

Preplanned Studies

Genome Analysis of *Plasmodium falciparum*: A Preliminary Observation — Sierra Leone, 2022–2023

Tianyu Wang¹; Haimo Shen¹; Bin Xu¹; Wanxuan Yang^{1,2}; Shenbo Chen^{1,#}; Junhu Chen^{1,2,3,4,#}

Summary

What is already known about this topic?

Sierra Leone, with a gross domestic product (GDP) per capita below \$300 and significant poverty, ranks among the world's least developed countries (LDCs). Despite its modest population of 8.6 million, the nation reports approximately 2.6 million malaria cases annually. Previously, there has been no reporting on the malaria genome data from this country.

What is added by this report?

In this study, we present the first reported whole-genome sequence analysis of 19 high parasite-density *Plasmodium falciparum* isolates from Sierra Leone, providing insights into the genomic epidemiology of this high-prevalence area. We found a high degree of relatedness among infections and substantial genetic diversity, consistent with the gradual reduction in overall case numbers. Moreover, our whole-genome analysis revealed that, beyond drug-resistance genes, gene families related to blood cell invasion, immune evasion, and others are undergoing directional selection. This suggests that the population in Sierra Leone has developed a relatively strong acquired immunity.

What are the implications for public health practice?

The genomic data not only facilitate the creation of single nucleotide polymorphism barcodes for case tracking but also enable the analysis of evolving transmission dynamics and selection pressures. Additionally, the samples from Sierra Leone exhibited higher selective pressures on resistance genes compared to those from Asia, a trend not commonly observed in other African samples. This suggests that less stringent healthcare systems and inconsistent treatment strategies can subject parasites to increased drug pressure, thereby accelerating the development of resistant strains.

Numerous databases cataloging the genomic diversity of malaria parasites have been developed recently (1–2). These resources enable comprehensive

analysis of genetic variation within hosts, superinfection, co-transmission, identification and monitoring of drug resistance loci, and understanding of malaria transmission dynamics. Although these databases support advanced analytical methods for examining epidemiological changes and identifying distinct regional epidemiological features, they do not encompass detailed data for every endemic country. In West Africa, Sierra Leone experiences year-round malaria transmission primarily due to *Plasmodium falciparum* (*P. falciparum*). Despite falling short of the 2025 target set by the Global Technical Strategy for Malaria 2016–2030, Sierra Leone recorded a decrease in malaria case incidence in 2022 compared to 2012. Specifically, malaria cases decreased from 2.7 million in 2019 to 2.6 million in 2022 despite the challenges posed by the coronavirus disease 2019 (COVID-19) pandemic (3). However, with a population of 8.6 million, the annual burden of 2.6 million cases remains excessively high. Additionally, until now, there has been no report of malaria genome data from the country, which limits the integration of genomic information with epidemiological and public health data in national malaria control programs and impedes genomic surveillance efforts. In this study, we present the first report of whole-genome sequence variations in 19 high parasite-density *P. falciparum* isolates from Sierra Leone, offering insights into the genomic epidemiology of an area with high prevalence. Our findings open avenues for further studies into the molecular epidemiology of a population characterized by a high rate of superinfection and a low correlation between infections. Additionally, we observed a reduced diversity in drug resistance genes in Sierra Leone samples, suggesting that the country's unstructured antimalarial drug policies have subjected the parasite population to increased directional selection.

Blood samples were collected from Freetown, the capital of Sierra Leone, between December 2022 and March 2023. This study received ethical approval from the Ethics Committee of Chinese Center for Disease

Control and Prevention. Participants were fully informed about the study's procedures, potential risks, and benefits before providing their written informed consent. Samples were taken from hundreds of clinical malaria patients at the Sierra Leone China Friendship Hospital (Freetown), all confirmed to be microscopically positive and polymerase chain reaction (PCR) verified for *P. falciparum* infection.

