BMC Genomics

RESEARCH Open Access

Comprehensive genomic study of *Plasmodium falciparum* in Sierra leone: genetic variation and resistance patterns

Xin-jie Zhan[1,2], Shen-Bo Chen[3,4,5,6], Tian-yu Wang[3,4,5,6], Yi-wen Duan[3,4,5,6], Hai-Mo Shen[3,4,5,6,7*], Jun-Hu Chen[3,4,5,6,8*] and Jing-jing Xie[2,9*]

Abstract

Background Malaria remains a significant global health concern, with *Plasmodium falciparum* being the most dangerous of the malaria-causing parasites. Sierra Leone, with year-round transmission of malaria, continues to experience high infection rates, largely due to *P. falciparum*. Although genomic databases have provided valuable insights into malaria transmission patterns, drug resistance, and population structure, data from Sierra Leone has been limited. This study aims to build on our previous report by incorporating new samples and providing a more comprehensive genomic analysis of *P. falciparum* in Sierra Leone, with a particular focus on genetic diversity, population structure, and drug resistance.

Methods We collected *P. falciparum* samples from Freetown, Sierra Leone, between December 2022 and March 2023. A total of 35 samples underwent sequencing using the MGISEQ and Illumina platforms, resulting in high-coverage genomic data. Population structure was assessed using PCA, NJ trees, and STRUCTURE analysis, alongside comparisons with a global dataset from the *pf3k* project. Genetic diversity was evaluated using metrics such as $\widehat{\pi}$, $\widehat{\theta}_{\omega}$, Tajima's D, and iHS. XPEHH was used to examine selection pressures across different regions.

Results Sequencing yielded over 376,450 high-quality SNPs across 35 Sierra Leone isolates, indicating substantial genetic variability. PCA, NJ trees, and STRUCTURE analysis revealed that the Sierra Leone isolates formed distinct clusters from both West African and Southeast Asian samples, with evidence of region-specific genetic adaptation. The low IBD we found suggested that infections were largely independent. Meanwhile the Tajima's D and iHS analysis showed strong directional selection in genes associated with immune evasion and drug resistance genes, exhibiting ongoing evolutionary pressure due to therapeutic behavior.

*Correspondence:
Hai-Mo Shen
shenhm@nipd.chinacdc.cn
Jun-Hu Chen
chenjh@nipd.chinacdc.cn
Jing-jing Xie
651543263@qq.com

Full list of author information is available at the end of the article

Conclusion This study provides us not only a genomic database but also a profile of *P. falciparum* in Sierra Leone, revealing high genetic diversity, strong selection pressure on drug resistance genes, and unique population structures. Our data emphasize the need for continued genomic surveillance to better manage malaria.

Keywords Plasmodium falciparum, Genomic diversity, Drug resistance, Population structure, Sierra Leone

Introduction

Malaria remains one of the world's most serious public health challenges, with *Plasmodium falciparum* being the primary cause of the disease, and many studies have focused on prevention and control efforts in sub-Saharan Africa [1]. Sierra Leone is one of the countries with the highest incidence of malaria. Like many West African countries, malaria transmission occurs throughout whole year, with the majority of cases being *P. falciparum*. Despite significant efforts by health authorities to control malaria, the country continues to report high rates of malaria cases, with more than 2.6 million cases reported in 2022. This extremely high rate highlights the significant burden of malaria on the country's population [2].

The introduction of genomic technologies has enabled researchers to explore the genetic diversity of malaria parasites worldwide [3]. Genomic studies have helped to reveal key questions such as population characteristics, multiple infections, co-transmission patterns, gene flow, and the spread of drug resistance [4–6]. Although the *P. falciparum* genome was published decades ago and genomic databases have become routine tools for tracking transmission patterns and identifying markers of drug resistance, data from Sierra Leone remain limited. The lack of comprehensive genomic surveillance has made it more difficult to integrate genomic information with epidemiological data. This, in turn, has limited the effectiveness of national malaria control strategies.

In our previous study, we performed the first whole-genome analysis of 19 high-density *P. falciparum* isolates from Sierra Leone [7]. This provided the first insight into the genetic diversity and population structure of the parasite in this highly endemic region. We found high levels of repeated infections and low levels of related infections, suggesting independent transmission events. Building on this work, this study expanded the dataset to include 35 isolates from the same region. Our goal was to provide a more detailed and comprehensive analysis of the genetic landscape. By including more samples and using sophisticated population analysis tools, we hope to further demonstrate the evolutionary pressures facing the parasite in Sierra Leone. Specifically, we examined genetic diversity, population structure, and selection on genes associated with immune evasion and drug resistance. This work enhances our understanding of *P. falciparum* in Sierra Leone and contributes to a deeper understanding of malaria genomics in areas of high transmission.

Materials and methods

Sample collection

Blood samples were collected by the 23rd batch of China Medical Team from clinical malaria patients at the Sierra Leone China Friendship Hospital (Freetown) between December 2022 and March 2023. Approval for this study was obtained from the Ethics Committee of the Chinese Center for Disease Control and Prevention. All participants were fully informed about the study's protocol, potential risks, and benefits, and they provided written informed consent. Blood samples were collected from patients who tested RDT and microscopically positive for *P. falciparum* infection, which was further confirmed through PCR as single-species infections.

Next-generation sequencing

To ensure high-quality genomic data, only samples with a high parasite density were selected. DNA was extracted from frozen blood samples using a QIAGEN DNeasy Blood & Tissue Kit (Qiagen, UK). As per the requirements of the Sierra Leone Ministry of Health, an initial batch of 50 samples was sequenced locally at the Sierra Leone-China Biological Laboratory, equipped with an MGISEQ-200 sequencer donated by BGI. Due to the lack of automated equipment, sequencing libraries were manually constructed. DNA was fragmented enzymatically to achieve the desired fragment size, using the MGIEasy Universal DNA Library Prep Kit (MGI, China). Fragments of approximately 430 bp were selected after end repair, A-tailing, and adapter ligation. Circularized libraries were prepared using the MGISEQ-200RS sequencing kit, and DNA nanoballs (DNB) were loaded for sequencing. The MGISEQ-200 platform generated an average of 39.6 million paired-end reads (150 bp) per sample. Adapter sequences and low-quality reads were removed using Trimmomatic-3.0 [8], and reads were mapped to the 3D7 reference genome using BWA [9]. Variant calling was performed with the GATK4 workflow [10], yielding 165,000 high-quality SNPs after filtering for missing calls exceeding 5%. However, only 19 of the 50 samples met the required ≥ 90% genome coverage for downstream analyses.

