haploid genome, 23000000bp was entered for n-value.

$$m = \left[23\text{x}10^{6}\text{bp}\right]\left[1.096\text{x}10^{-21}\frac{\text{g}}{\text{bp}}\right] = 2.5\text{x}10^{-14}\text{g}$$

$$m = 2.5\text{x}10^{-14}\text{g} = 0.000025\text{ng}$$

The mass of the haploid genome of the calibrator sample is 0.000025ng, and this mass contains single copy of each the target gene. Subsequently, the determination of the mass DNA that is needed for the copy number of interest ranging from $4\text{x}10^{3}$ to 0.04 copies was calculated. For instance, if single copy of each the target gene is obtained from 0.000025ng haploid DNA, then 4000 copies can be obtained from 0.1ng. The formula is:

[Copy number of interest] x [mass of haploid genome] = mass of DNA needed.

The above calculation was performed to determine the copy number of interest of the target genes in the six-fold serial dilution (Fig. 1). The standard curves constructed were run in the range of the calculated copy of interest. The copy number of interest of each standard dilution is shown in (Table 2.).

Table 2
Concentrations of the calibrator sample obtained from six-fold serial dilution and the corresponding copy of interest.

Serial dilution	Volume (µL)	Concentration (ng/µL)	Copy of interest (μL^{-1})
1	10	0.1	4000
2	10	0.01	400
4	10	0.0001	4
5	10	0.00001	0.4
6	10	0.000001	0.04

Melting Curves Analysis

Melting curves analysis allows the assessment of the specificity of PCR assay. It confirms that only the target templates were amplified. Specific PCR amplification generates single melting peak [5]. Conversely, multiple peaks are interpreted as nonspecific amplification or the formation of primer-dimers [44]. This interpretation, although generally supported, is prone to errors since the multiple peaks profiles are not always representative of unspecific amplification and or primer-dimers. Indeed, when a double-stranded DNA melts, the GC-rich regions do not dissociate directly, thus, these stable bases maintain their double-stranded conformation until the melting temperature (Tm) sufficiently increases to melt the bonds. Such

cases result in two melting phases and lead to multiple peaks [45]. Factors such as the presence of secondary cross structure in the amplicons sequence and amplicons sequence misalignment (AT-rich fragments) can also lead to multiple peaks profile [45] Primer-dimers are not desirable but can be controlled. They mostly occur in the non-template control (NTC) where primers are abundant and there is no template, although their presence in the NTC tube indicates that they also occur in the template reactions. A non-template control tube was included in each the reaction to check dimers formation by comparing the temperature profile of the NTC to that of the samples, as the melting temperatures of the templates are higher than primer-dimers melting temperature due the longer length of the amplicons.

The melting curves generated from the amplification of the *Pfcoronin* and *Pfcysteine* genes, showed no amplification in the NTC tubes and single peaks were observed in the test tubes (Fig. 2. A1 & A2), (Fig. 3. B1&B2). Although there was no peak in the NTC included in *Pfplasmepsin* reaction, smaller peaks at temperature ranging from 75°C to 76.5°C were observed in the melting profile of some test tubes of the *Pfplasmepsin* gene (Fig. 3. C1&C2), then corresponding to primer-dimers as the Tm of the amplicons were ranged from 78°C to 79°C. Additional peaks were observed (Tm ranges from 83 to 84), slightly higher than the Tm of the amplicons. These additional peaks might be a result of the amplicons structures, mostly due to the GC-rich content of the template sequence.

Amplification efficiency, and standard curves,

The amplification efficiency represents the rate at which amplicons are generated. PCR efficiency evaluates the performance of the real-time PCR reaction. The efficiency, $E = (10^{-1/slope} - 1) \times 100$, is calculated using the slope of the standard curves, and the efficiency equals 100% when the slope is -3.32. More negative slope indicates low efficiency (< 100%) while more positive slope results from sample quality or pipetting errors. Poor amplification efficiency or high difference in efficiency (≥ 4%) between replicates strongly affects the copy number calculation [3]. In practice the efficiency at 100% is recommended, however, the acceptable PCR efficiency is ranged from 90–110% [27, 5]. The amplification efficiencies for each reaction of the triplicate run of the standards dilutions and the test samples were automatically calculated by the Mic software after the reactions were completed. The high difference in the amplification efficiency observed between the replicates was 3%. The amplification efficiencies for all the reactions were in the acceptable range of the PCR efficiency.

A standard curve was automatically built by the Mic PCR software for each target gene amplification (Fig. 4.). The equation ($y = mx + b$) on the top of the corresponding graphs represents the formula of the linear regression line, where m is the slope, x is the DNA quantity and b represents the intercept. The slopes values and the efficiencies were − 3.58 (E = 90.23%) and − 3.32 (E = 100%) for *Pfcoronin* and *Pfcysteine* respectively (Fig. 4. A & B), while *Pfplasmepsin 2* had slope of -3.36 (E = 99.00%) (Fig. 4. C). The coefficient of correlation (R^2) determines the linearity of the data and should be > 0.99 50 [46], and for an assay to be valid, the quantity of the test samples should be within the range of the linear dilution series

[43]. The coefficient of correlations were R^2 = 0.9966; R^2 = 09980, and R^2 = 0.9882 for *Pfcoronin, Pfcysteine* and *Pfplasmesin 2* respectively.

Determination of the target genes copy number variation

The concentration of the amplified target genes was generated automatically for each test sample by the Mic qPCR software based on the concentration of the standard dilution. The copy numbers of the target genes in test samples were calculated from the generated concentration (ng/µl) of input DNA as described by Tajebe et al., [43], using the following formula:

$$\mathrm{DNACopynumber} = \frac{\left[6.023 \times 10^{23} \frac{\mathrm{copy}}{\mathrm{mol}}\right] \times \left[\mathrm{DNAamount}\left(\frac{\mathrm{g}}{\mu\mathrm{l}}\right)\right]}{[\mathrm{Ampliconlength(bp)}] \times \left[\frac{660\mathrm{g}}{\mathrm{mol}} \times \frac{1}{\mathrm{bp}}\right]}$$

Where Avogadro's number represents 6.023×10^{23} copy per mole; 660g per mole per base pair is the molecular weight of DNA. The amplicons lengths are 277bp, 233bp and 249bp for *Pfcoronin*, *Pfcystein*e and *Pfplasmepsin 2* respectively.

The above formula has also been described by Lee et al., [28] and by Kamau et al., [27] for copy number calculation. To determine copy number variation, the ratio of the target gene copies to the calibrator copies was calculated. In other words, the copy numbers variation was calculated by dividing the copy number of the target genes found in the test samples to copy numbers of the target genes in the calibrator samples. For example, in the *Pfcoronin gene*, the copy number of the calibrator at 0.1ng/µL input DNA, and the positive control (PC = 0.133ng/µL, (concentration determined by the PCR software), is computed as follow:

$$\mathrm{Calibratorcopynumber} = \frac{\left[6.023 \times 10^{23} \frac{\mathrm{copy}}{\mathrm{mol}}\right] \times \left[1\mathrm{x}10^{-10}\left(\frac{\mathrm{g}}{\mu\mathrm{l}}\right)\right]}{[277(\mathrm{bp})] \times \left[\frac{660\mathrm{g}}{\mathrm{mol}} \times \frac{1}{\mathrm{bp}}\right]}$$

Cal. = 329449732 copies μL^{-1}

$$\text{Positivecontrolcopynumber} = \frac{\left[6.023 \times 10^{23} \frac{\text{copy}}{\text{mol}}\right] \times \left[\left(1.33\text{x}10^{-10} \frac{\text{g}}{\mu\text{l}}\right)\right]}{[277(\text{bp})] \times \left[\frac{660\text{g}}{\text{mol}} \times \frac{1}{\text{bp}}\right]}$$

PC = 438168143.5 copies μL^{-1}

$$\text{TheratioPCtoCal} = \frac{\text{PC}}{\text{Cal.}} = \frac{438168143.5}{329449732} = 1.33\text{folddifference}$$

The result obtained from the ratio represents the fold change or copy number variation between the test samples relative to the calibrator sample. The copy number variation in the test samples was calculated the same way. The calibrator harbours single copy of the target genes, then if the test samples have more copies of the target genes than the calibrator, the result of the ratio was > 1. Conversely, the ratio ≤ 1 indicates equal or fewer copy numbers of the target genes in the test samples. The ration (R) was computed for each replicate, thus, for the three replicates run, three values of (R) were obtained from each sample. To narrow down the assignment of copy number variation, we defined a cutoff value as described by Leroy et al., [47]: Single copy, and two or multicopy were defined as, the ratio (R < 1.5 and R ≥ 1.5) respectively.

Data analysis

The equation of gene copy number calculation mentioned above was computed in Microsoft Excel 2019, to calculate the target genes copy numbers. The gene copies were calculated per sample from the replicates concentration generated. Descriptive statistics were applied to the three replicates results to calculate the average copy number, standard deviation, and standard error of average by using GraphPad Prism 9.00, software. To evaluate the significance of each test sample copy variation, the one-sample t-test was performed. The significance cutoff was defined as $P < 0.05$. All the graphics were performed using GraphPad software.