To ensure sequencing integrity, we selected samples with high parasite density. DNA was extracted from these frozen blood samples using the QIAGEN DNeasy Blood & Tissue Kit (Qiagen, UK), and sequencing libraries were constructed accordingly. These libraries were then sequenced on an MGISEQ-200 machine (BGI, CN), generating an average of 220 million paired-end reads, each 150 base pairs in length (Table 1). The sequenced reads underwent cleaning to remove adapter sequences, and low-quality reads using the Trimmomatic-3.0 tool and were then aligned to the annotated *P. falciparum* 3D7 reference genome, sourced from PlasmoDB.org, using BWA. Following the application of the GATK4 workflows, we refined our results to a set of 165,000 high-quality single nucleotide polymorphisms (SNPs) (4). Our workflow encompassed pre-processing steps such as marking duplicates, indel realignment, and base quality score

recalibration, along with variant discovery using the HaplotypeCaller tools for SNPs and indels calling, and variant filtering based on various quality metrics. SNPs that exhibited more than 5% missing calls in high-quality samples were excluded. Missing calls were defined as positions with fewer than two reads. We utilized a dataset of genotype calls released in 2016, comprising over 3,000 samples from various countries globally. SNP information and allele frequencies were retrieved from the pf3k Project. The complexity of infection within the hosts was assessed by calculating the within-sample F statistic value (FWS), using the bahlolab/moimix R package. FWS values for individual infections ranged from 0.6 to 0.94, with a mean of 0.84 and a median of 0.86. An FWS value below 0.95 indicates the presence of multiple genotypes within an infection. Our analysis demonstrated a comparably lower frequency of monoclonal infections among the samples from Sierra Leone.

We then evaluated the population structure using the pf3k global collection (5) and our 19 high-coverage, single-infection isolates. We estimated nucleotide diversity (π), Watterson's estimator ($\hat{\theta}_w$), genetic differentiation (FST), and Tajima's D value for the 5,600 genes located on chromosomes, employing ARLEQUIN-Ver3.5 (6). Principal component analysis

TABLE 1. Sequencing and mapping summary statistics for 19 samples from Sierra Leone.

Sample ID	RmDup_reads	Mapped ratio (%)	Average coverage depth (X)	Coverage >1x (%)	Coverage >10x (%)	F _{ws}
DL_34	41846007	9.86	9.044	91.21	48.51	0.763
PFGX_162	25515745	1.92	11.272	92.74	21.93	0.819
29104_57-73	31073980	7.77	12.586	90.29	13.93	0.942
29104_58-74	43909234	9.67	4.062	93.27	28.95	0.866
29104_59-75	39891198	10.07	6.010	89.12	8.47	0.755
29104_63-76	32070741	8.62	12.682	94.17	14.78	0.859
7509_81-86	30206296	1.12	10.461	92.43	20.68	0.865
7509_84	50154284	11.02	13.275	90.70	13.2	0.888
Ah_7	58998903	5.10	7.170	95.99	15.39	0.859
D_Pf5_1	36781350	8.58	13.164	91.80	25.82	0.764
sierra131	37367263	26.79	45.902	99.49	96.72	0.602
sierra135	33004273	9.62	6.190	92.92	13.77	0.932
sierra136	44491835	15.65	21.889	98.55	86.06	0.867
sierra146	56873546	7.47	10.314	92.51	39.85	0.808
sierra148	33913056	12.17	11.799	97.47	53.28	0.900
sierra55	30366090	10.97	7.691	96.06	23.62	0.881
sierra57	45635899	9.91	9.566	97.11	34.93	0.892
sierra60	42761031	13.03	10.208	97.11	38.07	0.857
sierra79	37871554	10.25	5.094	90.99	9.99	0.939

(PCA) was performed using the R package. A neighbor-joining (NJ) tree was constructed with MEGA6 software. We reviewed lists of putative resistance genes from previous studies and assessed the selection pressure on these genes in Sierra Leone compared to other reference populations. We also investigated gene families involved in merozoite invasion and modulation of the immune response.