To address this, the remaining 31 samples were transported to China, with approval from the Sierra Leone Ministry of Health, for re-sequencing on the Illumina platform. DNA was fragmented into 500 bp segments using the Covaris S2 system. Libraries were prepared and sequenced on an Illumina X-10 platform, generating an

average of 136 million reads per sample (range: 31–329 million). From this second sequencing round, 16 additional samples met the desired sequencing coverage threshold. In total, 35 samples with sufficient coverage were included in the final dataset.

The overall quality of sequencing data from the MGISEQ-200 platform was observed to be lower than that from the Illumina X-10 platform. To ensure uniform quality for downstream population genetic analyses, we applied consistent preprocessing across both datasets using Trimmomatic software. Specifically, we removed the first 15 bp from each read, irrespective of the reported base quality scores, to minimize platform-induced biases and improve data comparability.

Population structure, genetics analysis, and selection tests

We assessed population genetic using both the global dataset and the imported Sierra Leone isolates. For this analysis, we downloaded the publicly available *pf3k* project dataset [11], which comprises samples from 13 countries and nearly 3,000 isolates. From this dataset, we randomly selected 30 samples per country for comparative analysis. The complete list of countries used for this selection is provided in the supplementary table (Table S1). Due to variations in sequencing protocols, platforms, and laboratory conditions, it was necessary to extract a subset of SNPs common across all datasets. A total of 31,000 loci shared between our data and the reference datasets were selected for further analysis.

Principal Component Analysis (PCA) was conducted using the *pcadapt* R-package [12], while STRUCTURE analysis was used to assess genetic clustering based on allele ancestry [13]. The admixture model for values K as 3 to 12 was run, and the optimum number of clusters was determined using the delta K method [14]. STRUCTURE was run with 10,000 burn-in steps and 100,000 Markov Chain Monte Carlo (MCMC) iterations. SNPs with a minor allele frequency (MAF) below 5% were excluded from the analysis. To investigate the phylogenetic relationships among isolates, a maximum likelihood (ML) phylogenetic tree was constructed using RAxML-NG v0.8 [15]. The analysis employed the GTR + G substitution model, which accounts for reversible nucleotide substitution rates and site-specific rate heterogeneity. High-quality SNPs identified in the dataset were used as input for the analysis. Bootstrapping (1000 replicates) was performed to assess the robustness of the inferred tree topology, and the resulting tree was visualized using FigTree v1.4.4.

Pairwise IBD was checked using hmmIBD [16] with default parameters and auto-estimated allele frequencies. Within-sample genomic diversity, relative to the overall population, was determined using the F_{WS} fixation index in the bahlolab/moimix R-package [17]. Samples with F_{WS} approaching 1.0 were considered monoclonal. $F_{WS} < 0.95$ indicated mixed infections. Genotypic errors were evaluated by checking heterozygosity calls; loci with multiple segregating alleles were refined by retaining the major allele and discarding the minority allele to avoid errors due to genetic complexity or hypnozoite-induced relapses.

For high-quality SNPs, we estimated nucleotide diversity ($\widehat{\pi}$), Watterson's estimator ($\widehat{\theta}_{\omega}$), genetic differentiation (F_{ST}), and Tajima's D value for all genes using ARLEQUIN Ver3.5 [18]. Tajima's D tests were performed to assess deviations from neutral evolution and identify genes under selection.

Selection and haplotype analyses

To detect signals of recent or ongoing positive selection, we used integrated haplotype score (iHS) and cross-population extended haplotype homozygosity (XP-EHH) analyses implemented in Selscan-Ver1.10a [19]. iHS was calculated for each SNP by tracking the decay of haplotype homozygosity for both ancestral and derived haplotypes extending from each SNP. iHS values were standardized in frequency bins across the entire genome, identifying loci subject to moderate-frequency selective sweeps.

XP-EHH was used to compare site-specific extended haplotype homozygosity between the Sierra Leone isolates and reference populations. Initially, we aimed to compare SL with West Africa to highlight unique selection signals in Sierra Leone. However, this comparison did not yield significant results or identify key genes, likely due to the genetic similarity and shared selective pressures between Sierra Leone and other West African populations. To address this, we compared Sierra Leone with Asian populations, which provided a clearer contrast due to the distinct selective environments in these regions. This approach allowed us to identify loci under differential selection pressures between geographically distant populations, emphasizing the unique evolutionary dynamics in Sierra Leone.

Results

We successfully sequenced 35 *P. falciparum* samples from Sierra Leone, achieving a comprehensive dataset with substantial sequencing depth. Each sample generated a large number of reads, with a total ranging from approximately 27 million to 589 million reads, maintaining an average mapped ratio above 9X and 1X coverage exceeding 90% for most samples (Table S2). This robust sequencing effort revealed approximately 376,450 high-quality SNPs across the samples, indicating extensive genetic variability and potential regions of high evolutionary pressure or selection. The distribution of SNP numbers across the genome highlighted significant

variability, with some regions exhibiting dense clusters of SNPs, suggestive of potential hotspots for genetic diversity or selective pressure. Compared the total SNP numbers across several countries (Fig. 1A), we found a wide distribution of SNP numbers in Sierra Leone, with some samples showing extremely high values, significantly higher than any other country. This suggests a high genetic diversity or possible sequencing discrepancies that might be unique to the samples from this region. Meanwhile we calculated nucleotide diversity (π) for 5,600 genes in three groups: Asia, West Africa, and Sierra Leone (Fig. 1B). The mean π values for these groups were 0.0015, 0.0016, and 0.0022, respectively, while the median values were 0.0007, 0.0008, and 0.0006. The slightly higher mean π value in Sierra Leone suggests greater genetic diversity, potentially reflecting local transmission dynamics and historical selective pressures.

The analysis of F_{WS} values across 34 Sierra Leone isolates revealed a range of 0.583 to 0.99, with the majority falling below 0.95, consistent with high-transmission settings. This aligns with the high prevalence of complex infections observed in the region. For comparison, similar F_{WS} patterns have been reported in neighboring West African countries such as Guinea, Mali, and Côte d'Ivoire, as demonstrated in Amambua-Ngwa et al. [20]. These findings underscore the genetic complexity and high transmission intensity of *P. falciparum* in Sierra Leone.

