

Genomic analysis reveals independent evolution of *Plasmodium falciparum* populations in Ethiopia

Deriba Abera Beyene

Department of Biochemistry, College of Health Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya.

Caleb Kipkurui Kibet

International Centre for Insect Physiology and Ecology, Nairobi, Kenya.

Teshome Degefa

School of Medical Laboratory Science, Faculty of Health Sciences, Jimma University

Lucas Amenga-Etego

Western African Center for Cell Biology of Infectious Pathogens WACCBIP), University of Ghana, Accra, Ghana

Joel L Bargul

Department of Biochemistry, College Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Lemu Golassa (✉ Igolassa@gmail.com)

Addis Ababa University, Aklilu Lemma Institute of Pathobiology <https://orcid.org/0000-0002-1216-8711>

Research

Keywords: *Plasmodium falciparum*, Ethiopia, population structure, drug resistance, admixture, positive selection

Posted Date: February 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-75436/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on March 4th, 2021. See the published version at <https://doi.org/10.1186/s12936-021-03660-y>.

Abstract

Background: *Plasmodium falciparum* parasite populations in Ethiopia have been experiencing local selective pressures from drugs and immunity, leading to evolutionary adaptation. However, there was the paucity of data on the genomic characterization and the evolutionary adaptations of *P. falciparum* isolates from the central area of Ethiopia.

Method: Whole-genome analysis of 25 *P. falciparum* isolates from central Ethiopia, specifically from West Arsi, were studied to determine their genetic diversity, population structures, and signatures of selection in known drug resistance alleles against global isolates from Cambodia, Thailand, DR Congo, and Malawi.

Result: A total of 18,517 high-quality single-nucleotide polymorphisms (SNPs) were identified in Ethiopian *P. falciparum* isolates. About 84% of the Ethiopian *P. falciparum* isolates had an F_{WS} value >0.95 showing a dominant single genotype infection in most isolates at the time of collection with little potential for out-crossing as expected in areas with low transmission intensity. Within host diversity of Ethiopian infections was significantly different from East African ($p < 0.001$) but not Southeast Asian infections ($P > 0.05$). We observed significant population structure by PCA and population differentiation between Ethiopian parasites and East Africa ($F_{st} \sim 10\%$) and Southeast Asia populations ($F_{st} \sim 18\%$), suggesting limited gene flow and the independent evolution of the Ethiopian parasite population. Moreover, we found a total of 125 genes under balancing selection that included *ama1*, *trap*, *eba175*, and *lsa3* previously identified as targets of human host immunity. Recent directional selection analysis using integrated standardized haplotype score (IHS) did not detect any selection signatures in the *Pfcr*, *Pfdhfr*, *Pfdhps*, *Pfmdr1*, and *PfK13* genes. However, known drug resistance-conferring mutations analysis showed that at least one SNP marker was fixed in these genes, but not in *Pfdhps* and *PfK13*.

Conclusion: *Plasmodium falciparum* population in the central region of Ethiopia was structurally diverged from both Southeast Asian and other East African populations. Malaria infections in Ethiopia had low within-host diversity, and parasites carry fixed chloroquine resistance markers despite the withdrawal of this drug for the treatment of *P. falciparum*.

Background

Plasmodium falciparum malaria remains one of the major public health problems worldwide accounting for 228 million cases in 2018 compared to 231 million in 2017, while the number of deaths due to malaria decreased by just 2.5%, from 415,000 to 405,000 during the same period [1]. Sub-Saharan Africa (sSA) still accounts for 94% of global death. In Ethiopia, more than 75% of the total area is malarious and *P. falciparum* and *P. vivax* co-exist [2] making malaria control more complicated than in other African countries.

Across malaria-endemic regions, large-scale deployment of antimalarial drugs has led to the emergence of drug resistance to chloroquine (CQ) and Sulfadoxine/Pyrimethamine (SP) antifolate drugs [3–5]. Like many other countries, Ethiopia has switched from CQ to SP in 1998 and from SP to AL in 2004 [6] for the treatment of uncomplicated *P. falciparum* malaria in response to the development of parasite resistance. However, CQ remained the first-line drug for *P. vivax* treatment in the country [7] leading to a continued selection of CQ-resistance markers in *P. falciparum* as the result of indirect pressure from CQ and the presence of mixed infections of *P. falciparum* and *P. vivax*. Similar to CQ and SP, *P. falciparum* developed resistance to AL first at the Thai-Cambodian border [8] and recently in East Africa [9]. The rapid development of resistance in *P. falciparum* to series of the first-line antimalarial hinders malaria prevention, control, and elimination efforts.

Antimalarial drugs are known to pose tremendous selective pressure on *P. falciparum* leading to the worldwide spread of resistant parasites [10]. It was well noted that *P. falciparum* resistance to the two conventional antimalarial drugs, CQ and SP, has resulted in increased malaria morbidity and mortality across endemic settings. Apart from increased morbidity and mortality, selective sweeps of drug resistance mutations have reduced levels of polymorphism in *P. falciparum* as these resistant and sensitive strains continue to recombine in mosquitoes [11,12] with perhaps a reduced diversity around the selected loci. However, the greatly reduced level of diversity across the entire *P. falciparum* genome most likely resulted from a recent severe population bottleneck, which is most plausibly explained by the gorilla-to-human cross-species transmission event [13].

Based on the analysis made on 12 strains collected from different countries in Africa and Asia, the average diversity of *P. falciparum* at four-fold degenerate sites was estimated to be 8×10^{-4} per site [13]. However, published mutation rates for *P. falciparum* were in the range of $1-10 \times 10^{-9}$ mutations per site per replication cycle [14,15]. Depending on *P. falciparum* life cycle and assuming varying lengths of time that the parasites spend either in the vector or in the mammalian host, the parasites are likely to undergo at least 200 replication cycles per year suggesting that the observed level of genetic diversity in *P. falciparum* could have readily accumulated within the past 10,000 years [10].

Plasmodium falciparum parasite change and select its new genetic variants owing to drug exposure to cause disease and overcome challenges from host immunity and therapeutic interventions [16]. Indeed, high pressure from immunity and drugs are known to select adaptive parasite strains that maintain transmission [17] and therefore many *P. falciparum* genes encoding immune and drug targets are under natural selection and show signatures of balancing or directional selection [4,17–21]. This selection may vary due to differences in innate susceptibility of human populations, variations in ecological transmission, resulting in varying degrees of acquired immunity and/or drug pressure. Malaria parasites from low and high endemic regions have a distinct opportunity for transmission and host acquired immune responses [22]. For effective management of malaria control and intervention strategies, it is important to determine genetic variation patterns due to parasite adaptation to host environments and drug interventions. Balancing selection brings the favored alleles of parasites to an intermediate equilibrium where they are maintained as genetic polymorphisms, while the directional selection forces cause the parasite's genetic variants to increase in frequency and facilitate the occurrence of selective sweeps around the affected loci [23].