Among the numerous samples collected in Freetown, only 19 sequencing results achieved satisfactory coverage, meeting the required standards to date. This study generated between 176 and 245 million paired-end reads, each with an average length of 150 base pairs. Notably, the samples from Sierra Leone showed a higher frequency of SNPs compared with those from other regions. A PCA of SNP variants effectively revealed the global population structure (Figure 1A), indicating that *P. falciparum* predominantly clusters according to geographic origin, with African samples notably distinct from Asian ones. The primary axes of differentiation grouped the Sierra Leone samples together, aligning them with other African samples, a similarity also evident in the NJ tree

(Figure 1B). Further, we analyzed the chromosome-level structure in the CMB imported isolates using the identity by descent (IBD) fraction, a metric typically used to explore the relatedness among closely related parasite populations (7). The Sierra Leone samples displayed the highest pairwise IBD fractions (mean value=0.04) among pf3k populations, although these values remained relatively low, suggesting a high degree of transmission within the region.

Contrary to the results obtained from PCA and phylogenetic trees, the genetic diversity and balancing selection of genes did not correlate with geographical origin. The genetic diversity in isolates from Sierra Leone ($\pi=0.0028$) exceeded that in the global *P. falciparum* sample (π approximately 1.03×10^{-3}), which includes isolates from diverse regions such as Africa, America, Asia, and Oceania. Gene families such as *rifin*, *var*, *stevor*, *surfin*, and *pfmc* exhibited the highest diversity, consistent with previous studies (8) (Figure 2A). Notably, most of Tajima's D values were negative, averaging -1.02 with a median of -1.16 , suggesting an excess of low frequency polymorphisms indicative of either recent population expansion or