PCA and NJ-tree constructions elucidated the genetic relationships within the Sierra Leone samples as well as between these samples and reference populations from West Africa and Southeast Asia. The PCA depicted a distinct clustering of Sierra Leone samples, highlighting their unique genetic backgrounds when juxtaposed with other regions (Fig. 2A, B). Samples from West Africa displayed a broader spread across the PCA plot, indicating a higher genetic diversity which might reflect the diverse genotypes circulating due to intense malaria transmission and subsequent recombination events within this region. In the ML-tree (Fig. 2C), Asian samples formed a tight cluster, reflecting a lower genetic diversity which could suggest recent population bottlenecks or limited lineage expansions within this region. In contrast, African samples exhibited more phylogenetic diversity with multiple sub-clusters, consistent with the continent's status as a high-burden malaria region with extensive genetic variability among circulating strains.

STRUCTURE analysis, optimized at K = 5, further delineated the genetic ancestry of the Sierra Leone samples, revealing five principal genetic ancestries distinct from those observed in the reference populations (Fig. 2D). This pattern underscores the potential unique genetic adaptations of the Sierra Leone malaria parasite populations, which might be shaped by local transmission dynamics and environmental factors, such as the seasonality of malaria transmission and interactions with diverse host populations. Further investigation on different K value revealed that STRUCTURE could reliably differentiate Asian and African populations, but it struggled to resolve finer distinctions within African populations

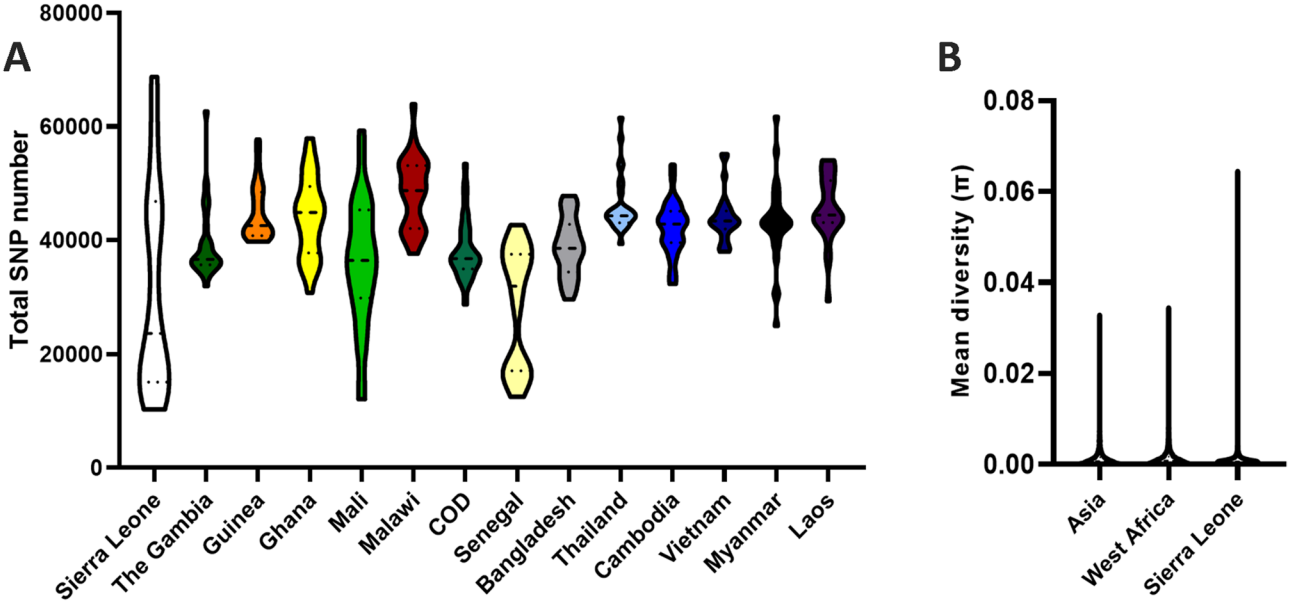

Fig. 1 Comparison of SNP Abundance and Nucleotide Diversity (π) Across Populations in Africa and Asia. **A** Total SNP counts for individual countries across Africa and Asia, showing variation in genetic diversity across regions. **B** Mean nucleotide diversity (π) for 5,600 genes in three population groups: Asia, West Africa, and Sierra Leone. The Sierra Leone group demonstrates higher mean nucleotide diversity, consistent with complex population dynamics and local selection pressures