Results

The 36 samples selected for copy number variation determination were successfully assessed. The average copy obtained from the three replicate run from each test sample after calculation of the ratio (test sample copy / Calibrator copy) was defined as the copy number variation. In *Pfcoronin* gene, significant high CNV (more than 2 copies) were observed in the isolate KE024/18 (5.29 copies), t(2) = 27.91, $P = 0.0013$, 95% CI [3.62–4.95]; KI016/17 (4.19 copies), t(2) = 17.56, $P = 0.0032$, 95% CI [2.41–3.97] and isolate KI053/17 (3.43 copies), t(2) = 8.54, $P = 0.0134$, 95% CI [1.20–3.65]. Three isolates

(KM002A/18, KMSB020/18, and OS062/18) were found to harbour (2.76, 2.70, and 2.76 copies) respectively (Table 3.). The sample t-test was significant for the three isolates, KM002A/18, t(2) = 8.46, *P* = 0.0137, 95% CI [0.86–2.66]; KMSB020/18, t(2) = 7.31, *P* = 0.0182, 95% CI [0.70– 2.71]; and OS062/18,, t(2) = 6.09, *P* = 0.0259, 95% CI [0.51–3.00]. Single copy of *Pfcoronin* (copy < 1.5) was found in sample OS031/16, (1.48 copies), t(2) = 3.08, *P* = 0.0908, 95% CI [-0.19–1.16]; OS115/18 (1.47copies), t(2) = 4.83, *P* = 0.0401, 95% CI [0.05–0.88]; and KE002/17 (1.48 copies),), t(2) = 2.57 *P* = 0.1235, 95% CI [-0.32–1.29].

Table 3
Pfcoronin gene copy number variation

Isolate ID	Cq[a]	Copy (μL^{-1})[b]	Copy number variation[C]
OS031/16	23,63 ± 0.13	4930	1.48 ± 0.27 (0.15)
KI016/17	23,63 ± 0.12	13839	4.19 ± 0.31 (0,18)
KI053/17	23,96 ± 0.24	11396	3,43 ± 0.48 (0.28)
KE002/17	23,02 ± 0.10	4917	1.48 ± 0.32 (0.18)
KE024/18	23,24 ± 0.08	17449	5.29 ± 0.26 (0.15)
KM002A/18	24,31 ± 0.21	9135	2.76 ± 0.36 (0.20)
KMSB020/18	24,35 ± 0.25	8934	2.70 ± 0.40 (0.23)
OS062/18	24,32 ± 0.29	9116	2.76 ± 0.50 (0.28)
OS115/18	22,52 ± 0.15	4865	1.47 ± 0.16 (0.09)

[a] Average ± Standard deviation. [b] Average. [C] Average ± Standard deviation (Standard error of average)

Cq: PCR cycle number at which samples' amplification curves reach the threshold.

Among the fourteen samples assessed for *Pfcysteine desulfurase* gene copy number variation, multicopy of the gene were determined in 12 samples. Isolates KE002/18, KI016/18, AL054/17, and AL076/18 had copy number (2.74, 2.89, 2.13 and 2.34 copies) respectively with significant *P-value*, (KE002/18, t(2) = 10.67, *P* = 0.0087, 95% CI [1.04–2.45]; KI016/18, t(2) = 29.77, *P* = 0.0011, 95% CI [1.62–2.16]; AL054/17, t(2) = 11.06, *P* = 0.0081, 95% CI [0.69–1.56]; and AL076/18,t(2) = 5.62, *P* = 0.0302, 95% CI [0.31–2.37]). The single copy isolate SA022/17 (1.45 copy) was statistically significant, t(2) = 9.52, *P* = 0.0108, 95% CI [0.24–0.65], while,no statistical significance was observed in the single copy isolate GE023/16 (1.06 copy), t(2) = 1.87, *P* = 0.2016, 95% CI [-0.12–0.30]. Threeisolates were reported to have copy (< 2) but exceeding the assigned cutoff of single copy (< 1.5), thus were considered as double copies. There isolates were AL027/18 (1.73 copies), ALR026/18 (1.83 copies), and NY023/17 (1.97 copies). The reported copy variation was found statistically significant for the three aforementioned isolates, t(2) = 6.52, *P* = 0.0227, 95% CI [0.25–1.22] for AL027/18; t(2) = 38.28, *P* = 0.0007, 95% CI [0.74–0.93] for ALR026/18, and t(2) = 22.46, *P* = 0.0220, 95% CI [0.78–1.16] for NY023/17. Two isolates were found to contain four copies of *Pfcysteine*, OS103/16 (4.42 copies) and KE022/18 (4.66 copies), both recorded

copies were significant, t(2) = 7.90, *P* = 0.0156, 95% CI [1.56–5.29], and t(2) = 20.42, *P* = 0.0024, 95% CI [2.88–4.43] for OS103/16 and KE022/18 respectively. KI099/18 and KE023/17 isolates showed three copies each (3.16 and 3.26 copies) respectively with significant single (for KI099/18, t(2) = 14.19, *P* = 0.0049, 95% CI [1.52–2.85], and for KE023/17, t(2) = 12.67, *P* = 0.0062, 95% CI [1.49–3.03]). The highest copy variation of *Pfcysteine* gene was detected in sample OS0149/18 which had 5.09 copies, t(2) = 7.85, *P* = 0.0148, 95% CI [1.85–6.33] (Table 4.).

Table 4
Pfcysteine desulfurase gene copy number variation

Isolate ID	Cq[a]	Copy (μL^{-1})[b]	Copy number variation[c]
OS103/16	15.52 ± 0.60	1736247	4.42 ± 0.75 (0.43)
NY023/17	23.08 ± 0.04	7760	1.97 ± 0.07 (0.04)
KE023/17	18.82 ± 0.29	128105	3.26 ± 0.30 (0.17)
GE023/16	23.00 ± 0.05	43179	1.09 ± 0.08 (0.04)
AL054/17	22.63 ± 0.52	8367	2.13 ± 0.17 (0.10)
ALR026/18	23.19 ± 0.01	7110	1.83 ± 0.03 (0.02)
AL027/18	23.54 ± 0.51	6826	1.73 ± 0.19 (0.11)
AL076/18	22.80 ± 0.20	9212	2.34 ± 0.41 (0.23)
KI099/18	22.45 ± 0.23	12529	3.19 ± 0.26 (0.15)
KI016/18	22.52 ± 0.05	11358	2.89 ± 0.11 (0.06)
SA022/17	22.36 ± 0.08	57160	1.45 ± 0.08 (0.08)
OS0149/18	24.03 ± 0.29	19964	5.09 ± 0.90 (0.52)
KE022/18	24.16 ± 0.10	18273	4.66 ± 0.31 (0.17)
KE002/18	24.04 ± 0.06	10773	2.74 ± 0.28 (0.16)

[a] Average ± Standard deviation. [b] Average. [C] Average ± Standard deviation (Standard error of average)

Cq: PCR cycle number at which samples' amplification curves reach the threshold.

Single copy of *Plasmepsin II* gene was reported in 7 samples over 13 samples assessed for *plasmepsin II* copy number variation (Table 5). The following isolates KB002/18, AL090/18, NY021/18, and MI009/16 had copy recorded (1.31, 1.05, 1.45, and 1.38 copy) respectively. The reported copy were not statistically significant KB002/18, t(2) = 2.21, *P* = 0.1574, 95% CI [-0.29–0.93]; AL090/18, t(2) = 3.67, *P* = 0.0669, 95% CI [-0.01–0.11]; NY021/18, t(2) = 3.43, *P* = 0.0753, 95% CI [-0.11–1.029], and MI009/16, t(2) = 1.81, *P* = 0.2113, 95% CI [-0.53–1.30]. Single copy was also found in isolates GE023/17 (1.35 copy), KI006/18

(1.04 copy) and NY066/17 (1.05 copy). The copy recorded in GE023/17 was significant t(2) = 6.35, *P* = 0.0.02939, 95% CI [0.11–0.58], while copy found in the KI006/18 and in NY066/17 were not, (t(2) = 1.92, *P* = 0.1946, 95% CI [-0.04–0.12], and (t(2) = 2.44, *P* = 0.1348, 95% CI [-0.04–0.14]), for KI006/18, and NY066/17 respectively. We reported two isolates containing 2 copies of *Pfplasmepsin II*: KI045/18 (1.87copies), t(2) = 2.88, *P* = 0.1028, 95% CI [-0.43–2.17], and KI004/18 (1.96 copies), t(2) = 8.28, *P* = 0.0143, 95% CI [0.46–1.46]. One isolate was showed four copies, SA027/17 (4.48 copies), t(2) = 22.26, *P* = 0.0020, 95% CI [2.81–4.15]. Moreover, although *Pfplasmepsin* II gene was found to be the less polymorphic gene in copy number variation, three isolates were reported with the most high copy variation observed in this study: these were, OS062/17 (9.63 copies), AL106/16 (9.18 copies), and KM019/18 (9.41 copies). The recorded copies were statistically significant, t(2) = 80.50, *P* = 0.0002, 95% CI [8.01–9.09] for OS062/17; t(2) = 28.61, *P* = 0.0012, 95% CI [6.95–9.41] for AL106/16; and t(2) = 90.51, *P* = 0.0001, 95% CI [8.01–8.81] for KM019/18, (Table 5.).

Table 5
Pfplasmepsin 2 gene copy number variation

Isolate ID	Cq[a]	Copy (μL^{-1})[b]	Copy number variation[c]
KB002/18	22.67 ± 0.26	48578	1.31 ± 0.24 (0.14)
AL090/18	24.22 ± 0.04	38944	1.05 ± 0.02 (0.01)
SA027/17	24.33 ± 0.23	16450	4.48 ± 0.27 (0.15)
KM019/18	24.31 ± 0.02	34502	9.41 ± 0.16 (0.09)
AL106/16	23.19 ± 0.08	33670	9.18 ± 0.49 (0.28)
NY021/18	23.64 ± 0.23	53608	1.45 ± 0.23 (0.13)
MI009/16	22.91 ± 0.29	51154	1.38 ± 0.36 (0.21)
NY066/17	26.37 ± 0.02	3894	1.05 ± 0.03 (0.02)
OS062/17	23.14 ± 0.06	35333	9.63 ± 0.18 (0.10)
KI045/18	25.46 ± 0.30	6882	1.87 ± 0.52 (0.30)
KI004/18	25.43 ± 0.14	7241	1.96 ± 0.20 (0.11)
KI006/18	24.27 ± 0.13	38457	1.04 ± 0.03 (0.02)
GE023/17	22.65 ± 0.10	49815	1.35 ± 0.09 (0.05)

[a] Average ± Standard deviation. [b] Average. [C] Average ± Standard deviation (Standard error of average)

Cq: PCR cycle number at which samples' amplification curves reach the threshold.