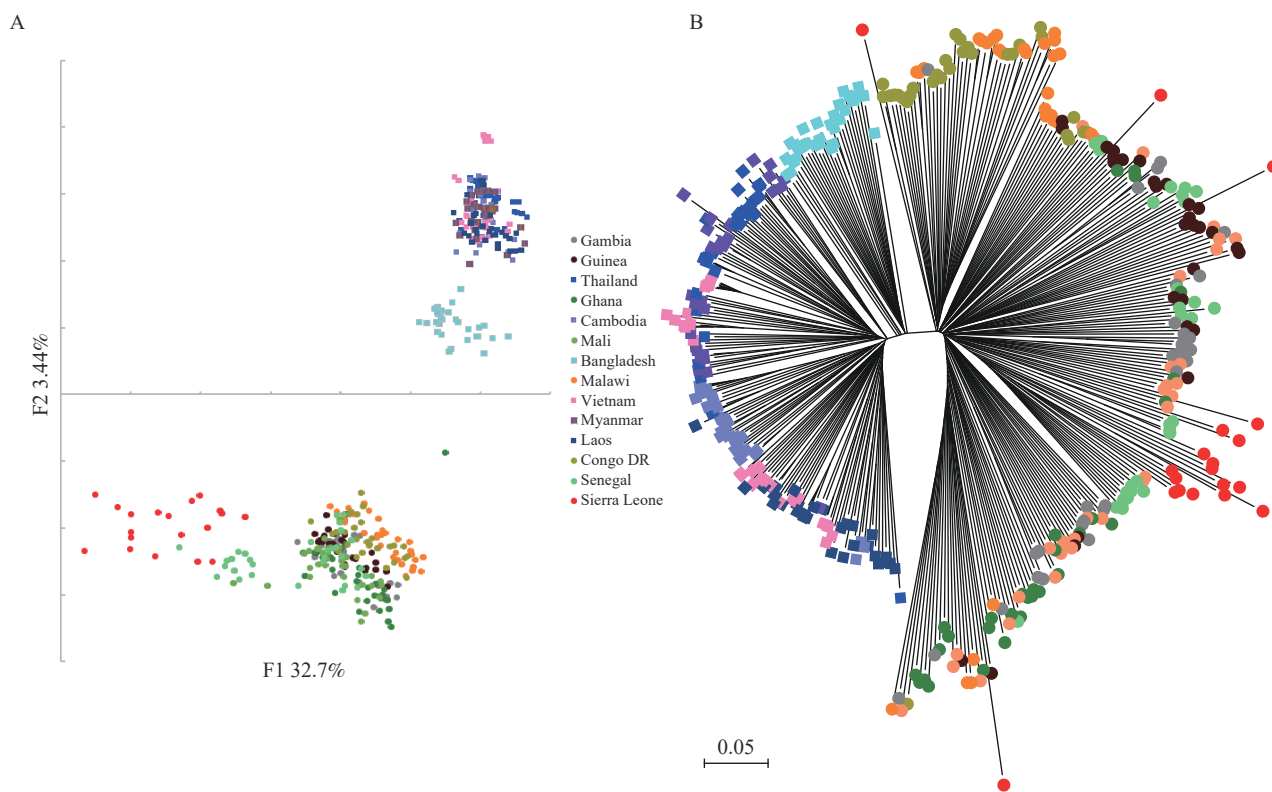


FIGURE 1. Parasite population structure in Sierra Leone samples compared to references (Cambodia, Ghana, Guinea, Mali, Bangladesh, Malawi, Laos, Senegal, DR Congo, Myanmar, Thailand, Vietnam, and the Gambia). (A) PCA plots showcasing genetic differentiation among populations; (B) Neighbor-joining tree depicting the relatedness of Sierra Leone isolates to reference populations, supported by 1,000 bootstraps.