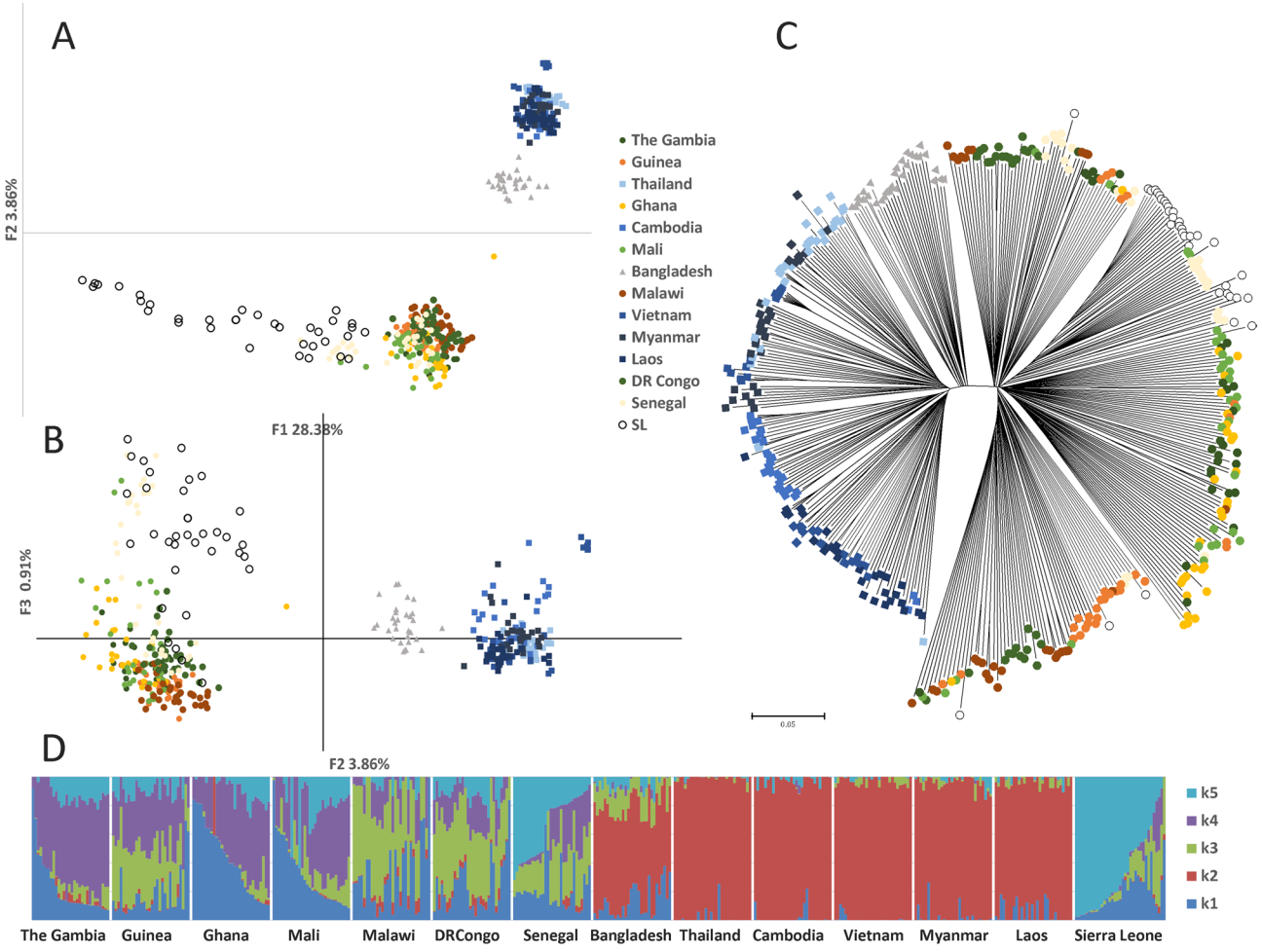

Fig. 2 Structure and Phylogenetic Relationships of *P. falciparum* Samples from Sierra Leone and Reference Populations. **A** PCA of Sierra Leone and reference populations from West Africa and Southeast Asia, based on F1 and F2 axes, showing clear separation between African and Asian populations. Sierra Leone samples (open circles) cluster separately, indicating a distinct genetic profile. **B** PCA plot highlighting F2 and F3 axes, further distinguishing Sierra Leone samples from reference. **C** ML-tree based on genome-wide SNP data, illustrating the genetic diversity across samples. Sierra Leone samples (open circles) form a separate cluster but show higher genetic diversity. **D** STRUCTURE analysis showing genetic ancestry inferred from K= 5. Sierra Leone samples exhibit unique genetic ancestry components, distinct from reference. The West African populations show greater genetic diversity with more mixed ancestry, while Southeast Asian populations cluster into fewer distinct genetic components

(Figure S1). The genetic structuring observed suggests that the Sierra Leone samples possess unique genetic configurations, setting them apart from other West African regions, reflecting localized evolutionary histories and adaptations.

IBD analysis among the Sierra Leone samples revealed extremely low relatedness, with a median IBD value of 0.021. This suggests high genetic heterogeneity and indicates that most infections are independently acquired. The vast majority of the sample pairs (74.96%) have an IBD value of 0.1, indicating that only a minimal proportion of their genomes are identical by descent (Fig. 3A). This suggests that while these individuals are infected by the same species, the infections are likely from diverse sources or different transmission chains. The highest IBD value observed is over 0.5, occurring in only 0.17% of sample pairs, which indicates an extremely rare instance of high genetic similarity, potentially suggesting a close transmission link or less diversity in those particular cases. The analysis of conserved genomic segments within these samples confirmed the minimal shared genetic material, highlighting the independent nature of malaria infections within this region (Fig. 3B). With 35 samples forming 595 pairs, around 35,000 highly similar segments were identified involving 2,900 genes. Notably, only 188 genes had segments appearing more than 50 times, including 70 *var* genes, 65 *rif* genes, and 15 *stevor* genes, suggesting that these VSA genes contain numerous conserved regions, which may be linked to their roles in immune evasion and other vital functions [21].

The analysis of evolutionary pressures using Tajima's D test across 5,600 genes of *P. falciparum* isolates from