Discussion

Plasmodium falciparum parasites experience strong pressure from antimalarial drugs treatments and from exposure to host immune responses. Single nucleotide variation and gene copy number variation are two alternatives that *P.falciparum* parasites exploit to circumvent environment pressure. Importantly, gene copy number variations have shown prominent role in the gaining of antimalarials resistance. The investigation of gene CNVs in the putative candidate genes related to drug pressure is crucial. In this study, we explored copy number variation in three putative genes of *P.falciparum* (*Pfcoronin, Pf cysteine desulfurase and Pf. Plasmepsin 2*). Previous studies have shown point mutation in the *Pfcoronin and Pfcysteine desulfurase* conferring reduced artemisinin susceptibility, and selection for lumefantrine drug respectively [31, 32], however, amplification of *Pfplasmepsin 2* copy number is commonly known as piperaquine resistance marker [22, 23]. It should be noted that the current study represents the first report investigating copy number variation in the aforementioned genes in Kenya since we did not find a report that describes CNV in the *Pfcoronin* and *Pfcysteine desulfurase* in the literature reviews. *Pfplasmepsin* CNVs are widely reported although no report was found to describe Kenya parasites *plasmepsin 2* CNVs. Indeed, most of the studies conducted in *P.falciparum* copy number variation in Kenya were focused on the multidrug resistance gene 1 (*Pfmdr1*) [48]. Therefore the lack of previous data describing CNVs of the three genes we investigated in this study limits the comparison of our results. Importantly, the finding of this report will serve as baseline data for future investigations.

The highest increased copy number detected in the *Pfcoronin* gene was five-fold amplification observed in one isolate (KE024/18) followed by 4 copies detected in KI016/17 (Fig 5. A), the others isolate (except two, which had single copy) showed fold changes >2.5 copies and therefore considered as 3 copies. In total 7 over 9 isolates have been found to increase *Pfcoronin* copy number. The profile of increased *Pfcoronin* gene copy number observed in this study leads us to postulate that this increased copy number may be a result of drugs pressure namely dihydroartemisinin-piperaquine combination which is the second frontline ACT drug used in Kenya since 2009 [49]. Indeed, *Pfcoronin* has shown point mutations (R100K, E107V and G50E) after *P.falciparum* has been pulsed intermittently with dihydroartemisinin-piperaquine over 4 year's periods [31]. The mutant parasites have shown reduced DHA susceptibility compared to wild type parasites [31]. These findings show evidence implicating the *Pfcoronin* gene in the response to DHA drug pressure, therefore these findings strengthen our hypothesis that the copy number observed in the *Pfcoronin* may be a response of drug pressure as CNVs contribute to parasite fitness.

Moreover, *Pfcoronin* protein in addition to being involved in drug response plays important role in the parasite. The protein is conserved, and is expressed in the merozoite's periphery and also localized in late schizogony [50]. The coronin protein contains β-propeller domain like *PfKelch13* protein [31] and it processes WD40 repeats (repetitive residues composed of tryptophan (W) – aspartic acid (D) at the C-terminus while having glycine (G) – histidine (H) at the N terminus) spanning the β-propeller domain which serves protein-DNA and protein-protein interaction and is also involved in multiples functions such as chemical inhibitors pathways, regulation of transcription, signal transduction [51, 52, 41] *Pfcoronin* is also implicated in vesicular transport, cytoskeletal motility F-actin binding and its structural organisation [31]. These functions highlight the prominent role of *Pfcoronin* protein in the *P.falciparum* parasite.

The isolates explored for *Pfcysteine desulfurase* gene copy number variation revealed important amplification. Overall 12 over 14 isolates were found to contain *cysteine desulfurase* gene copy polymorphism. The highest copy amplification observed was 5 copies of cysteine gene in the isolate OS0149/18 and KE022/18 (Fig 5. B), and the rest of the isolates had CNVs ranging from 4 copies to 1 copy. Only two isolates were detected with single copy. The results show important amplification profiles of the *cysteine desulfurase* gene in the study area. Amambua et, al., [32] reported increased frequency in the allele K65 in *Pfcysteine desulfurase* gene in the Gambian clinical isolates, and the IC_{50} for lumefantrine drug was significantly higher in isolates that harboured the K65. These findings imply the *cysteine desulfurase* in response to lumefantrine drug pressure. Since *Pfalciparum* parasites use point mutations and gene copy polymorphism to withstand antimalarial drugs pressure, therefore, the copy number variation observed in the *cysteine desulfurase* gene in this current study may be attributed to parasite exposure to arthemeter-lumefantrine drug which is in use as first frontline ACT in Kenya since 2004 [34] for the treatment of uncomplicated malaria. *Pfcysteine desulfurase* protein is a member of class-V pyridoxal-phosphate-dependent aminotransferase family [53]. The protein is involved in the Iron-sulphur biosynthesis and its delivery to different metabolic pathways for instance the Fe-S complex which is modulated in apicomplexans during drug resistance, and stress conditions [32] In *P.falciparum*, the iron homeostasis is substantive for blood-stage parasite development and is implied in the mechanisms of quinolones drug, _ the lumefantrine drug family [32]. Moreover, many important features were predicted in the cysteine desulfurase protein including *Aminotran_5 domain,* disordered regions covering low complexity motifs [41] those functions include adhesion, signal transduction, molecular recognition, interaction with phospholipid bilayers, protein modification, and modulation of protein translation [40, 54]

Pfplasmepsin 2 gene was found to be the less polymorphic gene in copy number variation, in total 6 isolates showed copy number variation. However, the highest CNVs detected in this study was found in the *Pfplasmepsin 2* gene, three isolates (OS062/17, KM019/18, and AL106/16) showed (9.63, 9.41 and 9.18 copies) of *plasmepsin 2* gene respectively (Fig 5. C). *Pfplasmepsin 2* and plasmepsin 3 genes copy number variation are widely reported and known as piperaquine resistance markers [23, 55]. Previous studies investigating the impact of plasmepsin genes CNVs in the resistance to piperaquine drug have allowed to better understand how a gene copy number polymorphism reduces drug efficacy and leads to drug resistance. Indeed, within the host red cells, malaria parasites degrade haemoglobin (Hb) generating peptides and amino acids which are subsequently used by the parasites as nutrient [56]. Plasmepsin 2 and plasmepsin 3 are food digestive vacuole enzymes and are directly implicated in haemoglobin degradation [56]. However, haemoglobin degradation generates a byproduct (*Heme*) which is toxic to the parasites [57, 56]. To overcome this toxicity, malaria parasites depolymerise the toxic heme to non-toxic *hemozoin* [57]. Piperaquine drug targets plasmepsin enzymes and prevent both haemoglobin degradation and the conversion of the toxic heme to non-toxic hemozoin [57]. Therefore, piperaquine treatment leads to the accumulation of undigested haemoglobin and toxic heme as well as reduction of hemozoin production [56] [57]. The lack of haemoglobin degradation starve parasites and finally, the toxic heme leads to parasites' death. However, amplification of plasmepsin 2 gene copy number causes

overexpression of plasmepsin 2 enzyme that overwhelms the inhibitory effect of piperaquine [57], that is, single copy of plasmepsin 2 is effectively neutralized by piperaquine preventing haemoglobin degradation, however, when there are multicopy of plasmepsin 2, piperaquine neutralises some copies, while extra copies of the enzyme maintain hemoglobin degradation (Fig 6.), consequently the parasites withstand piperaquine inhibitory effect. Trophozoites under piperaquine drug treatment have revealed large digestive food vacuoles containing undegraded hemoglobin [57]. This supports the inhibitory effect of piperaquine (the prevention of Hb digestion). Moreover, the mechanism we described here, illustrating the influence of plasmepsin gene amplification on the piperaquine inhibitory effect has been strengthened by Mukherjee et al., [58] in a study where the author found that inactivation plamepsins genes in *P.falciparum* results to hyper sensibility of piperaquine drug, highlighting the role of plasmepsins in piperaquine susceptibility. In the light of the mechanism described here, we speculate the possibility that CNVs detected in the *Pfcoronin*, as well as in the *Pfcysteine desulfuferase* and in *Pfplasmepsin* 2 gene may act in a similar mechanism like plasmepsin genes and thus evolve to drug resistance.

Mechanism of DNA copy number formation

P.falciparum genome is highly AT-rich, all of the 23Mb, the overall (AT)-content is estimated at 80.6% and can reach 95% in the intergenic, and introns regions [42]. The AT-content greatly contributes to copy number generation. It has been shown that genomic breakpoints frequently occur in AT-rich regions and in near homopolymeric traits of poly-(A) or poly-(T) bases [18]. The multidrug resistance gene 1 (pfmdr1) copy amplicon (Chr 5) and the GTP- cyclohydrolase I gene copy (Chr 12) have been commonly detected in DNA breakage sites within monomeric A/T tracts [59]. Huckaby et al., [60] have conducted an in-depth analysis of a resistant clone of *P.falciparum* genome copy number and found that all the resistance-conferring CNVs had long monomeric A/T tracks on their upstream and downstream boundaries. Indeed, the most supported mechanism of CNVs formation imply A/T-tracks, and DNA hairpins structure as essential structure in the initiation and acquisition of new copy formation. The A/T-tracks regions trigger homology mediated pathways across DNA breaks that are caused by the hairpins structure throughout the parasite' genome (Fig 7.). Hairpins trigger DNA breaks likely due to halting of the replication fork progression, or due to the recognition of hairpin-binding proteins or by enzymatic action and polymerase pausing impact. [60, 61]. Hairpins structures are known to cause DNA breakage and can lead to many replication forks collapsing [60]. The DNA breakage is fixed by utilizing the A/T tracks as microhomology-mediated repair that leads to copy number generation [60, 61].