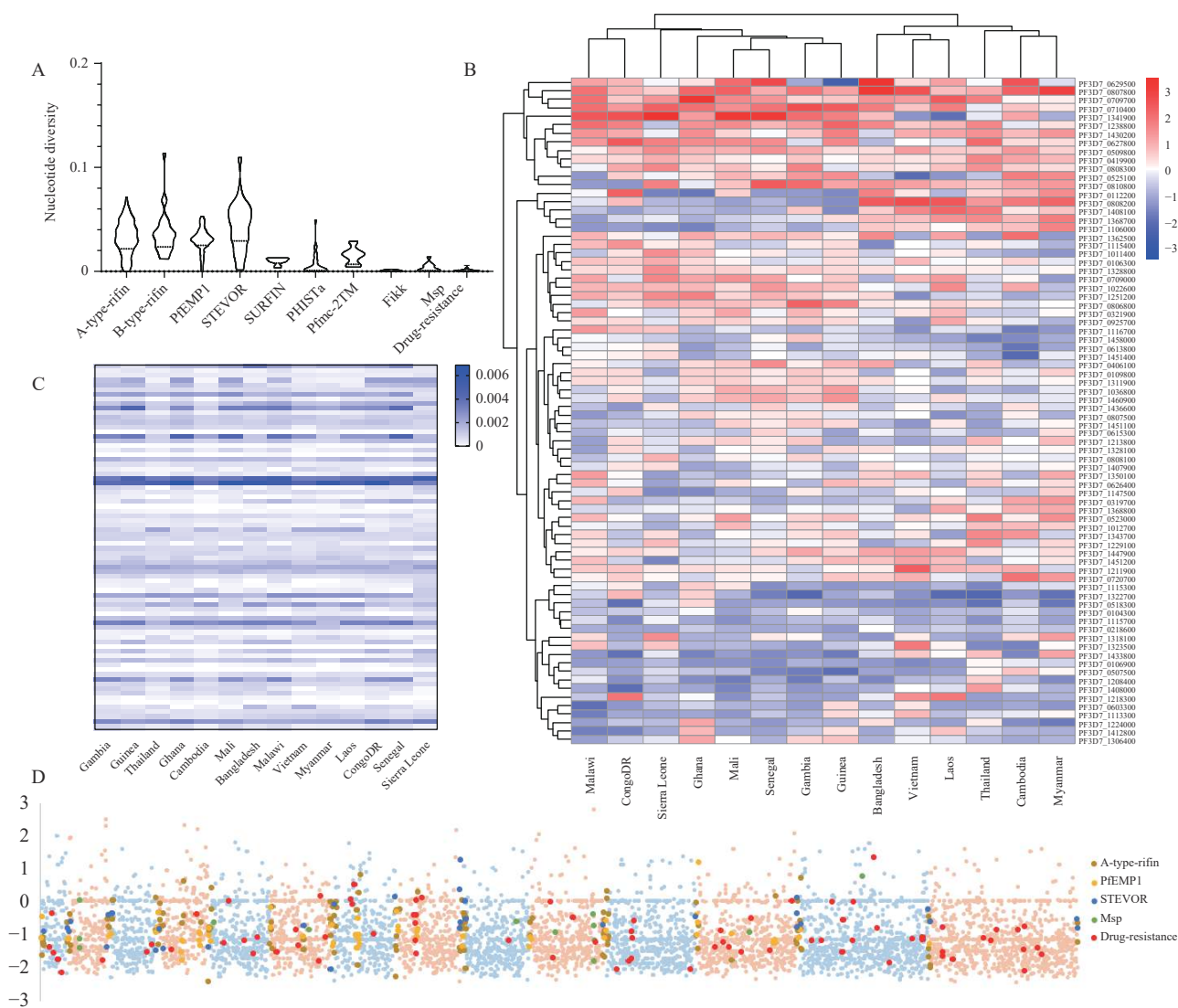


FIGURE 2. Genetic diversity and balanced selection of 5,600 genes in Sierra Leone and reference populations. (A) Relatively lower diversity of resistance genes; (B) Tajima's D value indicated that African samples exhibit clustered types and less selection pressure than Asian samples; (C) Compared to reference populations, Sierra Leone samples exhibit reduced diversity in drug resistance genes; (D) Whole-genome scans reveal that the gene families associated with invasion are also under directional selection.

purifying selection. This heightened diversity, particularly in gene families related to erythrocyte invasion and immune evasion, highlights the parasite's adaptive strategies in facing local environmental pressures and treatment regimes. Nevertheless, only a small fraction of genes, specifically 397 (7%), showed positive Tajima's D values, which are typically associated with reduced diversity due to directional selection. In contrast, Guinea, despite also being a West African population, displayed 1,495 genes (26%) with positive Tajima's D values and a higher mean value (-0.38) but lower genetic diversity ($\pi=0.0016$). These lower negative Tajima's D values are consistent with past findings, indicating recent population

reduction and declining infection rates.

Drawing from previous studies, we compiled a comprehensive list of potential drug-resistance genes (9–10). We anticipated that populations in Sierra Leone would exhibit higher overall selective pressure and lower genetic diversity compared to Asian populations, as indicated by Tajima's D test, particularly in genes resistant to artemisinin-based combination therapies (ACTs) (Figure 2B–C). For antifolate resistance genes, positive balancing selective pressures were observed in the Southeast Asian populations, in contrast to the negative balancing selection found in Africa. This difference could be due to the nearly two-decade-long discontinuation of these

drugs in Southeast Asia. Further analysis revealed distinctive pressure differences between the Sierra Leone population and reference populations. Notably, specific genes, such as the pantothenamide resistance gene (*acs11*) and those associated with imidazole piperazine resistance (e.g., *ugt* and *krs1*), continued to face selection pressure in Sierra Leone but showed positive balancing selection in Asia (11–12). These results underscore the complex and evolving nature of drug resistance patterns across different parasitic populations.