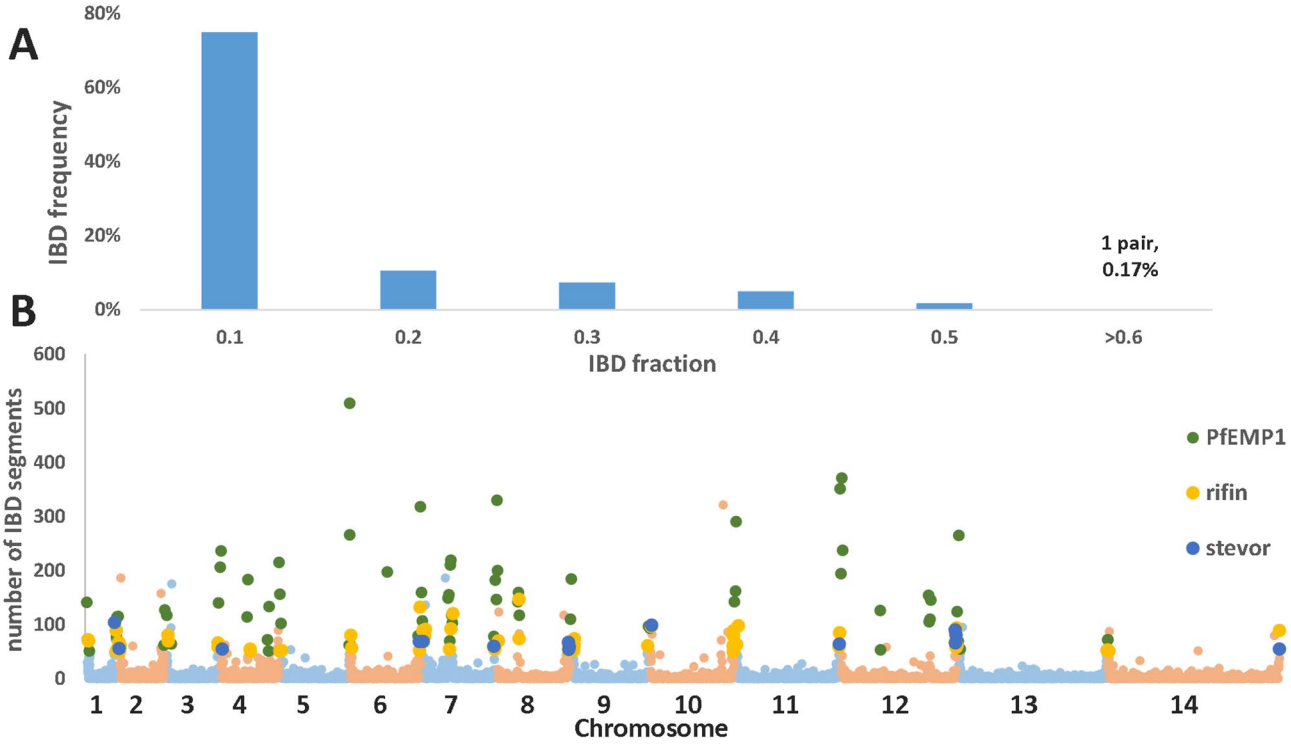

Fig. 3 Analysis of Identity-by-Descent (IBD) in *P. falciparum* Samples from Sierra Leone. **A** Distribution of IBD fraction across 595 sample pairs from Sierra Leone. The vast majority of sample pairs have an IBD fraction below 0.1, indicating low relatedness between infections, with only 1 pair (0.17%) exceeding 0.5. This suggests that most infections are independently acquired, reflecting high transmission and low genetic relatedness among the parasite population in Sierra Leone. **B** Distribution of highly conserved IBD segments across the 14 chromosomes, showing the number of IBD segments identified in genes associated with immune evasion, including *var*, *rifin*, and *stevor*. Notably, only 188 genes contained highly conserved segments that appeared more than 50 times, primarily concentrated in genes related to immune evasion, indicating the functional importance of these gene families in the population

Sierra Leone reveals predominantly negative values (Table S3), particularly within drug resistance loci, suggesting a significant influence of directional selection (Fig. 4A). We compared the distribution of Tajima's D values from isolates in Sierra Leone and West Africa (Fig. 4B). We have listed potential drug resistance genes identified based on prior studies and reviews on drug resistance compound screening (Table S6) [22–26], and these genes tend to show more negative Tajima's D values in Sierra Leone samples. This difference suggests that drug resistance genes may be under more intense selection pressure due to drug treatment regimes, which often lead to the rapid emergence of resistant strains [27]. In contrast, genes associated with invasion appear to be more evenly distributed across the spectrum of Tajima's D values. Further analysis with iHS identified significant recent positive selection on several SNPs across the genome (Fig. 4C, Table S4). Notably, high iHS values were observed in SNPs within the *pfcrt* gene, linked to chloroquine resistance, reflecting persistent selective pressures that may correspond to historical treatment strategies [28]. Comparative rsb analysis between Sierra Leone and Asian populations delineated differential selection pressures, with Sierra Leone samples showing widespread selection across the genome, affecting genes involved in immune evasion such as the *var*, *rifin*, and *stevor* families, and those associated with drug resistance (Fig. 4D, Table S5). These findings suggest ongoing evolution driven by both immune response and treatment practices. In contrast, Asian samples exhibited selection signals predominantly at the telomeric regions of chromosomes, enriching for VSA genes, which could indicate localized adaptation.

Our detailed investigation into the genetic basis of drug resistance in Sierra Leone malaria strains revealed distinct patterns of evolution in key mutations [29], underscored by negative Tajima's D values (Table 1). For example, the K76 T mutation in *pfcrt*, associated with chloroquine resistance [30], continues to show signs of positive selection (iHS of 1.68). Mutations such as N86Y and Y184 F in *pfmdr1* also showed significant iHS values [31], while sulfadoxine-pyrimethamine (SP) resistance-associated mutations such as S108 N in *pfdhfr* and A437G in *pfdhps* were prevalent [32], indicating persistent resistance in these genes. In our analysis of the *kelch13* gene, no validated markers of ACT resistance, such as C580Y, R561H, R622I, C469Y, or A675 V, were detected in our dataset. However, two non-synonymous

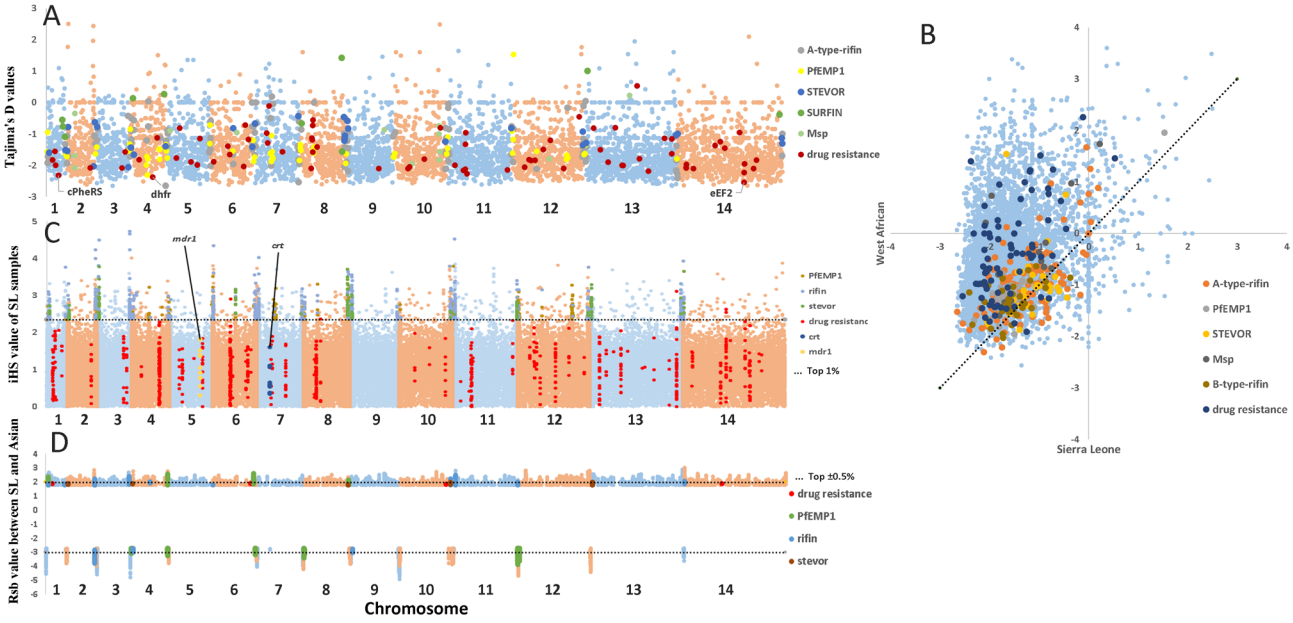

Fig. 4 Selection Pressure Analysis of *P. falciparum* Genes from Sierra Leone Isolates. **A** Tajima's D values calculated across 5,600 genes from isolates in Sierra Leone, highlighting directional selection on several key loci, including drug resistance genes such as *dhfr* and *cPheRS*, and immune evasion genes (*var*, *rifin*, *stevor*). Negative Tajima's D values across many genes suggest the influence of directional selection. **B** Comparative analysis of Tajima's D values between Sierra Leone and West African isolates. While most genes cluster around neutral evolution, drug resistance genes and some surface antigen families show stronger selection pressures in Sierra Leone. **C** iHS analysis of Sierra Leone samples identifies strong selection signals, particularly in immune evasion gene families and drug resistance loci. The higher iHS scores reflect recent positive selection on these loci. **D** Rsb analysis between Sierra Leone and Asian populations showed distinct selection patterns between regions, with the Sierra Leone population exhibiting widespread selection signals across genes related to immune evasion and drug resistance