Plasmodium falciparum population genomics has been highly studied in West African populations and showed signatures of balancing selection on multiple candidate vaccine antigens and strong directional selection around known drug resistance genes [19,24]. In contrast, there is little information about the

genomic variations of *P. falciparum* populations in the horn of Africa, including Ethiopia, where *P. falciparum* and *P. vivax* malaria co-exist and are heterogeneously distributed. A recent study reported *P. falciparum* populations in the horn of Africa, specifically in Ethiopia are unique and structurally diverged from other West, East, and central African *P. falciparum* populations [25]. These parasite populations share a chunk of genes with other sub-Saharan African *P. falciparum* populations across drug and immune targets and facilitate the spread of drug-resistant strains [25]. These studies call for in-depth analysis of Ethiopian parasite genomes to deepen understanding of genome diversity and natural selection in Ethiopia's unique human populations with co-species transmission dynamics.

Understanding the population genetic diversity of *P. falciparum* strains circulating in the specific region of central Ethiopia is very important to monitor the effectiveness of control schemes and provide baseline information for making informed decisions by the national malaria control program [26]. This study aimed to characterize the *P. falciparum* population in West Arsi, in central Ethiopia, using whole-genome analysis of data generated by Illumina next-generation sequencing.

Methodology

Study area and population

The study was conducted in West Arsi, Oromia (07° 17' 34.2 S, 038° 21' 46.3 W) located about 251 km from Addis Ababa, Ethiopia. This region with distinct wet and dry seasons has an altitude of about 1500 – 2300m above sea level with the human population of 176,671. The inhabitants of this malarious region have high levels of poverty worsened by the malaria diseases caused predominantly by *P. falciparum* and *P. vivax* with a seasonal and unstable pattern of transmission [7].

Sample collection and processing

Venous blood (2-5mL) was collected from July 2012 to December 2013 from consented *P. falciparum* malaria patients following standard procedures. Sequencing of 34 *P. falciparum* samples from leukocyte-depleted infected whole blood was done as described in [27] at the Wellcome Sanger Institute as part of the MalariaGEN *P. falciparum* Community Project (www.malariagen.net/projects). Freely available *P. falciparum* sequence data were accessed via the Pf3K project (<https://www.malariagen.net/data/pf3k-5>) for Southeast Asian and East African samples. We found only DR Congo and Malawi from East Africa and randomly took Cambodia and Thailand. Sample collection site with the number of samples greater or equal to Ethiopian samples with closer/similar *P. falciparum* samples collection year to Ethiopian sample were randomly selected. Then, 50 samples were randomly selected from each country.

Short sequence reads were generated on the Illumina HiSeq platform and aligned to Pf3D7 reference (version 3) by Burrows-Wheeler-aligners (BWA). SNP calling was done following a customized genome analysis tool kit (GATK) pipeline. Each sample was genotyped for polymorphic coding SNPs across the genome, ensuring a minimum of 5x paired-end coverage across each variant per sample. Polymorphic sites within hyper-variable, telomeric, and repetitive sequence regions were excluded. Biallelic high-quality SNPs with mapping quality (MQ) >20 and Variant Quality Score (VQSLOD) >=3 in the core region loci with a minor allele frequency of at least 2% and individual sample with less than 10% missing data and SNP-site missing less than 10% across the isolate was extracted and used for downstream analysis. After quality filtering, we left with 46, 50, 25, 50, 49 samples of Cambodia, DR Congo, Ethiopia, Malawi, and Thailand respectively.

Analysis of population genetic diversity and within-host infection diversity

The genome-wide F_{WS} (inbreeding coefficient within a population) metric was used to calculate within-host diversity as described in [21]. To derive F_{WS} ($=1 - H_W/H_S$), within isolate expected heterozygosity (H_W) was calculated from the relative allele frequencies for all genic SNPs, averaged across the genome and compared with the heterozygosity of local population (H_S). F_{WS} value ranged from zero to one, where zero indicates high diversity of infection, and one represents a single infection within the sample as compared to local population diversity. For this analysis, individual alleles with coverage of less than < 5 reads and positions with total coverage of < 20 reads were classified as undermined (missing). Isolates with greater than > 10% missing SNP data and SNPs with >10% missing isolate data were discarded. Isolates with F_{WS} scores of >0.95 were classified as single predominant genotype infections.

Population structure and admixture analysis

Principal component analysis (PCA) was used to estimate population structure using the *gI*PCA function in the open-source R statistical software version 3.6.2. The first 10 principal components axis (PCs) were calculated and the first three PCs which explained the majority of the variation in the data were retained. The data was thinned down by pruning SNPs with pairwise linkage disequilibrium (LD) by r^2 greater than 0.05 for determining the PCs. The pruned SNP loci employed in the *gI*PCA function was used to calculate an allele sharing matrix in custom R scripts. This function uses variance between and within groups to determine population genetic structure. A discriminant analysis of principal components (DAPC) [30] was used to transform the PCA data, and perform discriminant analysis on the retained principal components using the *adegenet* package in the R software version 3.6.2. Population admixture was determined based on spatial modeling of allele sharing among geographical coordinates of sampling sites. DAPC determines ancestry proportions and membership probability modeled on genetic variation across space to determine admixtures as described in [30].

Allele frequency and differentiation analysis

Analyses of allele frequency distributions between-population F_{ST} values [31] were calculated using VcfTools or *hierfstat* package from *adegenet* in R after excluding SNPs with greater than 10% missing data. For F_{ST} analysis, missing data were excluded on the SNP basis with the size of each population corrected to account for F_{ST} value difference due to population size variation.

Detection of signatures of natural selection

Within-population Tajima's D index[32] was calculated using Vcftools. Tajima D values were determined for each SNP and the average value for each gene was calculated. Genes with at least five SNP and positive TajimaD values > 1 were considered as genes under balancing selection.

The standardized integrated haplotype score (IHS) analysis was used to identify positive directional selection signatures by using phased SNP data with allele frequency $> 5\%$. IHS was determined using the *rehh* package in R software with default parameters[33] after imputing missing SNP data using Beagle version 5.2. The $|IHS| > 2.5$ (top 1% of the expected distribution) was used as cut off value [34] to report genes under recent directional selection as reported for genome analysis of West African *P. falciparum*[17].