DISCUSSION

The genomic diversity of parasites is instrumental in managing transmission and developing immunity. Investigating the genetic composition of parasite populations is vital to comprehending the epidemiology, dissemination, and behavior within endemic *P. falciparum* communities (13). Additionally, examining the population structure of genes subjected to immune selection provides an in-depth understanding of the relationship between transmission and immunity, which is crucial for vaccine development. Consequently, our study aimed to characterize the basic molecular features of the *falciparum* malaria population in Sierra Leone. We analyzed genomic variations in 19 isolates from Freetown, gathered in early 2023, and compared them with global reference strains from the same period.

Our sample collection encountered specific challenges inherent to high-endemic malaria regions, where inhabitants frequently contract the disease. Repeated malaria infections typically confer a level of protective immunity, resulting in lower parasitic densities in the bloodstream. While this immunity reduces the severity of the illness, it complicates genomic research by producing lower-quality DNA samples. Consequently, of the 57 DNA samples obtained from numerous clinical cases, only 19 met the stringent criteria necessary for high-quality genomic sequencing. What's more, although our results provide valuable insights, they may be influenced by potential confounding factors such as sample selection bias or differences in exposure to antimalarial drugs. These elements could affect the genetic diversity and selection pressures observed. Furthermore, alternative factors, such as genetic drift or localized outbreaks, might also explain the skewed allele frequencies noted. Acknowledging these limitations, including sample size and potential selection bias, is essential as they could

impact our interpretations of genetic diversity and drug resistance.

Based on our analysis of the malaria epidemic in West Africa, as well as insights drawn from extensive public databases over the last decade (14), we anticipate a high degree of genetic relatedness among infections in Sierra Leone, despite continuing high genetic diversity. Additionally, we project an increase in inbreeding, consistent with the observed gradual decline in overall case numbers. Our hypothesis suggests directional selection within gene families related to erythrocyte invasion and immune evasion (Figure 2D), with a more pronounced pressure on drug-resistant genes compared to those found in Asian samples. Generally, our findings align well with typical characteristics of the West African population, aside from the IBD ratio. We noted a lower proportion of monoclonal infections, likely due to a high rate of malaria transmission via infected mosquito bites during the outbreak. A lower monoclonal ratio typically indicates frequent superinfection or co-transmission, common in high-endemic areas (15). Additionally, the resistance genes in the Sierra Leone samples exhibited lower diversity and faced higher selection pressure than those in the Asian samples, a pattern not usually seen in African samples. We speculate that less regulated healthcare practices and varied treatment strategies in the country may have exposed parasites to increased drug resistance pressures. Whole genome scans further revealed that, in addition to drug resistance genes, gene families involved in blood cell invasion, immune evasion, and others are also subject to directional selection, suggesting that the Sierra Leone population possesses relatively strong acquired immunity.

Our study demonstrates elevated genetic diversity in the *P. falciparum* population of Sierra Leone compared to other regions. This increased diversity is particularly notable in gene families associated with erythrocyte invasion and immune evasion, highlighting the parasite's adaptive strategies in response to local environmental pressures and treatment regimens. Moreover, drug resistance patterns observed in Sierra Leonean samples — characterized by reduced diversity and heightened selection pressure compared with those from Asian regions — warrant significant concern. These findings imply that antimalarial drugs may have a more profound effect on the parasite population in Sierra Leone, potentially due to specific local treatment strategies or drug usage patterns. Our data emphasize the critical need for ongoing genomic surveillance of *P. falciparum* in high-burden areas like Sierra Leone to

tailor malaria control measures effectively. Understanding the distinctive genetic makeup of this population is essential for devising localized public health interventions and ensuring continued genomic monitoring to combat malaria effectively. Further studies are needed to delve deeper into these genetic variations and assess their effects on clinical outcomes and treatment efficacy.

Conflicts of interest: No conflicts of interest.

Acknowledgements: The staff of the China-Sierra Leone Friendship Hospital (Freetown, Sierra Leone) for their diligent collection of blood samples from individuals infected with *P. falciparum*.