In *P.falciparum*, the copy number of drug-target protein may amplify under drug pressure, and selection, for example, drug of fitness cost lead the parasite to conserve the beneficial amplicon during the repairing processes of DNA breakage [61, 60]. It has been shown CNVs that occur in the field parasites under drug pressure do not always de-amplify to single copy of the target protein, the parasites conserve some number of copies and these copies will be quickly re-amplified when parasites encounter the same drug-like before, this allows parasites to rigorously resist to drug pressure [61]. Moreover, it has been reported that CNVs influence both the expression of genes inside and outside CNVs regions [12], highlighting a co-

regulation of genes in the CNVs vicinity thereby suggesting that CNVs act as substantive elements of transcriptional regulation. Amplification of gene copy number play important role in *P.falciparum* adaptation to its environment, however, deletion of specific genes in the parasites' population may also render the rapid diagnosis tests inadequate [61], both CNVs amplification and gene deletion strongly impact the level of gene expression. Moreover, it should be noted that *P.falciparum* genome is subject to natural rearrangement of DNA, consequently, gene copy number variations may occur even in the absence of drug pressure, therefore, CNVs are not always representative of drug resistance. More remains to be known about the distribution CNVs, their inheritance as well as the rate at which *P.falciparum* develops or reverts CNVs to the iniatial state.

Conclusion

Gene copy number variations are important features in the adaption of *P.falciparum* parasites to their environment. CNVs strongly contribute to drug resistance. Here, we investigated CNVs in three candidate genes of *P.falciparum* in clinical samples from Mbita isolates, one of the high intensive malaria transmission settings. The findings showed differential levels of copy polymorphism in the three genes investigated, which may be a result of antimalarial drugs pressure, therefore the findings call for continuous surveillance drug resistance in the study area. Importantly, this report represents the first study exploring CNVs in Mbita field isolate parasites. The results will serve as baseline data for future investigation.

Declarations

Acknowledgements

We would like to reiterate our gratitude to Dr Jeremy K. Herren's research team for their valuable assistance, and we are grateful to the management staff of the International Centre of Insect Physiology and Ecology (ICIPE) for their availability and assistance. We reiterate our sincere thanks to the staff of the Emerging Infectious Disease (EID) laboratory and Molecular Biology and Bioinformatics Unit (MBBU). We thank all the volunteers recruited in this study and parents/guardians as well.

Author contributions

V.A.M., J.K.H., J.K. supervised the work; H.D., E.E.M., V.A.M. designed the experiments; H.D., E.E.M. conducted the experiments; J.K.H. provided samples; H.D., V.A.M. were involved in the data analysis; H.D. wrote the manuscript; H.D., V.A.M., J.K.H., J.K., and E.E.M. read, commented, reviewed the final manuscript.

Declaration of interests

The authors declare no conflict of interest.

Data accessibility

Data are available under justifiable request. Corresponding author, H.D

Funding

This study was supported by the Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI): P.O Box 62000-00200, Nairobi (KENYA). Physical Address: JKUAT – Juja Campus.

References

1. Simam, J., Rono, M., Ngoi, J., Nyonda, M., Mok, S., Marsh, K., Bozdech, Z., & Mackinnon, M. (2018). Gene copy number variation in natural populations of Plasmodium falciparum in Eastern Africa. *BMC Genomics*, *19*(1), 1–15. https://doi.org/10.1186/s12864-018-4689-7
2. Volkman, S. K., Sabeti, P. C., Decaprio, D., Neafsey, D. E., Schaffner, S. F., Milner, D. A., Daily, J. P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M. T., Lukens, A., Derr, A., Stange-thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., … Wirth, D. F. (2006). A genome-wide map of diversity in Plasmodium falciparum. *Nature genetics, 39*(1), 113–119. https://doi.org/10.1038/ng1930
3. Fernandez-Jimenez, N., Castellanos-Rubio, A., Plaza-Izurieta, L., Gutierrez, G., Irastorza, I., Castaño, L., Vitoria, J. C., & Bilbao, J. R. (2011). Accuracy in copy number calling by qPCR and PRT: A matter of DNA. *PLoS ONE*, *6*(12), 1–7. https://doi.org/10.1371/journal.pone.0028910
4. Shebanits, K., Günther, T., Johansson, A. C. V., Maqbool, K., Feuk, L., Jakobsson, M., & Larhammar, D. (2019). Copy number determination of the gene for the human pancreatic polypeptide receptor NPY4R using read depth analysis and droplet digital PCR. *BMC Biotechnology*, *19*(1), 1–8. https://doi.org/10.1186/s12896-019-0523-9
5. D'haene, B., Vandesompele, J., & Hellemans, J. (2010). Accurate and objective copy number profiling using real-time quantitative PCR. *Methods*, *50*(4), 262–270. https://doi.org/10.1016/j.ymeth.2009.12.007
6. Mehrotra, M. (2016). PCR-based detection of DNA copy number variation. *Methods in Molecular Biology*, *1392*, 27–32. https://doi.org/10.1007/978-1-4939-3360-0_3
7. Stankiewicz, P., & Lupski, J. R. (2010). *Structural Variation in the Human Genome and its Role in Disease.Annual Review of Medecine, 61(1), 437–455.* https://doi.org/10.1146/annurev-med-100708-204735
8. Olsson, M., Kierczak, M., Karlsson, Jablonska, J., Leegwater, P., Koltookian, M., Abadie, J., Dufaure De Citres, C., Thomas, A., Hedhammar, Tintle, L., Lindblad-Toh, K., & Meadows, J. R. S. (2016). Absolute quantification reveals the stable transmission of a high copy number variant linked to autoinflammatory disease. *BMC Genomics*, *17*(1), 1–9. https://doi.org/10.1186/s12864-016-2619-0
9. Tam, G. W. C., Redon, R., Carter, N. P., & Grant, S. G. N. (2009). The Role of DNA Copy Number Variation in Schizophrenia. *Biological Psychiatry*, *66*(11), 1005–

1012. https://doi.org/10.1016/j.biopsych.2009.07.027

10. Cheeseman, I. H., Gomez-Escobar, N., Carret, C. K., Ivens, A., Stewart, L. B., Tetteh, K. K. A., & Conway, D. J. (2009). Gene copy number variation throughout the Plasmodium falciparum genome. *BMC Genomics*, *10*, 1–11. https://doi.org/10.1186/1471-2164-10-353
11. Samarakoon, U., Gonzales, J. M., Patel, J. J., Tan, A., Checkley, L., & Ferdig, M. T. (2011). The landscape of inherited and de novo copy number variants in a plasmodium falciparum genetic cross. *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-457
12. Mackinnon, M. J., Li, J., Mok, S., Kortok, M. M., Marsh, K., Preiser, P. R., & Bozdech, Z. (2009). Comparative transcriptional and genomic analysis of Plasmodium falciparumfield isolates. *PLoS Pathogens*, *5*(10). https://doi.org/10.1371/journal.ppat.1000644
13. Alano, P. (1995). *Plasmodium falciparum: Parasites Defective in Early Stages of Gametocytogenesis*.
14. Biggs, B. A., Kemp, D. J., & Brown, G. V. (1989). Subtelomeric chromosome deletions in field isolates of Plasmodium falciparum and their relationship to loss of cytoadherence in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(7), 2428–2432. https://doi.org/10.1073/pnas.86.7.2428
15. Nair, S., Nkhoma, S., Nosten, F., Mayxay, M., French, N., Whitworth, J., & Anderson, T. (2010). Genetic changes during laboratory propagation: Copy number At the reticulocyte-binding protein 1 locus of Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *172*(2), 145–148. https://doi.org/10.1016/j.molbiopara.2010.03.015
16. Ribacke, U., Mok, B. W., Wirta, V., Normark, J., Lundeberg, J., Kironde, F., Egwang, T. G., Nilsson, P., & Wahlgren, M. (2007). Genome-wide gene amplifications and deletions in Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *155*(1), 33–44. https://doi.org/10.1016/j.molbiopara.2007.05.005
17. Lim, P., Alker, A. P., Khim, N., Shah, N. K., Incardona, S., Doung, S., Yi, P., Bouth, D. M., Bouchier, C., Puijalon, O. M., Meshnick, S. R., Wongsrichanalai, C., Fandeur, T., Le Bras, J., Ringwald, P., & Ariey, F. (2009). Pfmdr1 copy number and artemisinin derivatives combination therapy failure in falciparum malaria in Cambodia. *Malaria Journal*, *8*(1), 1–9. https://doi.org/10.1186/1475-2875-8-11
18. Su, X. Z., Lane, K. D., Xia, L., Sá, J. M., & Wellems, T. E. (2019). Plasmodium genomics and genetics: New insights into malaria pathogenesis, drug resistance, epidemiology, and evolution. *Clinical Microbiology Reviews*, *32*(4), 1–29. https://doi.org/10.1128/CMR.00019-19
19. Rosenthal, P. J. (2004). Cysteine proteases of malaria parasites. *International Journal for Parasitology*, *34*(13–14), 1489–1499. https://doi.org/10.1016/j.ijpara.2004.10.003
20. Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A. C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Ménard, S., Rogers, W. O., … Ménard, D. (2014). A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, *505*(7481), 50–55. https://doi.org/10.1038/nature12876
21. Picot, S., Olliaro, P., De Monbrison, F., Bienvenu, A. L., Price, R. N., & Ringwald, P. (2009). A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite

resistance and treatment outcome in falciparum malaria. *Malaria Journal*, *8*(1). https://doi.org/10.1186/1475-2875-8-89