Table 1 The evaluation of key mutations in drug resistance genes highlighted distinct patterns of evolution

Gene	Description	Mutation	Ratio(%)	Tajima's D	Highest |iHS|	Mutation associated resistance
pfcrt	chloroquine resistance transporter	C72S	0	−0.97438	1.68131	chloroquine
		V73S	0			chloroquine
		M74I	0			chloroquine
		N75E	0			chloroquine
		K76 T	23			key marker for chloroquine resistance
pfmdr1	multidrug resistance protein 1	N86Y	3	−1.98965	1.65961	chloroquine and mefloquine
		Y184 F	49			altered susceptibility to multiple antimalarials
		D1246Y	0			altered drug susceptibility
pfdhfr	dihydrofolate reductase	N51I	0	−2.37805	'-	sulfadoxine-pyrimethamine
		C59R	0			sulfadoxine-pyrimethamine
		I164L	0			sulfadoxine-pyrimethamine
		S108 N	86			sulfadoxine-pyrimethamine
pfdhps	dihydropteroate synthase	S436 A	26	−1.41477	1.92893	sulfadoxine
		A437G	86			sulfadoxine
		K540E	6			sulfadoxine, combined with dhfr mutations
		A581G	0			sulfadoxine
		A613S	20			sulfadoxine
k13	kelch13 gene	C580Y	0	−1.79161	0.741735	artemisinin
		R561H	0			
		R622I	0			
		C469Y	0			
		A675 V	0			
		K189 T	40			
		Y493H	9			

mutations, K189 T (40%) and Y493H (9%), were identified. While these mutations are not currently associated with resistance phenotypes, their high prevalence warrants further investigation into their potential functional impacts.

Discussion

Even without the impact of COVID-19 on the public health system, Sierra Leone remains an area of high transmission of *P. falciparum*, and persistent malaria infection provides ideal conditions for the evolutionary adaptation of the parasite. Based on our data, we propose that the heavy burden of malaria on the local population, combined with repeated infections and high transmission rates, may have resulted in a genetic landscape shaped by both immune pressure and drug treatment. Compared to previous sequencing reports, this more comprehensive genomic study could accurately reveal the local adaptation of *P. falciparum* and demonstrate the characteristic of the Sierra Leone population based on methods such as IBD or selection pressure testing.

We propose that the high malaria prevalence in Sierra Leone is closely associated with a dynamic and diverse genetic landscape, which may be caused by frequent reinfections and intensive transmission rates [33]. The low F_{WS} values observed in our samples suggest a complex infection with multiple parasite genotypes. This may point to frequent reinfection events rather than chronic or monoclonal infections [34]. In addition, the low IBD values between samples, with nearly 75% of sample pairs being uncorrelated, suggest that these infections are mostly independent and may have originated from different transmission events. This high degree of genetic diversity, coupled with the elevated SNP numbers compared to other regions, highlights the extensive transmission in Sierra Leone, where the introduction of new parasite genotypes through superinfection is common. As showed by previous studies [35–37], we took a high prevalence in Sierra Leone as a result of the frequent introduction of genetically distinct strains, driven by intense local transmission dynamics.

Previous studies have suggested that genes associated with immune evasion and drug resistance in *P. falciparum* are subject to directional selection [38, 39], and our data supports this notion. Genes involved in immune evasion, such as those in the *rifin*, *var*, and *stevor* families, exhibit clear signs of strong selective pressure in our analysis, as evidenced by negative Tajima's D values and significant iHS scores. The iHS values further confirm that specific alleles in these immune evasion gene families are under strong positive selection, suggesting that these genes play a crucial role in the parasite's ability to persist despite immune challenges. Moreover, our XP-EHH analysis between Sierra Leone and other regions, particularly Southeast Asia, highlights distinct evolutionary pressures, with immune evasion genes being under stronger selection in Sierra Leone. This finding contrasts with drug resistance genes, which, while still under selection, exhibited less extreme Tajima's D values, suggesting that drug pressure, though present, may not be the primary driver of selection in this population. Some studies have suggested that immune evasion mechanisms are the dominant force shaping the parasite's genetic landscape [40, 41], and our data stands with this view.

The analysis of drug resistance genes, such as pfcrt and pfmdr1, revealed distinctly negative Tajima's D values and high iHS scores, consistent with strong directional selection. However, it is important to note that most genes in the genome also exhibited negative Tajima's D values and significant iHS and rsb signals, which could reflect broader demographic processes, such as population expansion or bottlenecks, rather than selection specific to resistance genes. This emphasizes the need for careful interpretation of selection signals. For drug resistance genes like pfcrt, these patterns may not solely reflect ongoing drug use but could also result from historical drug pressures. These findings highlight the enduring impact of both historical and current treatment practices on the evolution of P. falciparum populations. While immune evasion genes appear to be the dominant force driving selection, the persistent selection pressures on drug resistance loci underscore their continued relevance in shaping the genetic landscape of Sierra Leone populations.

Our data validates the presence of key drug resistance mutations previously reported in this region [42], including K76 T in *pfcrt*, N86Y and Y184 F in *pfmdr1*, and S108 N and A437G in *pfdhfr* and *pfdhps*, all of which show strong signs of directional selection with negative Tajima's D values. The persistence of mutations like K76 T in *pfcrt*, despite the dramatic reduction and abandonment of chloroquine use across Africa, suggests potential cross-resistance with newer therapies and highlights the residual effects of past drug use. Notably, the high iHS values for K76 T in *pfcrt* (1.68) and Y184 F in *pfmdr1* further indicate ongoing positive selection due to continued reliance on artemisinin-based combination therapies (ACTs) [43, 44]. Similarly, the mutations S108 N in *pfdhfr* and A437G in *pfdhps*, both exhibiting high prevalence and significant selection pressure, underscore the entrenched nature of SP resistance in the region. The lack of the C580Y mutation in *kelch13*, associated with artemisinin resistance, suggests that artemisinin resistance is not yet a significant issue in Sierra Leone, consistent with reports indicating low access and adherence to ACTs.