Results

Sequencing of *P. falciparum* and analysis of allele frequency

High-quality sequence data obtained from 25 *P. falciparum* clinical isolates (Additional file 1) collected from the West Arsi of Ethiopia enabled the identification of 672,956 biallelic SNPs with less than 10% missing SNPs data and $< 10\%$ sample missing data in the individual isolate. All isolates had 95.95% (645715/672,956) SNPs call. Sequences from the intergenic regions had lower read coverage compared to those sequences in the coding regions, and as a result, 78.92% (531120/672,956) of all SNPs called were located within genes. Of 5,058 genes analyzed, 3,370 genes had at least one SNP (Table 1, Additional file 2 Figure 1A). About 18517 SNPs were polymorphic marker in at least one sample in Ethiopian ($n=25$) samples of which 43.4% (8,037/18,517) were non-synonymous coding SNPs, 22.8% (4,222/18,517) synonymous coding SNPs, 26.6% (4,932/18,517) in intergenic regions, 3.6% (666/18,517) other intragenic regions and 3.5% (656/18,517) SNPs in intron region. Similarly, *P. falciparum* populations from Cambodia ($n=46$), DR Congo ($n=50$), Malawi ($n=50$), and Thailand ($n=49$) had 32,854, 68,476, 79,250 and 30,427 biallelic polymorphic SNPs marker in at least one sample, respectively (Table 1, Additional file 2 Figure 1B). The proportion of non-synonymous coding to synonymous coding and the intragenic to intergenic SNPs were ~ 2 or above in all populations (Table 1).

In general, all populations had a high percentage of non-synonymous coding SNPs at polymorphic marker consistent with previous findings[17]. SNPs with minor allele frequency (MAF) $< 5\%$ were common in all analyzed *P. falciparum* populations following the exclusion of monomorphic SNPs in each population. Further, SNPs with minor allele frequency of $< 5\%$ occurred more frequently in samples from Malawi than in Ethiopian isolates (Additional File 2-Figure 2).

Genomic diversity of *P. falciparum* infections

F_{WS} scores ranged from 0.837 to 0.997 (mean = 0.97, median = 0.99) for Ethiopian *P. falciparum* infections whereas the F_{WS} values in Cambodia ranged from 0.702 to 0.999 (mean = 0.962, median = 0.995), from 0.483 to 0.998 (mean = 0.94, median = 0.994) in Thailand, from 0.321 to 0.998 (mean = 0.94, median = 0.994) in DR Congo and from 0.194 to 0.997 (mean = 0.747, median = 0.762) in Malawi (Figure 1; Additional file 3).

The F_{WS} value of > 0.95 suggests that the individual samples predominantly contained a single genotype and could have other additional genotypes in lower proportions. In this study, F_{WS} values of > 0.95 were observed in 84%, 79.6%, 78%, 50%, and 36% of samples from Ethiopia, Thailand, Cambodia, DR Congo, and Malawi respectively.

The mean F_{WS} scores of the Ethiopian *P. falciparum* population were not significantly different from Cambodia's (Welch two Sample t-test, $P = 0.42$) and Thailand's (Welch two-sample t-test, $p = 0.083$) at 95% confidence intervals. However, mean F_{WS} was significantly higher in Ethiopia compared to DR Congo (Welch two-sample t-test, $p = 5.603e^{-06}$) and Malawi (Welch two-sample t-test, $p = 3.242e^{-08}$) at 95% confidence intervals.

Population structure and admixtures

Analysis using PCA revealed the presence of four clear major population groups of isolates, which were coincident with their geographical origins (Figure 2A-C).

Similarly, the findings from admixture analysis were consistent with the PCA clustering. The isolates from the three regions were distinguished. This admixture analysis showed that four major components could be differentiated with a cluster value of $K = 5$. Multiple parasite subpopulations were observed in Malawi and DR Congo parasite populations suggestive of gene flow between these two populations (Figure 3). There was no detectable gene flow between the isolates from Ethiopia and East African or Southeast Asia.

The clustering of Ethiopian *P. falciparum* isolates was consistent with the fixation index (F_{ST}) values with or without correcting for sample size. The F_{ST} values of Ethiopian isolates versus those from the two other East African regions (DR Congo and Malawi) ranged from 0.08 to 0.09, while the F_{ST} value of Ethiopian *P. falciparum* versus the two southeast Asian regions (Thailand and Cambodia) was 0.18 (Table 2).

Signatures of selection in the *P. falciparum* isolates

The Ethiopian isolates had the average Tajima's D value of 0.18 across the entire genome (One sample t-test, $p < 2 \times 10^{-16}$). 1,450 genes had at least one SNP with TajimaD value > 1 of which 125 genes had at least five SNPs with Tajima D values > 1 of which 125 genes had at least five SNPs with Tajima D values > 1 (Additional file 4). These genes include *apical membrane antigen-1 (ama1)*, *erythrocyte binding antigen-175 (eba175)*, *merozoites surface protein-1 (msp1)*, *thrombospondin related anonymous protein (trap)*, *duffy binding like merozoites surface protein (dblmsp)*, and *cytoadherence linked asexual protein 2 (clag2)*, that were previously reported for the balancing selection [24,29].

The standardized integrated haplotype homozygosity score (IHS) was applied to investigate genome-wide evidence for recent positive directional selection due to drug pressure, immune impact, or other mechanisms. Using |IHS| score of >2.5 (top 1% of the expected distribution) as a threshold for hits, 36 genes with at least one SNP that could be under significant positive selection were identified, and out of these, 15 genes had at least two SNPs (Table 3).

Thirteen (13) out of the above 15 genes under positive directional selection showed both positive balancing and directional selections (Table 4) and these genes include the vaccine candidate gene *SURF4.2* on chromosome 4 and *CLAG8* (cytoadherence linked asexual protein 8) on chromosome 8 [36].

Interestingly, our analysis failed to detect selection signals in drug-resistance genes such as *Pfcr1*, *Pfmdr1*, *Pfdhfr*, and *Pfdhps*. The reason could be that IHS may not be suitable for detecting positive selection for those SNPs that have reached or are near fixation in the local *P. falciparum* population [34].

Prevalence of mutations conferring antimalarial drug resistance in *P. falciparum*

Table 5 shows inter-population differences in the prevalence of drug resistance genes observed among the *P. falciparum* global datasets analyzed. In tandem with previous studies [20,37] that suggest temporal differences in the geographical distribution of antimalarial drug resistance mutations, we observed that CQ-resistance alleles (*Pfcr1-K76T*, *Pfcr1-A220S*, and *Pfcr1-Q271E*) were fixed in Ethiopia, Cambodia, and Thailand, regions where malaria transmission rates are comparably low. But, the prevalence of these same alleles was 0% in Malawi and ranged from 66% to 72% in DR Congo.

Similarly, drug resistance mutations in *Pfmdr1* (*Pfmdr1-N86Y* and *Pfmdr1-Y184F*) were also variable among populations. For instance, the Ethiopian parasite population showed the presence of 14% *Pfmdr1-N86Y* and 100% *Pfmdr1-Y184F* gene mutations, whereas *Pfmdr1-N86Y* was detected in 48% of DR Congo isolates and in 3% of Malawi's. Also, the *Pfmdr1-Y184F* drug resistance marker was detected in 58% of the *P. falciparum* population in Cambodia, 32% in DR Congo, 35% in Malawi, and 6% in Thailand's parasite isolates.