22. Ansbro, M. R., Jacob, C. G., Amato, R., Kekre, M., Amaratunga, C., Sreng, S., Suon, S., Miotto, O., Fairhurst, R. M., Wellems, T. E., & Kwiatkowski, D. P. (2020). Development of copy number assays for detection and surveillance of piperaquine resistance-associated plasmepsin 2/3 copy number variation in Plasmodium falciparum. *Malaria Journal*, *19*(1), 1–10. https://doi.org/10.1186/s12936-020-03249-x
23. Witkowski, B., Duru, V., Khim, N., Ross, L. S., Saintpierre, B., Beghain, J., Chy, S., Kim, S., Ke, S., Kloeung, N., Eam, R., Khean, C., Ken, M., Loch, K., Bouillon, A., Domergue, A., Ma, L., Bouchier, C., Leang, R., … Ménard, D. (2017). A surrogate marker of piperaquine-resistant Plasmodium falciparum malaria: a phenotype-genotype association study. *The Lancet Infectious Diseases*, *17*(2), 174–183. https://doi.org/10.1016/S1473-3099(16)30415-7
24. Duah, N. O., Matrevi, S. A., De Souza, D. K., Binnah, D. D., Tamakloe, M. M., Opoku, V. S., Onwona, C. O., Narh, C. A., Quashie, N. B., Abuaku, B., Duplessis, C., Kronmann, K. C., & Koram, K. A. (2013). Increased pfmdr1 gene copy number and the decline in pfcrt and pfmdr1 resistance alleles in Ghanaian Plasmodium falciparum isolates after the change of anti-malarial drug treatment policy. *Malaria Journal*, *12*(1), 1–10. https://doi.org/10.1186/1475-2875-12-377
25. Witkowski, B., Nicolau, M. L., Soh, P. N., Iriart, X., Menard, S., Alvarez, M., Marchou, B., Magnaval, J. F., Benoit-Vical, F., & Berry, A. (2010). Plasmodium falciparum isolates with increased pfmdr1 copy number circulate in West Africa. *Antimicrobial Agents and Chemotherapy*, *54*(7), 3049–3051. https://doi.org/10.1128/AAC.00209-10
26. Kamau, E., Tolbert, L. D. S., Kortepeter, L., Pratt, M., Nyakoe, N., Muringo, L., Ogutu, B., Waitumbi, J. N., & Ockenhouse, C. F. (2011). Development of a highly sensitive genus-specific quantitative reverse transcriptase real-time PCR assay for detection and quantitation of plasmodium by amplifying RNA and DNA of the 18S rRNA genes. *Journal of Clinical Microbiology*, *49*(8), 2946–2953. https://doi.org/10.1128/JCM.00276-11
27. Kamau, E., Alemayehu, S., Feghali, K. C., Saunders, D., & Ockenhouse, C. F. (2013). Multiplex qPCR for Detection and Absolute Quantification of Malaria. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071539
28. Lee, C., Lee, S., Shin, S. G., & Hwang, S. (2008). Real-time PCR determination of rRNA gene copy number: Absolute and relative quantification assays with Escherichia coli. *Applied Microbiology and Biotechnology*, *78*(2), 371–376. https://doi.org/10.1007/s00253-007-1300-6
29. Farrugia, C., Cabaret, O., Botterel, F., Bories, C., Foulet, F., Costa, J. M., & Bretagne, S. (2011). Cytochrome b gene quantitative PCR for diagnosing Plasmodium falciparum infection in travellers. *Journal of Clinical Microbiology*, *49*(6), 2191–2195. https://doi.org/10.1128/JCM.02156-10
30. Hou, Y., Zhang, H., Miranda, L., & Lin, S. (2010). Serious overestimation in quantitative pcr by circular (supercoiled) plasmid standard: Microalgal pcnaas the model gene. *PLoS ONE*, *5*(3). https://doi.org/10.1371/journal.pone.0009545

31. Demas, A. R., Sharma, A. I., Wong, W., Early, A. M., Redmond, S., Bopp, S., Neafsey, D. E., Volkman, S. K., Hartl, D. L., & Wirth, D. F. (2018). Mutations in plasmodium falciparum actin-binding protein coronin confer reduced artemisinin susceptibility. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50), 12799–12804. https://doi.org/10.1073/pnas.1812317115

32. Amambua-Ngwa, A., Jeffries, D., Amato, R., Worwui, A., Karim, M., Ceesay, S., Nyang, H., Nwakanma, D., Okebe, J., Kwiatkowski, D., Conway, D. J., & D'Alessandro, U. (2018). Consistent signatures of selection from genomic analysis of pairs of temporal and spatial Plasmodium falciparum populations from the Gambia. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-28017-5

33. Wamea, K. (2019). No Evidence of Plasmodium falciparum k13 Artemisinin Resistance-Conferring Mutations over a 24-Year Analysis in Coastal Kenya but a Near Complete Reversion to Chloroquine-Sensitive Parasites. *Antimicrobial Agents and Chemotherapy*. 63(12):1-12. https://doi.org/10.1128/AAC.01067-19

34. Kishoyian, G., Njagi, E. N. M., Orinda, G. O., & Kimani, F. T. (2018). Chloroquine Sensitivity and Prevalence of Chloroquine-resistant Genes pfcrt and pfmdr-1 in Western Kenya after Two Decades of Chloroquine Withdrawal. *Annals of Medical and Health Sciences Research*, 331–335. https://www.amhsr.org/abstract/chloroquine-sensitivity-and-prevalence-of-chloroquineresistant-genes-pfcrt-and-pfmdr1-in-western-kenya-after-two-decades-4808.html

35. Touray, A. O., Mobegi, V. A., Wamunyokoli, F., & Herren, J. K. (2020). Diversity and Multiplicity of P. falciparum infections among asymptomatic schoolchildren in Mbita, Western Kenya. *Scientific Reports*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-62819-w

36. Wanyua, S., Ndemwa, M., Goto, K., Tanaka, J., K'Opiyo, J., Okumu, S., Diela, P., Kaneko, S., Karama, M., Ichinose, Y., & Shimada, M. (2013). Profile: The Mbita Health and Demographic Surveillance System. *International Journal of Epidemiology*, *42*(6), 1678–1685. https://doi.org/10.1093/ije/dyt180

37. Omondi, R., & Kamau, L. (2018). Reduced Plasmodium Infection in Primary School Children Following Universal Distribution of Insecticide Treated Bed Nets in Kasipul, Homa-Bay County, Kenya. *International Journal of Sciences: Basic and Applied Research*, *4531*(January), 26–36. https://www.researchgate.net/publication/326368468

38. Minakawa, N., Dida, G. O., Sonye, G. O., Futami, K., & Njenga, S. M. (2012). Malaria vectors in Lake Victoria and adjacent habitats in Western Kenya. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0032725

39. Noor, A. M., Gething, P. W., Alegana, V. A., Patil, A. P., Hay, S. I., Muchiri, E., Juma, E., & Snow, R. W. (2009). The risks of malaria infection in Kenya in 2009. *BMC Infectious Diseases*, *9* (May 2014). https://doi.org/10.1186/1471-2334-9-180

40. Gisselberg, J. E., Dellibovi-Ragheb, T. A., Matthews, K. A., Bosch, G., & Prigge, S. T. (2013). The Suf Iron-Sulfur Cluster Synthesis Pathway Is Required for Apicoplast Maintenance in Malaria Parasites. *PLoS Pathogens*, *9*(9). https://doi.org/10.1371/journal.ppat.1003655