However, the continued presence of older resistance alleles, even as drug policies evolve, reflects the enduring impact of historical drug use. Based on the data we have

found and our work in local hospitals, we believe that limited access to new antimalarial treatments in Sierra Leone increases the likelihood that people will rely on older drugs, thereby maintaining selection pressure on these alleles [45, 46]. This highlights the importance of continuing to track the adaptive dynamics of drug resistance in this high-transmission area.

Given these results, we also recognize the urgent need to continue genomic surveillance to understand the changing landscape of *P. falciparum* in Sierra Leone. Although we see that high transmission rates, immune evasion, and drug resistance will continue to shape the genetic characteristics of the parasite, local health authorities should adopt adaptive strategies to respond to the continued threat of malaria.

Conclusion

As previously reported, our study demonstrates the role of transmission dynamics, immune evasion, and drug resistance in shaping the genetic diversity of *P. falciparum* in Sierra Leone. The genome-wide data provided a clearer picture of the genetic landscape of local parasite. High genetic diversity and low IBD values highlight widespread transmission and frequent reinfection. Immune evasion genes, including *var*, *rifin*, and *stevor*, continue to face strong directional selection, indicating ongoing adaptation to host immune responses. The mutation frequencies in drug resistance genes, such as *pfcrt* and *pfdhfr*, are consistent with previous studies and also demonstrate unique diversity in the Sierra Leone population. As with the data we reported in China [47, 48], we once again emphasize the need for continued genetic surveillance to better control malaria and address the ongoing threat of drug resistance.

Abbreviations

ACTs	Artemisinin-based combination therapies
DRC	the Democratic Republic of the Congo
DNB	DNA nanoball
FST	fixation index
FWS	Within-host infection fixation index
IBD	Identical by descent
iHS	Integrated haplotype score
MAF	Minor allele frequency
NJ	Neighbour-Joining trees
PCA	Principal component analysis
SNP	Single nucleotide polymorphism
SP	Sulfadoxine-pyrimethamine
VSAs	Variant surface antigens
XP-EHH	Cross-population extended haplotype homozygosity
$\widehat{\theta}_{\omega}$	Watterson's estimator
$\widehat{\pi}$	Nucleotide diversity

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11771-y.

Supplementary Material 1. Figure S1. STRUCTURE bar plots for K = 4 to K = 8 showing population genetic clustering across 14 countries. Each vertical bar represents one individual, and colors indicate proportional membership in each inferred ancestral population (K). The populations are grouped by country, including West African (e.g., The Gambia, Guinea, Ghana, Mali, Malawi, COD, Senegal, Sierra Leone) and Asian regions (e.g., Thailand, Vietnam, Myanmar, Laos, Bangladesh, Cambodia). STRUCTURE analysis at increasing K values reveals clear separation between African and Asian populations, while finer substructure within African populations, including Sierra Leone, remains difficult to resolve. Despite this, Sierra Leone samples show a distinct genetic profile compared to other West African populations, suggesting localized genetic adaptation and transmission history.

Supplementary Material 2. Table S1. The complete list of pf3k reference countries used for population analysis. Table S2. Sequencing and mapping summary statistics for 35 Sierra Leone samples. Table S3. 5,572 *P. falciparum* genes with population genetic statistics (nucleotide diversity, Watterson's estimator, Tajima's D and FST test). Table S4. Genes with top 1% integrated haplotype score. Table S5. Genes with top 1% (±0.5%) cross population extended haplotype homozygosity value in Sierra Leone/ Asian populations. Table S6. List of Potential Drug Resistance Genes Identified from Literature Screening and Compound Resistance Studies.

Acknowledgements

We thank the staff of the 23rd batch of China Medical Team in Sierra Leone for collecting blood samples from P. falciparum infected individuals. We would also like to express our gratitude to the members of the China-Sierra Leone Friendship Hospital (Freetown, Sierra Leone) for their invaluable assistance during the sampling process.

Authors' contributions

Conceived and designed the experiments: HMS JHC. Performed the experiments: SBC TYW YWD. Analyzed the data: HMS. Contributed the reagents/materials/analysis tools: XJZ JJX HMS. Wrote the paper: XJZ JJX HMS.

Funding

This work was financially supported in part by the Three-Year Initiative Plan for Strengthening Public Health System Construction in Shanghai (2023–2025) Key Discipline Project (No. GWVI-11.1-12), the National Sharing Service Platform for Parasite Resources (Grant No. TDRC-2019-194-30), and the Foundation of National Science and Technology Major Program (Grant no. 2012ZX10004-220, 2008ZX10004-011, and 2009ZX10004-302). The funding bodies had no role in the design of the study, collection, analysis, and interpretation of data, or in writing of the manuscript.

Data availability

Generated Statement: All data supporting these findings is contained within the manuscript and supplementary tables. All MGISEQ and Illumina raw sequencing reads have been submitted to the NCBI Short Read Archive (BioProject no. PRJNA1159984).

Declarations

Competing interests

The authors declare no competing interests.

Author details

[1]Department of laboratory, People's Hospital of Ningxiang City, Changsha, Hunan Province, China
[2]The 23rd batch of China Medical Team in Sierra Leone, Jui, Freetown, Sierra Leone
[3]National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research), Shanghai, People's Republic of China
[4]National Health Commission of the People's Republic of China (NHC) Key Laboratory of Parasite and Vector Biology, Shanghai, People's Republic of China
[5]World Health Organization (WHO) Collaborating Center for Tropical Diseases, Shanghai, People's Republic of China
[6]National Center for International Research on Tropical Diseases, Shanghai, People's Republic of China

[7]Sierra Leone-China Friendship Biological Safety Laboratory, Jui, Freetown, Sierra Leone
[8]Hainan Tropical Diseases Research Center (Hainan Sub-Center, Chinese Center for Tropical Diseases Research), Haikou 571199, China
[9]The Affiliated Nanhua Hospital, Department of Infectious Diseases, Hengyang Medical School, University of South China, Hengyang, Hunan, China

Received: 24 September 2024 / Accepted: 2 June 2025
Published online: 01 July 2025