Sulfadoxine/pyrimethamine drug resistance mutations were also present in *Pfdhfr* and *Pfdhps* genes in all analyzed *P. falciparum* populations. The major pyrimethamine resistance-conferring alleles such as *Pfdhfr-N51I* and *Pfdhfr-C59R* were also identified in all parasite populations with fixed or near fixation in frequency. *Pfdhfr-S108N* was fixed in other *P. falciparum* populations, except in Ethiopia. The variable prevalence of drug resistance-conferring alleles were also observed in *Pfdhps* (*Pfdhps-S436A*, *Pfdhps-G437A*, *Pfdhps-K540N*, and *Pfdhps-A581G*), for the parent drug sulfadoxine resistance.

In terms of artemisinin resistance, the African population-specific *Pfk13-K189T* mutation was observed in Ethiopia (in 20% of the samples), DR Congo (17%), and Malawi (13%). This mutation was previously identified in African *P. falciparum* populations [20,37]. As previously reported [8] the validated and most characterized artemisinin resistance-conferring mutation *Pfk13-C580Y* was identified in Cambodia (36% of the samples) as well as in Thailand (26%), but not in Africa.

Discussion

The transmission dynamic coupled with the unique history, ecology, and demography of Ethiopia raise interest in the genetics of its parasite population. High-resolution whole-genome SNP data was used to analyze *P. falciparum* parasite genetic diversity in the central region of Ethiopia and compared with similar parasite data from mainland Africa (DR Congo and Malawi) and Southeast Asian parasites from Cambodia and Thailand. We observe similar MAF across all five parasite populations with over-representation of low frequency (<5%) variants as previously reported [19,21]. Interestingly, mean F_{WS} values were significantly higher in the Ethiopian parasite isolates as compared to the other African populations but not the Southeast Asian parasite populations. F_{WS} is a genome-wide metric that averages heterozygosity across the genome in comparison with heterozygosity within the local parasite population [21]. Hence it is a measure of within-host diversity of infections that allows us to gauge the potential for inbreeding (or outcrossing). The higher F_{WS} values (> 0.95) in Ethiopia (*P. falciparum* prevalence of 0.02) [38] and East Asian infections is underscored by the low malaria transmission rates in these settings which supports a higher inbreeding and clonal propagation of infections (Figure 1; Additional file 3). Unlike the other East African countries (DR Congo and Malawi) where transmission intensities are higher [20,37], and there was a good distribution of F_{WS} values with the majority of infections being polyclonal with high potential for outcrossing (Figure 1). These findings are supported by similar studies that link lower F_{WS} values to in west African where transmission is high [18]. However, of note, is the possibility for high F_{WS} values to occur in areas of high transmission intensity if *P. falciparum* circulates in a geographically isolated community which limits the chance of outcrossing with other genetically distinct *P. falciparum* parasites as observed in the previous study [21].

An analysis of parasite population structure within and between continents revealed a higher degree of population structure between Ethiopian parasites and other east African parasites and between Southeast Asia and East Africa. However, neither PCA (Figure 2) nor admixture analysis (Figure 3) could resolve parasite populations in DR Congo and Malawi. These observations are corroborated by several studies that report regional and inter-continental level structure in global *P. falciparum* parasite populations [21]. However, the separation of Ethiopian parasites from the two East African populations is worth noting. Notwithstanding the increased human mobility between Addis Ababa and the rest of Africa, particularly East Africa, there remain important barriers to gene flow between parasite populations in central Ethiopia and the rest of the sub-region. We believe one of the factors that severely limit gene flow between Ethiopia and its neighbors is the local malaria transmission intensity as a function of poor vectoral capacity determined by the ecological landscape (highlands).

Against the backdrop of this unique eco-epidemiology of *P. falciparum* malaria in Ethiopia, we explored the mechanisms of natural selection by Tajima D and IHS. We identified many antigenic genes to be under balancing selection with Tajima D value greater than one in Ethiopia. These genes included known vaccine candidates such as *ama1*, *trap*, *msp1*, *eba175*, and *clag2* (Additional file 4) which were previously identified in different populations that vary in transmission intensity [24,29,39], to be under balancing selection. Besides, we identified 15 genes under positive directional selection by IHS, which includes *SURFIN* and *PHIST* families previously suggested to be targets of immunity [24]. We believe the low seasonal transmission in Ethiopia maintains significant immune selection pressure on the infection reservoir than drug pressure due to clinical malaria. Therefore, the candidate vaccine antigen loci under balancing

selection may be largely due to immune modulation and not positive adaptive selection influenced by drug pressure. This is supported by our failure to detect selection signatures in known drug target genes such as *Pfcr1*, *Pfmdr1*, *Pfdhfr*, *Pfdhps*, and *Pfkelch13*. Our data showed that the ability of IHS to detect selection in these drug resistance genes in Ethiopia may be because the frequency of polymorphisms in these loci are either fixed or near fixation in the Ethiopian population (Table 5). These findings are supported by a previous study in Ethiopia which showed that the CQ-resistant haplotype (CVIET) was fixed [7]. The continued use of CQ in Ethiopia for the treatment of *P. vivax* malaria may account for the high prevalence of CQ resistant markers. Also, *Pfmdr1* mutations have been demonstrated to mediate AL resistance. Therefore, the high prevalence of *Pfmdr1* mutations may signal poor efficacy of AL as the first treatment for *P. falciparum* malaria in Ethiopia. Variable prevalences of CQ-resistant polymorphisms were observed only in DR Congo and not in Malawi, evidence that supports the complete reversal of CQ susceptibility in Malawi as reported by Ochola et al [20].

Undoubtedly, artemisinin resistance has taken root in Southeast Asia. Despite 36% and 26% prevalence of *PfKelch13-C580Y* mutation in Cambodia and Thailand samples respectively, no validated *PfKelch13* mutation was found in the African samples. However, an uncharacterized *Pfkelch13* mutation (*PfK13-K189T*) found at prevalence >10% in all the African datasets and previously reported in other studies [37], may be important but its role in artemisinin resistance is unknown. One study [8] reported that mutation in *Pfkelch13* at amino acid positions less than 441 may not play any role in mediating artemisinin resistance. But, validated *Pfkelch13-R561H* mutation for artemisinin resistance was recently reported in other East African *P. falciparum* population [9].

Conclusion

Overall, our study revealed a comparably low genetic diversity of *P. falciparum* parasites in Ethiopia. The majority of infections were of low complexity, demonstrated significant population structure with Ethiopian parasites diverged from parasite populations within the sub-region. We highlight limited gene flow between parasite populations in the East African sub-region and Ethiopia. We also reported balancing selection in antigenic loci known to be targets of immunity and adaptive positive selection in *SURFIN* and *PHIST* gene families that are potential vaccine antigens. Though selection analysis did not pick up any adaptive mutations in known drug-resistant genes, we reported fixation of the CQ-resistance *Pfcr1-K76T* genotype and others in Ethiopia and the wild-type genotype (*K*) in Malawi. We reported no *PfKelch13* validated mutations in Ethiopia, DR Congo, and Malawi except a *PfK13-K189T* African specific uncharacterized mutation. Further molecular studies involving deeper sampling of Ethiopian parasite populations are essential to understand the genetic diversity, gene flow, and temporal evolution of drug resistance loci within Ethiopia. Our findings can be used to support national malaria control decision making for optimal impact in further reducing malaria transmission in Ethiopia.