Funding: Supported by the National Research and Development Plan of China (2020YFE0205700, 2018YFE0121600), the Shanghai Municipal Health Commission Planning (20214Y0214), the National Sharing Service Platform for Parasite Resources (TDRC-2019-194-30), and a grant from the Bill & Melinda Gates Foundation (INV-003421). The funding agencies did not have any involvement in the study design, data collection, analysis, interpretation, or manuscript preparation.

doi: 10.46234/ccdcw2024.071

Corresponding authors: Shengbo Chen, chensb@nipd.chinacdc.cn; Junhu Chen, chenjh@nipd.chinacdc.cn.

¹ National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research); NHC Key Laboratory of Parasite and Vector Biology; WHO Collaborating Centre for Tropical Diseases; National Centre for International Research on Tropical Diseases, Shanghai, China; ² School of Basic Medical Sciences and Forensic Medicine, Hangzhou Medical College, Hangzhou City, Zhejiang Province, China; ³ School of Global Health, Chinese Centre for Tropical Diseases Research, Shanghai Jiao Tong University School of Medicine, Shanghai, China; ⁴ Hainan Tropical Diseases Research Center (Hainan Sub-Center, Chinese Center for Tropical Diseases Research), Haikou City, Hainan Province, China.

Submitted: March 12, 2024; Accepted: April 21, 2024

REFERENCES

- MalariaGEN, Abdel Hamid MM, Abdelraheem MH, Acheampong DO, Ahouidi A, 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>.
- Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, et al. Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. Genome Res 2016;26(9):1288 – 99. <https://doi.org/10.1101/gr.203711.115>.
- World Health Organization. WHO guidelines for malaria, 14 March 2023. Geneva: World Health Organization; 2023. <https://iris.who.int/handle/10665/366432>.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297 – 303. <https://doi.org/10.1101/gr.107524.110>.
- MalariaGEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. eLife 2016;5:e08714. <https://doi.org/10.7554/eLife.08714>.
- Excoffier L, Lischer HEL. 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. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
- Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. Malar J 2018;17(1):196. <https://doi.org/10.1186/s12936-018-2349-7>.
- Bachmann A, Scholz JAM, Janßen M, Klinkert MQ, Tannich E, Bruchhaus I, et al. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. Malar J 2015;14:274. <https://doi.org/10.1186/s12936-015-0784-2>.
- Tumwebaze PK, Conrad MD, Okitwi M, Orena S, Byaruhanga O, Katairo T, et al. Decreased susceptibility of *Plasmodium falciparum* to both dihydroartemisinin and lumefantrine in northern Uganda. Nat Commun 2022;13(1):6353. <https://doi.org/10.1038/s41467-022-33873-x>.
- 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. [https://doi.org/10.1016/s2666-5247\(20\)30094-x](https://doi.org/10.1016/s2666-5247(20)30094-x).
- de Vries LE. Steepen the Curve: The Potential of Pantothenamides in Malaria Control. Nijmegen: Radboud University. 2021. <https://repository.uibn.ru.nl/handle/2066/240243>
- Menard D, Dondorp A. Antimalarial drug resistance: a threat to malaria elimination. Cold Spring Harb Perspect Med 2017;7(7):a025619. <https://doi.org/10.1101/cshperspect.a025619>.
- Winzeler EA. Malaria research in the post-genomic era. Nature 2008;455(7214):751 – 6. <https://doi.org/10.1038/nature07361>.
- Chan K, Tusting LS, Bottomley C, Saito K, Djouaka R, Lines J. Malaria transmission and prevalence in rice-growing versus non-rice-growing villages in Africa: a systematic review and meta-analysis. Lancet Planet Health 2022;6(3):e257 – e69. [https://doi.org/10.1016/S2542-5196\(21\)00349-1](https://doi.org/10.1016/S2542-5196(21)00349-1).
- Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. Nature 2012;487(7407):375 – 9. <https://doi.org/10.1038/nature11174>.