References

1. World Health Organization. World malaria report 2023. World Health Organization; 2023.
2. WHO guidelines for malaria [internet]. Geneva: World Health Organization; 2021. PMID: 36580567.
3. Winzeler EA. Malaria research in the post-genomic era. Nature. 2008;455(7214):751–6.
4. MalariaGEN; Abdel Hamid MM, et al. Pf7: an open dataset of *plasmodium falciparum* genome variation in 20,000 worldwide samples. Wellcome Open Res. 2023;8:22. https://doi.org/10.12688/wellcomeopenres.18681.1.
5. Manske M, et al. Analysis of plasmodium falciparum diversity in natural infections by deep sequencing. Nature. 2012;487(7407):375–9.
6. Neafsey DE, Taylor AR, MacInnis BL. Advances and opportunities in malaria population genomics. Nat Rev Genet. 2021;22(8):502–17.
7. Wang T, et al. Genome analysis of plasmodium falciparum: A preliminary Observation—Sierra leone, 2022–2023. China CDC Wkly. 2024;6(17):368.
8. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
9. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.
10. McKenna A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
11. Project MPfC. Genomic epidemiology of Artemisinin resistant malaria. Elife. 2016;5:e08714.
12. Luu K, Bazin E, Blum MG. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 2017;17(1):67–77.
13. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–20.
14. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61.
15. Kozlov AM, et al. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453–5.
16. Schaffner SF, et al. HmmIBD: software to infer pairwise identity by descent between haploid genotypes. Malar J. 2018;17(1):196.
17. Lee S, Bahlo M. moimix: an R package for assessing clonality in high-throughput sequencing data. Moimix R Package Assess. Clonality High-Throughput Seq. Data. 2016.
18. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. Mol Ecol Resour. 2010;10(3):564–7.
19. Szpiech ZA, Hernandez RD. Selscan: an efficient multi-threaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31(10):2824–7. p. msu211. https://doi.org/10.1093/molbev/msu211.
20. Amambua-Ngwa A, et al. Major subpopulations of plasmodium falciparum in sub-Saharan Africa. Science. 2019;365(6455):813–6.
21. Dzikowski R, Templeton TJ, Deitsch K. Variant antigen gene expression in malaria. Cell Microbiol. 2006;8(9):1371–81.
22. Ehrlich HY, Jones J, Parikh S. Molecular surveillance of antimalarial partner drug resistance in sub-Saharan africa: a spatial-temporal evidence mapping study. Lancet Microbe. 2020;1(5):e209–17.
23. Tumwebaze PK, et al. Decreased susceptibility of plasmodium falciparum to both Dihydroartemisinin and lumefantrine in Northern Uganda. Nat Commun. 2022;13(1):6353.
24. Miotto O, et al. Genetic architecture of artemisinin-resistant plasmodium falciparum. Nat Genet. 2015;47(3):226–34.
25. Horn D, Duraisingh MT. Antiparasitic chemotherapy: from genomes to mechanisms. Annu Rev Pharmacol Toxicol. 2014;54(1):71–94.
26. Blasco B, Leroy D, Fidock DA. Antimalarial drug resistance: linking plasmodium falciparum parasite biology to the clinic. Nat Med. 2017;23(8):917–28.
27. Amelo W, Makonnen E. Efforts made to eliminate Drug-Resistant malaria and its challenges. Biomed Res Int. 2021;2021(1):5539544.
28. Roux AT, et al. Chloroquine and sulfadoxine–pyrimethamine resistance in Sub-Saharan Africa—A review. Front Genet. 2021;12:668574.
29. Hodoameda P, Duah-Quashie NO, Quashie NB. Assessing the roles of molecular markers of antimalarial drug resistance and the host pharmacogenetics in drug-Resistant malaria. J Trop Med. 2022;2022(1):3492696.
30. Setthaudom C, et al. Role of plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes on in vitro chloroquine resistance in isolates of plasmodium falciparum from Thailand. Am J Trop Med Hyg. 2011;85(4):606.
31. Shafik SH, et al. Mechanistic basis for multidrug resistance and collateral drug sensitivity conferred to the malaria parasite by polymorphisms in PfMDR1 and PfCRT. PLoS Biol. 2022;20(5):e3001616.
32. Chaturvedi R, et al. Geographical spread and structural basis of sulfadoxine-pyrimethamine drug-resistant malaria parasites. Int J Parasitol. 2021;51(7):505–25.
33. Castelli F, Tomasoni LR. New insights on malaria. New Microbiol. 2022;45:83–98.
34. Mobegi VA, 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(6):1490–9.
35. Wu SL, et al. Spatial dynamics of malaria transmission. PLoS Comput Biol. 2023;19(6):e1010684.
36. Shi B, et al. Inference and prediction of malaria transmission dynamics using time series data. Infect Dis Poverty. 2020;9(04):84–96.
37. Roh ME, et al. High genetic diversity of plasmodium falciparum in the low-transmission setting of the Kingdom of Eswatini. J Infect Dis. 2019;220(8):1346–54.
38. Kaczanowski S. Detection of positive selection acting on protein surfaces at the whole-genome scale in the human malaria parasite plasmodium falciparum. Infect Genet Evol. 2023;107:105397.
39. Kassegne K, et al. Genome-wide analysis of the malaria parasite plasmodium falciparum isolates from Togo reveals selective signals in immune selection-related antigen genes. Front Immunol. 2020;11:552698.
40. Whitlock AO, Juliano JJ, Mideo N. Immune selection suppresses the emergence of drug resistance in malaria parasites but facilitates its spread. PLoS Comput Biol. 2021;17(7):e1008577.
41. Adejoh J, et al. Molecular and epigenetics mechanisms for the immune control of plasmodium parasites infection: a comprehensive review. Asian J Biol Sci. 2023;16(4):636–69.
42. Leski TA, et al. Prevalence of malaria resistance-associated mutations in plasmodium falciparum Circulating in 2017–2018, bo, Sierra Leone. Front Microbiol. 2022;13:1059695.
43. Banek K, et al. Factors associated with access and adherence to artemisinin-based combination therapy (ACT) for children under five: a secondary analysis of a National survey in Sierra Leone. Malar J. 2021;20:1–13.
44. Banek K, et al. Exploring barriers and facilitators of adherence to artemisinin-based combination therapies for the treatment of uncomplicated malaria in children in freetown, Sierra Leone. Healthcare (Basel). 2021;9(9):1233. https://doi.org/10.3390/healthcare9091233.
45. Hossain MS, et al. Examining the disparities of anti-malarial drug consumption among children under the age of five: a study of 5 malaria-endemic countries. Malar J. 2023;22(1):370.
46. Kabba JA, et al. Prescribing for patients seeking maternal and child healthcare in Sierra Leone: a multiregional retrospective cross-sectional assessments of prescribing pattern using WHO drug use indicators. Risk Manag Healthc Policy. 2020;13:2525–34. https://doi.org/10.2147/RMHPS256648.
47. Liu Y, et al. Retrospective analysis of plasmodium Vivax genomes from a pre-elimination China inland population in the 2010s. Front Microbiol. 2023;14:1071689.
48. Shi S-M, et al. Genome-wide scans for Ghanaian plasmodium falciparum genes under selection from local and Chinese host populations. Front Cell Infect Microbiol. 2021;11:630797.

Publisher's Note