List Of Abbreviations

ACT: Artemisinin Combined therapy; AL: Artemether-Lumefantrine; ART: Artemisinin; CQ: Chloroquine; FWS: Within infection diversity fixation index; FST: Population differentiation fixation index; HW: Heterozygosity within infection; HS: Heterozygosity within a local population; IHS: Standardized Integrated haplotype Score; IRS: Indoor Residual Spray; ITN: Insecticide-treated Nets; *P. falciparum*: *Plasmodium falciparum*; *Pfcr1*: *Plasmodium falciparum* chloroquine; *Pfdhfr*: *Plasmodium falciparum* dihydrofolate reductase; *Pfdhps*: *Plasmodium falciparum* pteroyl synthase; *PfK13*: *Plasmodium falciparum* kelch-13; *P.vivax*: *Plasmodium vivax*; SNP: Single Nucleotide Polymorphisms; SP: Sulfadoxine/Pyrimethamine

Declarations

Ethics approval and consent to participate

The use of human subjects and scientific merit for this study was approved by the institutional Ethical Review Board (IRB) and the scientific committee of Addis Ababa University and AHRI-ALERT (Armauer Hansen Research Institute and the Africa Leprosy Rehabilitation and Training Hospital). Written informed consent was obtained from all adult subjects and the parent or legal guardians of minors.

Consent for publication

Not applicable

Availability of data and materials

Datasets generated and/or analyzed during the study are available through the MalariaGEN Pf3K Project. The *P. falciparum* genome sequences used in this study are available in the ENA and SRA databases (see Additional file 1 for accession numbers)

Competing interests

The authors declare that they have no competing interests.

Funding

This project was funded by the Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME). DELGEME is a training program sponsored by the Wellcome Trust Developing Excellence in Leadership, Training and Science Africa (DELTA Africa) initiative in partnership with the Department of International Development (DFID) and the Alliance for Accelerating Excellence in Science in Africa (AESAs).

Authors' contributions

DA participated in study design, data analysis, interpretation, and writing manuscript, CK, TD, LAE, and JLB reviewed the manuscript, and LG participated in data collection, DNA extraction, genome sequencing, and manuscript reviewing.

Acknowledgments

Plasmodium falciparum genome sequencing was done at the Wellcome Sanger Institute as part of the MalariaGEN *P. falciparum* Community Project (www.malariagen.net/projects). We thank the MalariaGEN *P. falciparum* Community Project and Pf3K Project for allowing access to sequence data.

JLB is supported by DELTAS Africa Initiative grant # DEL-15-011 to THRIVE-2. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust grant # 107742/Z/15/Z and the UK government. The views expressed herein do not necessarily reflect the official opinion of the donors.

References

1. WHO. World Malaria Report 2019. Geneva: World Malaria Report; 2019.
2. Lo E, Hemming-Schroeder E, Yewhalaw D, Nguyen J, Kebede E, Zemene E, et al. Transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum* in Ethiopia and prevalence of antimalarial resistant genotypes. *PLoS Negl Trop Dis*. 2017;11:1–25.
3. Abdifatah AJ, Wanna C, Kesara N-B. *Plasmodium falciparum* drug resistance gene status in the Horn of Africa: A systematic review. *African J Pharm Pharmacol*. 2018;12:361–73.
4. Amambua-Ngwa A, Jeffries D, Amato R, Worwui A, Karim M, Ceessay S, et al. Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from the Gambia. *Sci Rep*. 2018;8:1–10.
5. Mekonnen SK, Aseffa A, Berhe N, Teklehaymanot T, Clouse RM, Gebru T, et al. Return of chloroquine-sensitive *Plasmodium falciparum* parasites and emergence of chloroquine-resistant *Plasmodium vivax* in Ethiopia. *Malar J*. 2014;13:1–9.
6. National five-year strategic plan for malaria prevention & control in Ethiopia. https://extranet.who.int/countryplanningcycles/sites/default/files/planning_cycle_repository/ethiopia/five_year_malaria_control_sp_2006_2010.pdf. Accessed 20 December 2019.
7. Golassa L, Enweji N, Erko B, Aseffa A, Swedberg G. High prevalence of *Pfcr1*-*CVIET* haplotype in isolates from asymptomatic and symptomatic patients in South-central Oromia, Ethiopia. *Malar J*. 2014;13:1–8.
8. Ashley E., M.Dhorda, R.M.Fairhurst, Amaratunga C, Lim P, Suon S, et al. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *new Engl J Med*. 2014;371:411–23.
9. Uwimana A, Legrand E, Stokes BH, Ndikumana JM, Warsame M, Umulisa N, et al. Emergence and clonal expansion of in vitro *Pfkelch13 R561H* mutant parasites in Rwanda. *Nat Med*. 2020;1–14.
10. Loy DE, Liu W, Li Y, Learn GH, Plenderleith LJ, Sundararaman SA, et al. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. *Int J Parasitol*. 2017;47:87–97.
11. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, et al. A selective sweep driven by pyrimethamine treatment in Southeast Asian malaria parasites. *Mol Biol Evol*. 2003;20:1526-36.
12. Volkman SK, Sabeti PC, Decaprio D, Neafsey DE, Schaffner SF, Milner DA, et al. A genome-wide map of diversity in *Plasmodium falciparum*. 2007;39:113–9.
13. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, et al. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun*. 2016;7:1-14.
14. Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia N V., Luna FL, et al. Mitotic Evolution of *Plasmodium falciparum* Shows a Stable Core Genome but Recombination in Antigen Families. *PLoS Genet*. 2013;9:1–15.
15. Paget-Mcnicol S, Saul A. Mutation rates in the *dihydrofolate reductase* gene of *Plasmodium falciparum*. *Parasitology*. 2001;122:497–505.
16. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, Zhou Y, et al. Erratum: A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog*. 2006;2:0562–77.
17. Oyebola KM, Idowu ET, Olukosi YA, Awolola TS, Amambua-Ngwa A. Pooled-DNA sequencing identifies genomic regions of selection in Nigerian isolates of *Plasmodium falciparum*. *Parasites and Vectors*. 2017;10:1–7.
18. Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, et al. Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. *BMC Genomics*. 2015;16:1–11.
19. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol*. 2014;31:1490–9.
20. Ocholla H, Preston MD, Mipando M, Jensen ATR, Campino S, Macinnis B, et al. Whole-genome scans provide evidence of adaptive evolution in Malawian *Plasmodium falciparum* isolates. *J Infect Dis*. 2014;210:1991–2000.
21. Manske M, Miotto O, Campino S, Auburn S, Zongo I, Ouedraogo J, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*. 2012;487:1–19.
22. Mackinnon MJ, Read AF. Virulence in malaria: An evolutionary viewpoint. *Philos Trans R Soc B Biol Sci*. 2004;359:965–86.
23. Biswas S, Akey JM. Genomic insights into positive selection. *Trends Genet*. 2006;22:437–46.