41. Diarra et al., Diarra, H., Makhulu, E. E., Odhiambo, P. O., Irekwa, R. M., Kinyua, J., Herren, J. K., & Mobegi, V. A. (2021). Molecular Investigation of Genetic Signatures of Selection in Plasmodium falciparum Actin-Binding Protein Coronin, Cysteine Desulfurase, and Plasmepsin 2 Gene in Mbita Field Isolates, Western Kenya. Open Journal of Genetics, 11(04), 120–144. https://doi.org/10.4236/ojgen.2021.114011
42. Karine et, A. (2008). Genomics and Integrated Systems Biology in Plasmodium falciparum: A Path to Malaria Control and Eradication. *Parasite Immunology*, *23*(1), 1–7. https://doi.org/10.1111/j.1365-3024.2011.01340.x.Genomics
43. Tajebe, A., Aemero, M., Francis, K., & Magoma, G. (2015). Identification of chloroquine resistance Pfcrt-K76T and determination of Pfmdr1-N86Y copy number by SYBR Green I qPCR. *Asian Pacific Journal of Tropical Biomedicine*, *5*(3), 208–220. https://doi.org/10.1016/S2221-1691(15)30008-3
44. Cano, J., Descalzo, M. Á., Moreno, M., Chen, Z., Nzambo, S., Bobuakasi, L., Buatiche, J. N., Ondo, M., Micha, F., & Benito, A. (2006). Spatial variability in the density, distribution and vectorial capacity of anopheline species in a high transmission village (Equatorial Guinea). *Malaria Journal*, *5*, 1–10. https://doi.org/10.1186/1475-2875-5-1
45. Nick, D. (2014). Interpreting Melt Curves : An Indicator, Not a Diagnosis. *Integrated DNA Technologies, Inc*, 1–7. https://www.idtdna.com/pages/decoded/decoded-articles/core-concepts/decoded/2014/01/20/interpreting-melt-curves-an-indicator-not-a-diagnosis
46. Santolamazza, F., Avellino, P., Siciliano, G., Yao, F. A., Lombardo, F., Ouédraogo, J. B., Modiano, D., Alano, P., & Mangano, V. D. (2017). Detection of Plasmodium falciparum male and female gametocytes and determination of parasite sex ratio in human endemic populations by novel, cheap and robust RT-PCR assays. *Malaria Journal*, *16*(1), 1–11. https://doi.org/10.1186/s12936-017-2118-z
47. Leroy, D., Macintyre, F., Adoke, Y., Ouoba, S., Barry, A., Mombo-Ngoma, G., Ndong Ngomo, J. M., Varo, R., Dossou, Y., Tshefu, A. K., Duong, T. T., Phuc, B. Q., Laurijssens, B., Klopper, R., Khim, N., Legrand, E., & Ménard, D. (2019). African isolates show a high proportion of multiple copies of the Plasmodium falciparum plasmepsin-2 gene, a piperaquine resistance marker. *Malaria Journal*, *18*(1), 1–11. https://doi.org/10.1186/s12936-019-2756-4
48. Ngalah, B. S., Ingasia, L. A., Cheruiyot, A. C., Chebon, L. J., Juma, D. W., Muiruri, P., Onyango, I., Ogony, J., Yeda, R. A., Cheruiyot, J., Mbuba, E., Mwangoka, G., Achieng, A. O., Ng'Ang'A, Z., Andagalu, B., Akala, H. M., & Kamau, E. (2015). Analysis of Major Genome Loci Underlying Artemisinin Resistance and pfmdr1 Copy Number in pre-and post-ACTs in Western Kenya. *Scientific Reports*, *5*, 1–6. https://doi.org/10.1038/srep08308
49. Ogutu, B. R., Onyango, K. O., Koskei, N., Omondi, E. K., Ongecha, J. M., Otieno, G. A., Obonyo, C., Otieno, L., Eyase, F., Johnson, J. D., Omollo, R., Perkins, D. J., Akhwale, W., & Juma, E. (2014). Efficacy and safety of artemether-lumefantrine and dihydroartemisinin-piperaquine in the treatment of uncomplicated Plasmodium falciparum malaria in Kenyan children aged less than five years: Results of an open-label, randomized, single-centre study. *Malaria Journal*, *13*(1), 1–10. https://doi.org/10.1186/1475-2875-13-33

50. Bane, K. S., Lepper, S., Kehrer, J., Sattler, J. M., Singer, M., Reinig, M., Klug, D., Heiss, K., Baum, J., Mueller, A. K., & Frischknecht, F. (2016). The Actin Filament-Binding Protein Coronin Regulates Motility in Plasmodium Sporozoites. *PLoS Pathogens*, *12*(7), 1–26. https://doi.org/10.1371/journal.ppat.1005710
51. Li, D., & Roberts, R. (2001). WD-repeat proteins: Structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Sciences*, *58*(14), 2085–2097. https://doi.org/10.1007/PL00000838
52. Schapira, M., Tyers, M., Torrent, M., & Arrowsmith, C. H. (2017). WD40 repeat domain proteins: A novel target class? *Nature Reviews Drug Discovery*, *16*(11), 773–786. https://doi.org/10.1038/nrd.2017.179
53. Schwartz, C. J., Djaman, O., Imlay, J. A., & Kiley, P. J. (2000). The cysteine desulfurase, IscS, has a major role in vivo Fe-S cluster formation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(16), 9009–9014. https://doi.org/10.1073/pnas.160261497
54. Lill, R., Broderick, J. B., & Dean, D. R. (2015). Special issue on iron-sulfur proteins: Structure, function, biogenesis and diseases. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1853*(6), 1251–1252. https://doi.org/10.1016/j.bbamcr.2015.03.001
55. Amato, R., Lim, P., Miotto, O., Amaratunga, C., Dek, D., Pearson, R. D., Almagro-Garcia, J., Neal, A. T., Sreng, S., Suon, S., Drury, E., Jyothi, D., Stalker, J., Kwiatkowski, D. P., & Fairhurst, R. M. (2017). Genetic markers associated with dihydroartemisinin-piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study. *The Lancet Infectious Diseases*, *17*(2), 164–173. https://doi.org/10.1016/S1473-3099(16)30409-1
56. Bopp, S., Magistrado, P., Wong, W., Schaffner, S. F., Mukherjee, A., Lim, P., Dhorda, M., Amaratunga, C., Woodrow, C. J., Ashley, E. A., White, N. J., Dondorp, A. M., Fairhurst, R. M., Ariey, F., Menard, D., Wirth, D. F., & Volkman, S. K. (2018). Plasmepsin II–III copy number accounts for bimodal piperaquine resistance among Cambodian Plasmodium falciparum. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04104-z
57. Qidwai, T. (2020). Exploration of copy number variation in genes related to anti-malarial drug resistance in Plasmodium falciparum. *Gene*, *736*(October 2019), 144414. https://doi.org/10.1016/j.gene.2020.144414
58. Mukherjee, A., Gagnon, D., & Wirth, D. F. (2018). Inactivation of Plasmepsins 2 and 3 Sensitizes Plasmodium falciparum to the Antimalarial Drug Piperaquine Angana. *Antimicrobial Agents and Chemotherapy, 62(4).* doi:10.1128/aac.02309-17 . https://pubmed.ncbi.nlm.nih.gov/29439977/.
59. Anderson, T. J. C., Patel, J., & Ferdig, M. T. (2009). Gene copy number and malaria biology. *Trends in Parasitology*, *25*(7), 336–343. https://doi.org/10.1016/j.pt.2009.04.005
60. Huckaby, A. C., Granum, C. S., Carey, M. A., Szlachta, K., Al-Barghouthi, B., Wang, Y. H., & Guler, J. L. (2019). Complex DNA structures trigger copy number variation across the Plasmodium falciparum genome. *Nucleic Acids Research*, *47*(4), 1615–1627. https://doi.org/10.1093/nar/gky1268

61. Guler, J. L., Freeman, D. L., Ahyong, V., Patrapuvich, R., White, J., Gujjar, R., Phillips, M. A., DeRisi, J., & Rathod, P. K. (2013). Asexual Populations of the Human Malaria Parasite, Plasmodium falciparum, Use a Two-Step Genomic Strategy to Acquire Accurate, Beneficial DNA Amplifications. *PLoS Pathogens*, *9*(5). https://doi.org/10.1371/journal.ppat.1003375

Figures

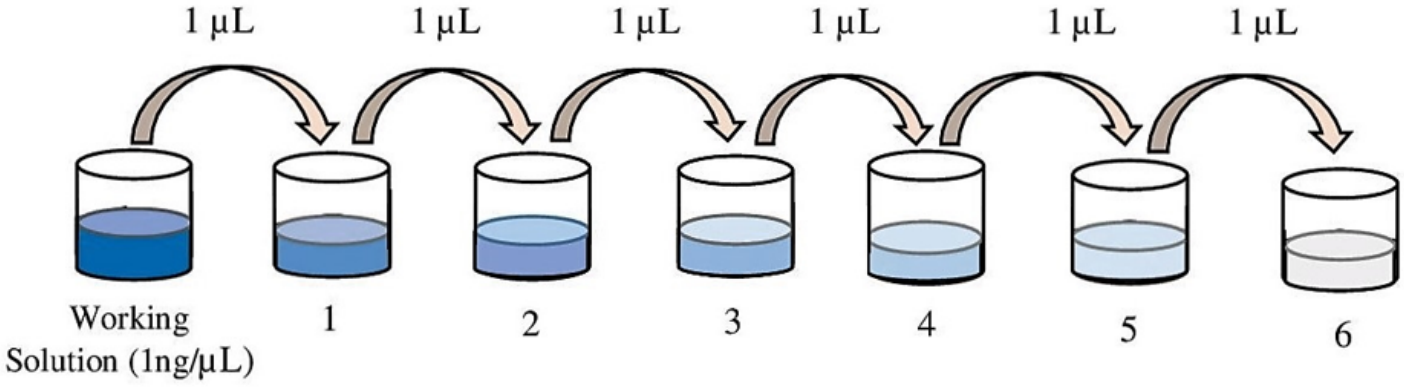

Figure 1

Serial dilution of the calibration sample.

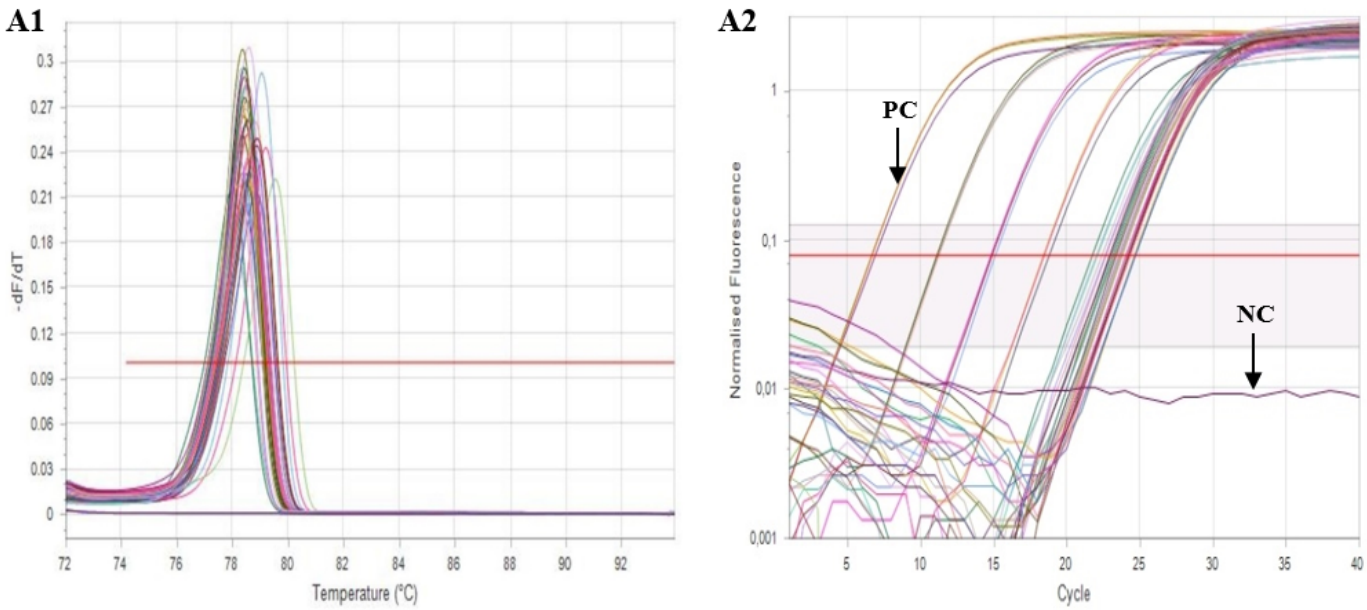

Figure 2

Melting curve and normalised fluorescence profile of the *Pfcoronin* gene amplification. A1 = melting curve A2 = normalised fluorescence. **PC*** = Positive Control. **NC*** = Negative Control.