24. Amambua-Ngwa A, Park DJ, Volkman SK, Barnes KG, Bei AK, Lukens AK, et al. SNP genotyping identifies new signatures of selection in a deep sample of west African *Plasmodium falciparum* malaria parasites. *Mol Biol Evol.* 2012;29:3249–53.
25. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, et al. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. *Science.* 2019;365:813–6.
26. Jalei AA, Chaijaroenkul W, Na-bangchang K. *Plasmodium falciparum* drug resistance gene status in the Horn of Africa : A systematic review. *African J Pharm Pharmacol.* 2018;12:361–73.
27. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, et al. An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole-genome high-throughput sequencing. *PLoS One.* 2011;6:4–11.
28. Nei M, Li W-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Genetics.* 1979;76:5269–73.
29. Ye R, Tian Y, Huang Y, Zhang Y, Wang J, Sun X, et al. Genome-wide analysis of genetic diversity in *Plasmodium falciparum* isolates from China – Myanmar border. *Genetics.* 2019;10:1–8.
30. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 2010;11:1–15.
31. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *evolution.* 1984;38:1358–70.
32. Tajima F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123:585–95.
33. Gautier M, Vitalis R. *Rehh* 2.0: a reimplementation of the R package *rehh* to detect positive selection from haplotype structure. 2017;17:78-90.
34. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *Plos Biol.* 2006;4:0446–58.
35. Volkman SK, Sabeti PC, Decaprio D, Neafsey DE, Schaffner SF, Milner DA, et al. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat Genet.* 2007;39:113–9.
36. Iriko H, Kaneko O, Otsuki H, Tsuboi T, Su X Zhuan, Tanabe K, et al. Diversity and evolution of the rhoph1/clag multigene family of *Plasmodium falciparum*. *Mol Biochem Parasitol.* 2008;158:11–21.
37. Verity R, Aydemir O, Brazeau NF, Watson OJ, Hathaway NJ, Mwandagaliwa MK, et al. The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the Democratic Republic of Congo. *Nat Commun.* 2020;11:1–10.
38. Wolde MF, Berhanu MY, Bikila LE. Analysis of malaria prevalence in Arsi Negelle health. *Int Sch Journals.* 2016;3:90–5.
39. Amambua-Ngwa A, Tetteh KKA, Manske M, Gomez-Escobar N, Stewart LB, Deerrhake ME, et al. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genet.* 2012;8:1–14.

Tables

Table 1: Distribution of polymorphic SNP marker effects and their relative proportion in each *P. falciparum* population

Country of Origin	Non-synonymous coding	Synonymous coding	Intergenic	Intron	Other intragenic	Proportion of Non-synonymous to synonymous coding	Proportion of intragenic to intergenic	Total
SNP in Cambodia	13271	6063	11059	1494	806	2.19	1.96	32693
SNP in DR Congo	29762	14450	19685	3034	1255	2.06	2.46	68186
SNP in Ethiopia	8037	4226	4932	656	666	1.90	2.75	18517
SNP in Malawi	31756	15623	26538	3608	1384	2.03	1.97	78909
SNP in Thailand	12194	5489	10466	1342	792	2.22	1.89	30283

Table 2: Pairwise population Divergence (measured by F_{ST}) among *P. falciparum* populations. Ethiopian *P. falciparum* highly diverged from both Southeast Asian and East African *P. falciparum* populations.

Country	Cambodia	DR Congo	Ethiopia	Malawi	Thailand
Cambodia	0	0.16	0.18	0.17	0.06
DR Congo	0.16	0	0.08	0.02	0.16
Ethiopia	0.18	0.08	0	0.09	0.18
Malawi	0.17	0.02	0.09	0	0.17
Thailand	0.06	0.16	0.18	0.17	0

Table 3: Genes with at least two SNPs that had a recent positive directional selection in *P. falciparum* of Ethiopia, identified using the integrated haplotype score at a significance threshold of $P < 0.01$. SNPs and ID stand for single nucleotide polymorphisms and gene identification numbers, respectively.

Chromosomes	Number of SNPs	Genes Name or ID.	Product description
1	9	PF3D7_0104100	Conserved Plasmodium membrane protein, unknown function
1	3	PF3D7_0113600	Surface-associated interspersed protein 1.2 (SURFIN 1.2)
4	2	PF3D7_0424300	Erythrocyte binding antigen-165
4	14	SURF4.2	Surface-associated interspersed protein 4.2 (SURFIN 4.2)
4	5	PF3D7_0425200	Plasmodium exported protein (hyp15), unknown function
4	4	PF3D7_0425250	Plasmodium exported protein (PHIST), unknown function
7	5	PF3D7_0713900	Conserved Plasmodium protein, unknown function
7	3	CRMP2	Cysteine repeat modular protein 2
8	3	CLAG8	Cytoadherence linked asexual protein 8
10	2	PF3D7_1004800	ADP/ATP carrier protein, putative
12	2	PF3D7_1201400	Plasmodium exported protein, unknown function
13	7	PF3D7_1301800	Surface-associated interspersed protein 13.1 (SURFIN 13.1)
13	3	PF3D7_1308400	Conserved Plasmodium protein, unknown function
14	2	PF3D7_1434500	Dynein-related AAA-type ATPase, putative
14	2	PF3D7_1477500	Plasmodium exported protein (PHISTb), unknown function

Table 4: Genes under both recent positive directional Selection and positive balancing selections in Ethiopian *P. falciparum* populations. ID stands for a gene identification number.

Chromosomes	Gene Name/ID.	Product Description
1	PF3D7_0104100	Conserved <i>Plasmodium</i> membrane protein, unknown function
1	PF3D7_0113600	Surface-associated interspersed protein 1.2 (SURFIN 1.2)
4	PF3D7_0424300	Erythrocyte binding antigen-165, pseudogene
4	SURF4.2	Surface-associated interspersed protein 4.2 (SURFIN 4.2)
4	PF3D7_0425200	<i>Plasmodium</i> exported protein (hyp15), unknown function
4	PF3D7_0425250	<i>Plasmodium</i> exported protein (PHIST), unknown function
7	PF3D7_0713900	Conserved Plasmodium protein, unknown function
8	CLAG8	Cytoadherence linked asexual protein 8
10	PF3D7_1004800	ADP/ATP carrier protein, putative
12	PF3D7_1201400	<i>Plasmodium</i> exported protein, unknown function
13	PF3D7_1301800	Surface-associated interspersed protein 13.1 (SURFIN 13.1)
13	PF3D7_1308400	Conserved Plasmodium protein, unknown function
14	PF3D7_1434500	Dynein-related AAA-type ATPase, putative

Table 5: Drug resistance-conferring allele's frequency across the 5 *P. falciparum* populations

Genes	Chromosome	Position	mutation site	Ethiopia	Cambodia	DR Congo	Malawi	Thailand
<i>DHFR</i>	4	748577	<i>I164L</i>	0	0.5	0	0	0.84
<i>DHFR</i>	4	748410	<i>S108N</i>	0	1	1	1	1
<i>DHFR</i>	4	748262	<i>C59R</i>	0.86	1	0.86	0.99	1
<i>DHFR</i>	4	748239	<i>N51I</i>	1	0.95	1	1	0.95
<i>MDR1</i>	5	961625	<i>D1246Y</i>	0	0	0.17	0	0
<i>MDR1</i>	5	958145	<i>N86Y</i>	0.14	0	0.48	0.03	0
<i>MDR1</i>	5	961566	<i>F1226Y</i>	0	0.04	0	0	0.59
<i>MDR1</i>	5	958440	<i>Y184F</i>	1	0.58	0.32	0.35	0.06
<i>CRT</i>	7	405600	<i>I356T</i>	0	0.52	0.27	0	1
<i>CRT</i>	7	405362	<i>N326S</i>	0.98	0.51	0	0	1
<i>CRT</i>	7	405838	<i>R371I</i>	0	0.8	0.71	0	1
<i>CRT</i>	7	404407	<i>A220S</i>	1	1	0.663	0	1
<i>CRT</i>	7	403625	<i>K76T</i>	1	1	0.66	0	1
<i>CRT</i>	7	404836	<i>Q271E</i>	1	1	0.7143	0	1
<i>DHPS</i>	8	549685	<i>G437A</i>	0.08	0.13	0.08	0.01	0
<i>DHPS</i>	8	549995	<i>K540N</i>	0	0.4	0	0	0.03
<i>DHPS</i>	8	549681	<i>S436A</i>	0	0.2	0.11	0.02	0.17
<i>DHPS</i>	8	550117	<i>A581G</i>	0.02	0.4	0.03	0.02	0.82
<i>K13</i>	13	1726432	<i>K189T</i>	0.2	0	0.17	0.13	0
<i>K13</i>	13	1725259	<i>C580Y</i>	0	0.36	0	0	0.26

Figures

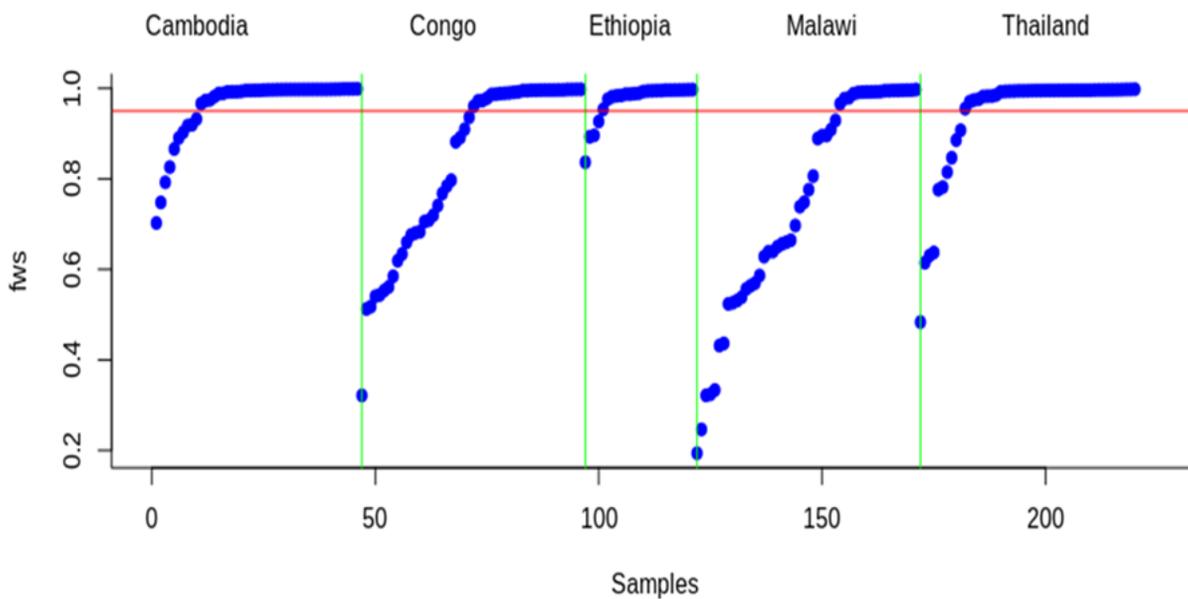


Figure 1

Ranked distribution of FWS values by population. Redline marks FWS = 0.95, above which an infection is considered clonal. A vertical green line separates the *P. falciparum* population.