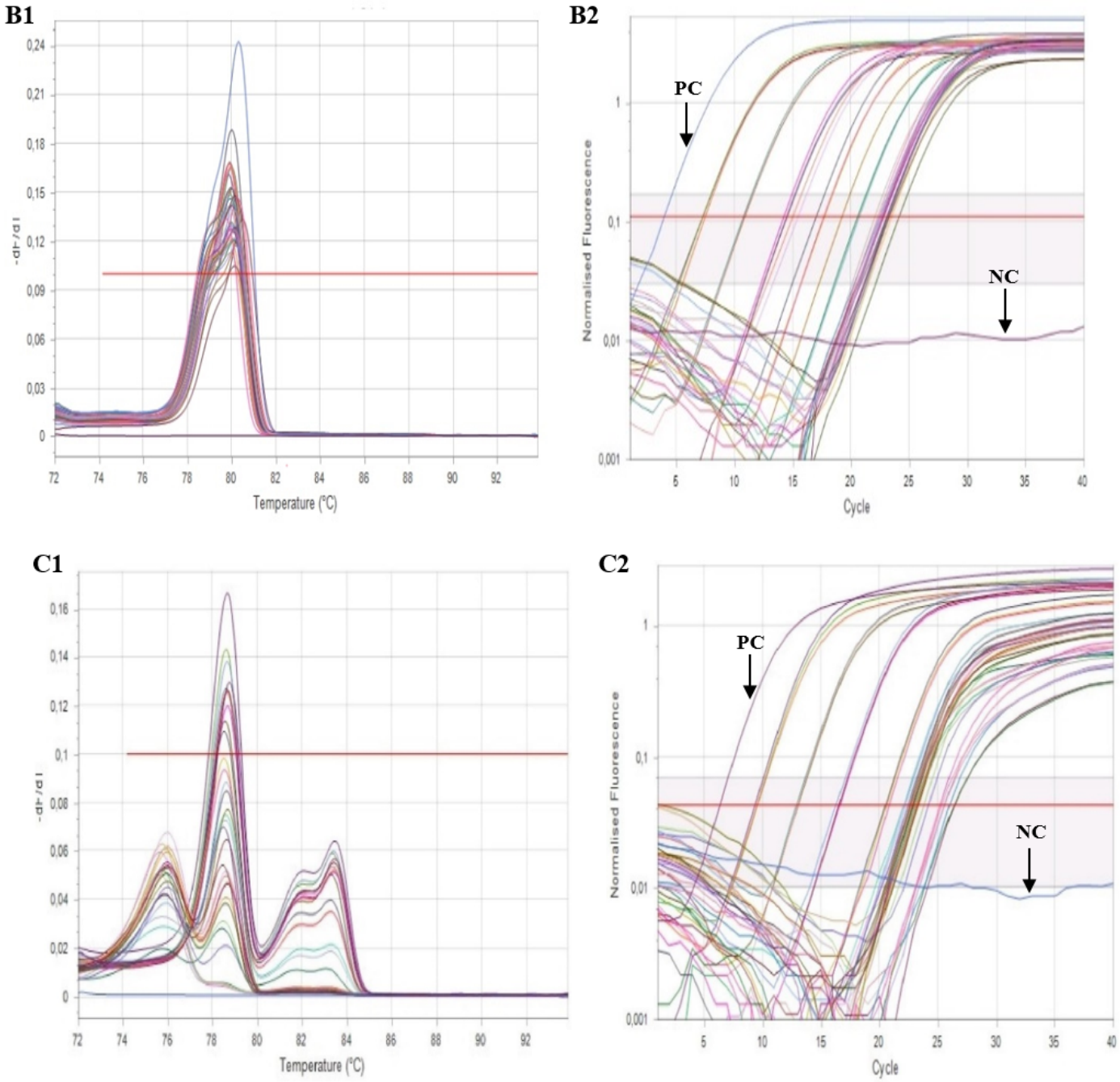

Figure 3

Melting curves and normalised fluorescence profiles of the *Pfcysteine desulfurase* gene (B1&B2), and *Pfplasmepsin 2* gene (C1&C2). **PC*** = Positive Control. **NC*** = Negative Control.

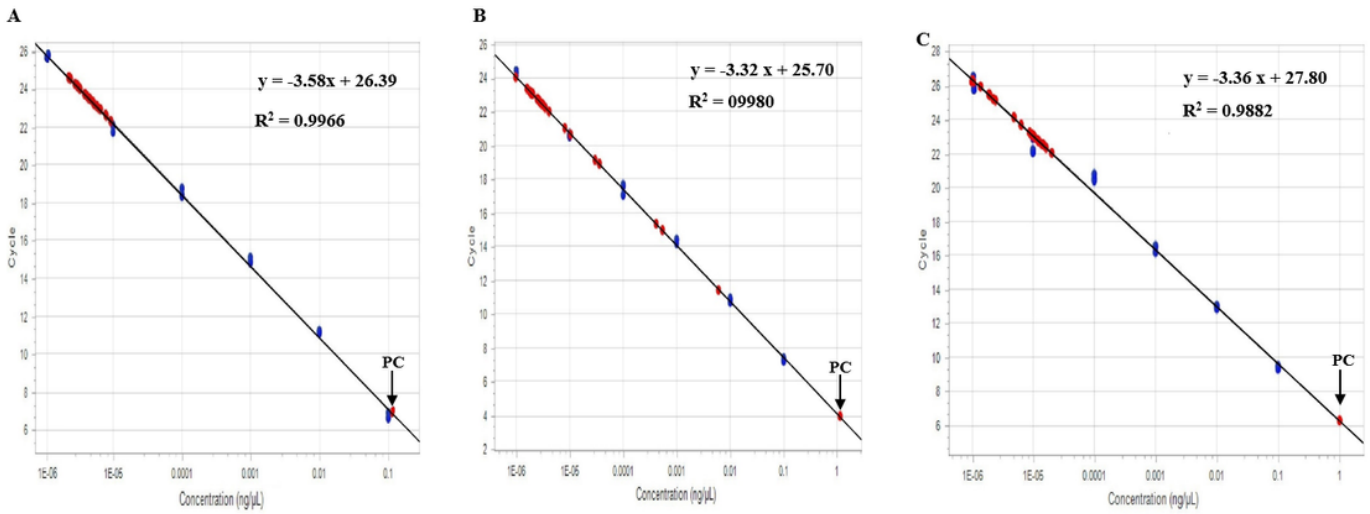

Figure 4

Standard curves. A= *Pfcoronin.* B = *Pfcysteine desulfurase*, C = *Pfplasmepsin 2.* The blue **PC* =** Positive Control. The **blue dots** represent the standards dilution, and **red dots** represent the test samples.