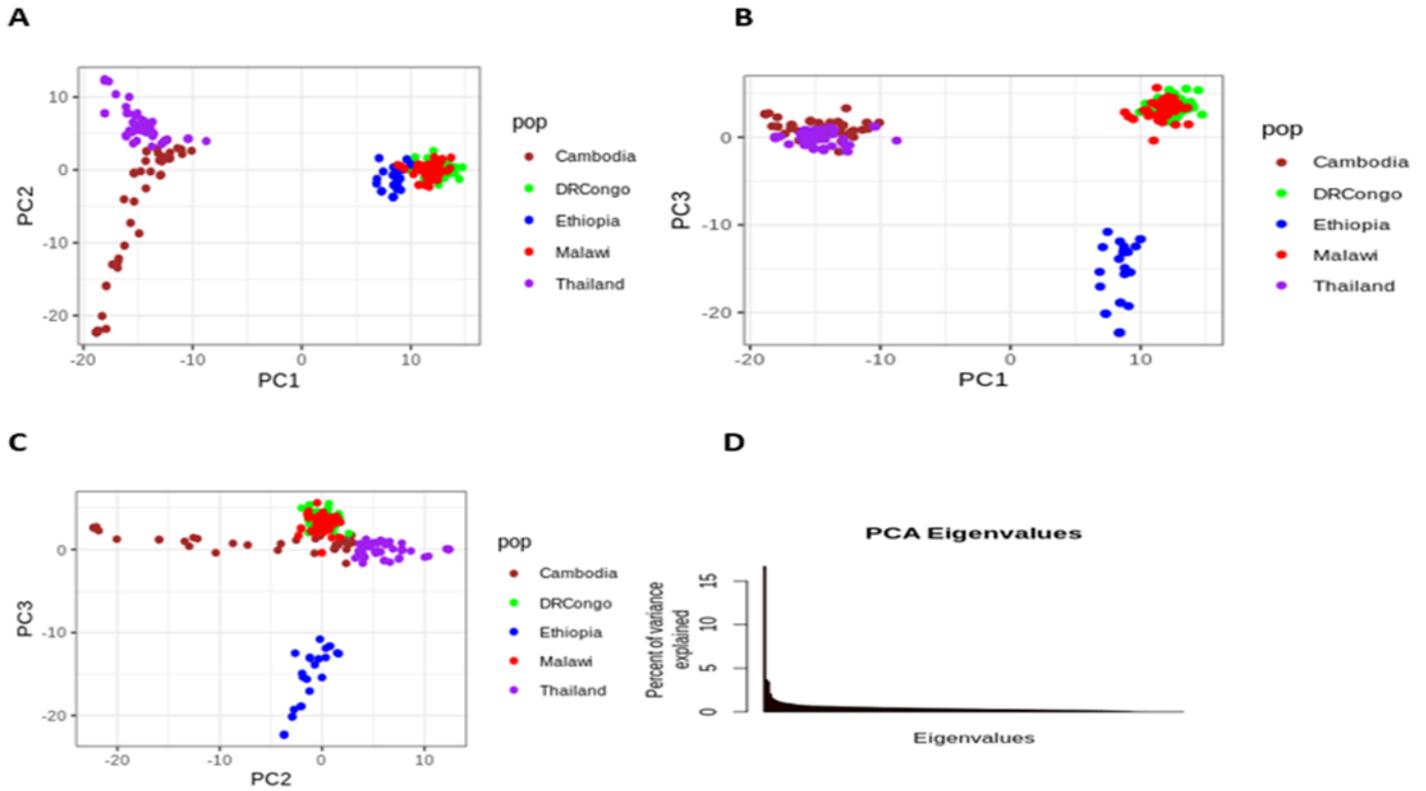


Figure 2
 Principal component analysis A) Principal component axis1 and axis2 identified Southeast Asian from African *P. falciparum* population B) Principal component axis1 and axis3 identified Ethiopian, Southeast Asian and other East African *P. falciparum* populations C) Principal Component axis2 and axis3 identified Ethiopian *P. falciparum* population from Southeast Asian and African *P. falciparum* population D) Percent of variance explained by each principal component axis. PC1 explained 16.7%, PC2 explained 3.7% and PC3 explained 3.4% of the variance in the data. Pop stands for the *P. falciparum* country of origin.

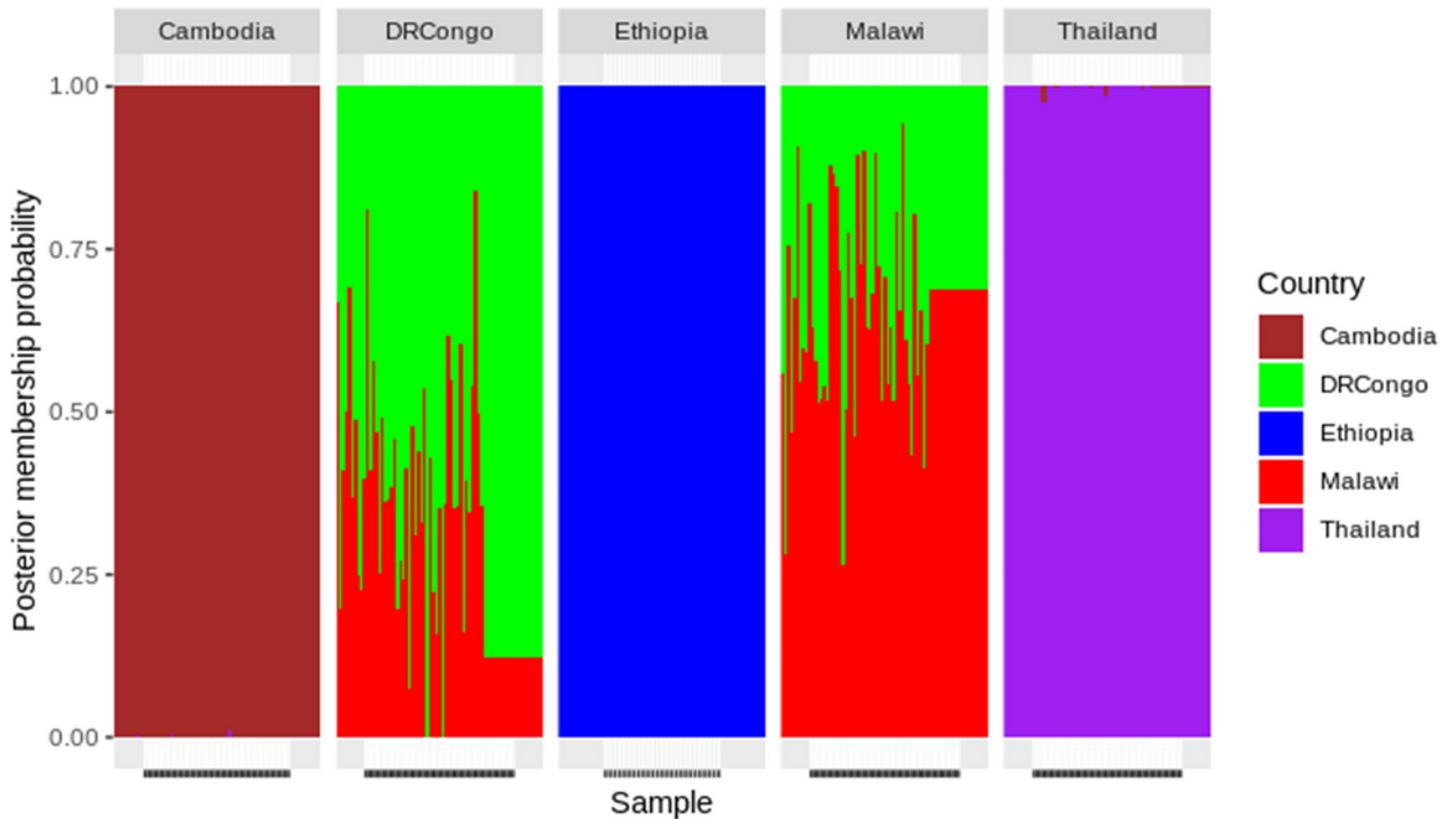


Figure 3

An admixture analysis: X-axis represent the assigned samples to country listed above the figure after modeling allele sharing probability, while Y-axis represent the membership probability of samples. Colour coded country listed at the right of figure indicated the origin of *P. falciparum* sample collection. Analysis of admixture identified four major components based on an optimized cluster value of $K = 5$, and multiple subpopulations of *P. falciparum* were found in DR Congo and Malawi. There was a bidirectional gene flow between Malawian and Congolese *P. falciparum* populations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xls](#)
- [AdditionalFile2.doc](#)
- [Additionalfile3.xls](#)
- [Additionalfile4.xls](#)