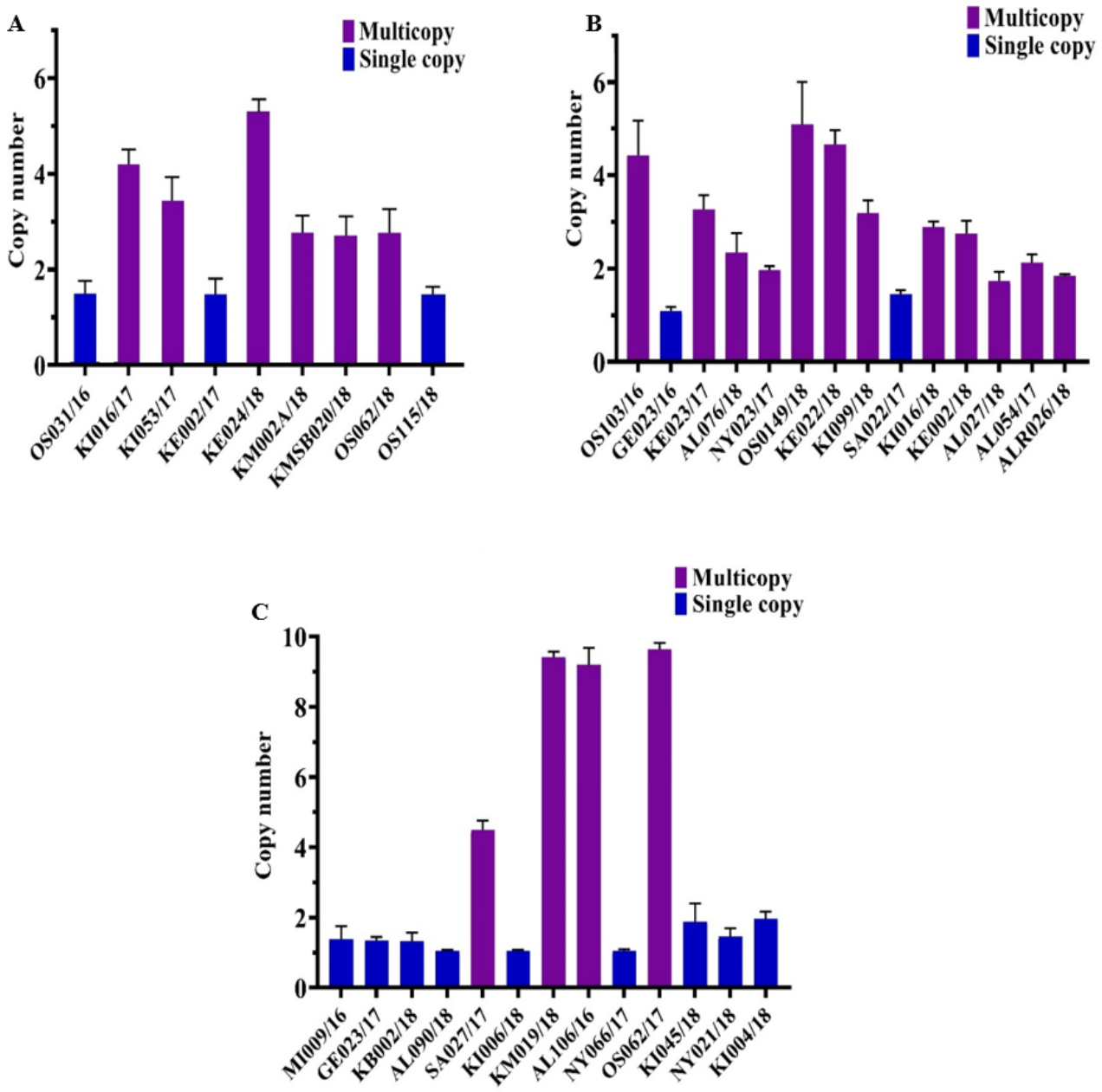

Figure 5

Copy number variation. A = *Pfcoronin.* B = *PfCysteine desulfurase*, and C = *Pf.plasmepsin 2*. The X axes represent explored isolates ID.

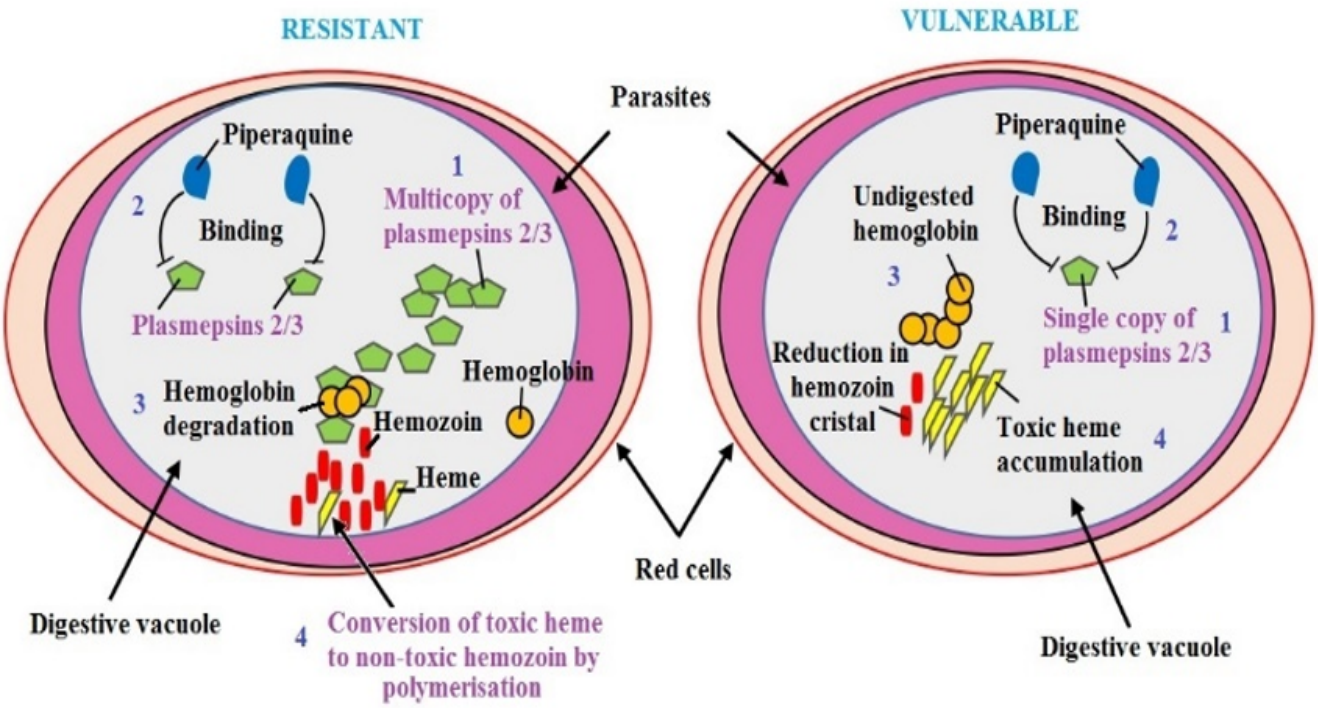

Figure 6

Illustration of plasmepsin 2 gene copy amplification in the mechanism of piperaquine resistance. **RESISTANT parasite:** (1) extra copy of plasmepsin 2 (multicopy) trigger haemoglobin degradation while piperaquine neutralizes some copies (2). (3) Hb digestion produces heme which is converted to hemozoin (4). **VULNERABLE parasite:** (1) Single copy of plasmepsin 2 is well-neutralised (2), therefore, undigested Hb and the toxic heme accumulate (3) while non-toxic hemozoin cristal decreases (4).

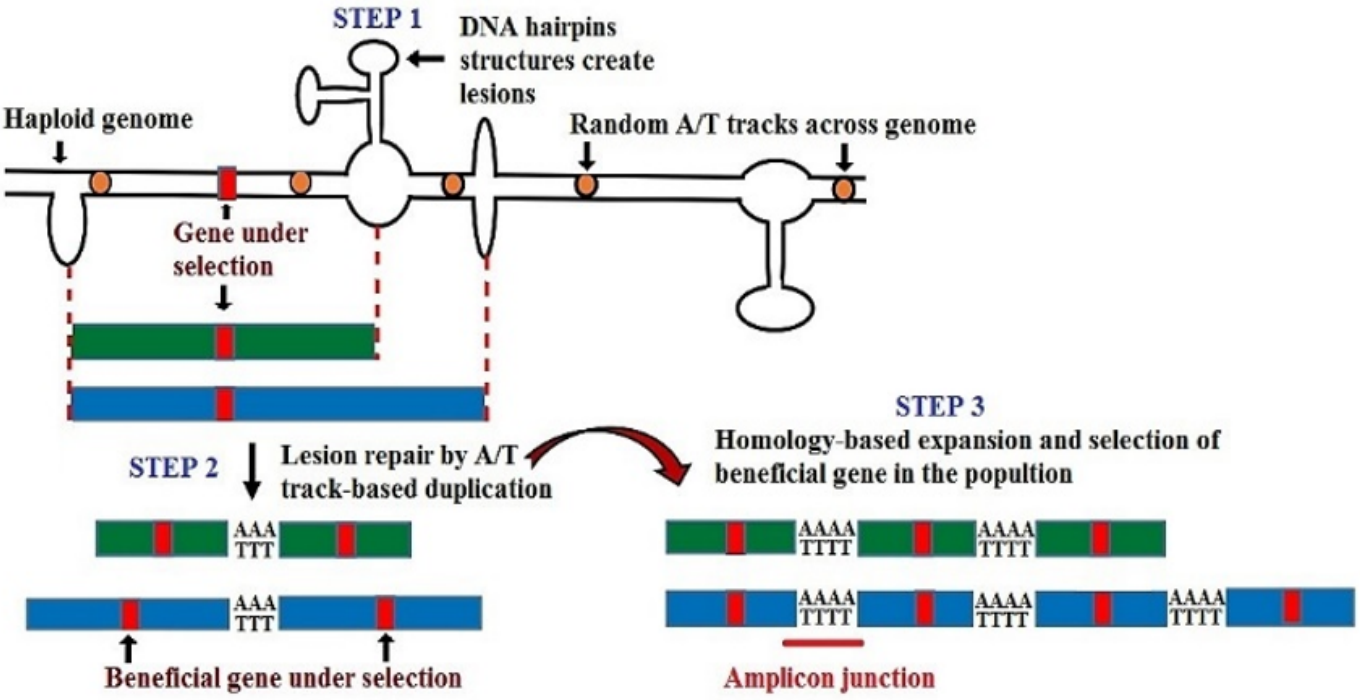

Figure 7

Illustration of a model of copy number formation in *P.falciparum.* **STEP 1,** DNA hairpin structures cause double-strand breaks across the genome presumably by the halting of the replication fork progression, by recognition of hairpin-binding proteins or by enzymatic action and polymerase pausing impact. **STEP 2**, A/T-tracks (orange circles) trigger homology mediated pathways across DNA breaks generating CNVs (green and blue horizontal bars). Red vertical dotted lines represent the copy breakpoints that are randomly generated throughout the genome. **STEP 3**, homology-based expansion acts to conserve beneficial copies. The regenerated CNVs can harbour beneficial genes (red bars within the horizontal bars), under conditions like drug selection, the beneficial copies are conserved.