scientific reports

Check for updates

OPEN

Unveiling mismatch of RTS S AS01 and R21 Matrix M malaria vaccines haplotype among Ethiopian *Plasmodium falciparum* clinical isolates

Aynalem Mandefro[1,2]✉, Alebachew Messele Kebede[1], Mitchel Katsvanga[3], Fatoumatta Cham[3], Eniyou Oriero[3], Alfred Amambua-Ngwa[3] & Lemu Golassa[1]✉

Malaria vaccines, RTS, S/AS01 and R21/Matrix which are based on the *Plasmodium falciparum* circumsporozoite protein (*Pfcsp*) have been approved by WHO for broad use in children in Africa. However, the extensive genetic diversity of *Pfcsp* limited its effectiveness, as vaccine efficacy reduced against non-vaccine strains. Using Oxford Nanopore Technology, we conducted amplicon sequencing of the full-length *Pfcsp* gene from 96 clinical isolates collected from three health centers in Ethiopia and compared the results against a reference genome. The result showed absence of population differentiation among the Ethiopian isolates. The N-terminal region was relatively conserved, with a KLKQP motif was present across all isolates. However, mutation at position A98G and an insertion of amino acids (DGNNNNGDNGREGKDEDKR) were identified in this region. The number of NANP and NVDP repeats of the central region per haplotype ranged from 39 to 42. Additionally, the Th2R and Th3R epitopes in the C-terminal region exhibited extensive polymorphism with at least one amino acid substitution compared to the reference strains. Notably, none of the Ethiopian *Pfcsp* haplotypes matched the vaccine haplotype. Furthermore, haplotype network and phylogenetic tree analyses shown considerable similarity among local and global isolates. The findings of this study revealed a high *Pfcsp* genetic diversity highlighting the need for further studies to inform allele selection for universal or region-specific vaccine development as this may influence vaccine efficacy.

Keywords Vaccines, Circumsporozoite protein, Genetic diversity, *Plasmodium falciparum*

Malaria continues to pose a significant global health burden, with reported cases steadily increasing. In 2023, an estimated 263 million cases and 597,000 deaths were recorded, representing an 11 million case increase compared to 2022[1]. This rise in malaria cases, even during the elimination era, highlights the urgent need for additional tools to combat the disease and achieve global elimination goals[2].

Malaria vaccines are a cost-effective intervention that complements existing control measures and can accelerate efforts to eliminate the disease in endemic regions[3–5]. Since the 1980s, the development of malaria vaccines has progressed through multiple phases of clinical trials, recently culminating in the approval of two pre-erythrocytic vaccines[3,4,6–8]. In October 2021, WHO approved RTS, S/AS01 as the first malaria vaccine for children in areas with moderate to high *P. falciparum* transmission[9]. This was followed by the approval of a second vaccine, R21/Matrix-M, for use in children in regions with low to moderate prevalence[4,10]. Both vaccines have proven safe and effective in preventing malaria in children; however, the R21/Matrix-M vaccine demonstrates superior efficacy, offering up to 77% protection compared to the 36.3% efficacy observed with RTS, S[10,11]. Despite targeting the *P. falciparum* circumsporozoite protein (*Pfcsp*) gene, the structural design and composition of the two vaccines differ slightly[8,12].

[1]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia. [2]College of Natural and Computational Science, Hawassa University, Hawassa, Ethiopia. [3]Medical Research Council Unit the Gambia, London School of Hygiene and Tropical Medicine, Banjul, The Gambia. ✉email: anaafman@gmail.com; lemu.golassa@aau.edu.et

Pfcsp is the dominant surface protein coating the infectious sporozoite, playing a critical role in sporozoite functionality and hepatocyte invasion[3,13,14].The *Pfcsp* gene encodes a protein of 412 amino acids, divided into three regions: the conserved N-terminal region, the central repeat region, and the C-terminal region[8,13]. The central region, primarily composed of tandem NANP repeats with a smaller number of NVDP repeats, contains the most immunogenic B-cell epitopes[15]. In contrast, the C-terminal region includes two polymorphic epitope regions, Th2R and Th3R, recognized by CD4 + and CD8 + T cells, respectively[16].

Pfcsp exhibits significant global genetic polymorphism. Studies conducted in malaria-endemic regions have revealed substantial variations in the vaccine antigen, with a low frequency of variants matching the vaccine haplotype[8,13,16]. This extensive genetic diversity presents a major challenge to developing a universally effective malaria vaccine. Furthermore, it has been well-documented that vaccine efficacy decreases against non-matched vaccine haplotypes, indicating that allele-specific immunity plays a critical role in providing protection[17,18]. For instance, Neafsey et al.[19] demonstrated that RTS, S vaccine efficacy dropped from 50.3% for parasites perfectly matching the vaccine haplotype to 33.4% for parasites with any amino acid mismatches, emphasizing the importance of allele-specific immunity.

In Ethiopia, despite substantial efforts to eliminate malaria by 2030, cases have surged dramatically. In 2023, 9.56 million cases were reported, compared to just 1.3 million cases in 2021[1,20]. To address this alarming increase, the introduction of vaccines alongside other intervention measures is critical to prevent further escalation. However, there is currently no comprehensive data on the genetic diversity or variation of the *Pfcsp* gene within Ethiopia. Therefore, it is imperative to investigate the genetic diversity of *Pfcsp* in clinical isolates from different regions of the country. This research is important for ensuring the successful deployment of malaria vaccines in the near future. Moreover, as no prior studies have been conducted in Ethiopia, the findings of this study will serve as a valuable baseline for future research.

Method

Sample collection, DNA extraction, and quantification

Blood samples were collected from consenting individuals aged 5 years and older who visited Metehara, Zenzelema, and Kolla Shelle health centers in Ethiopia (Fig. 1) between June 2022 and March 2023. A total of 120

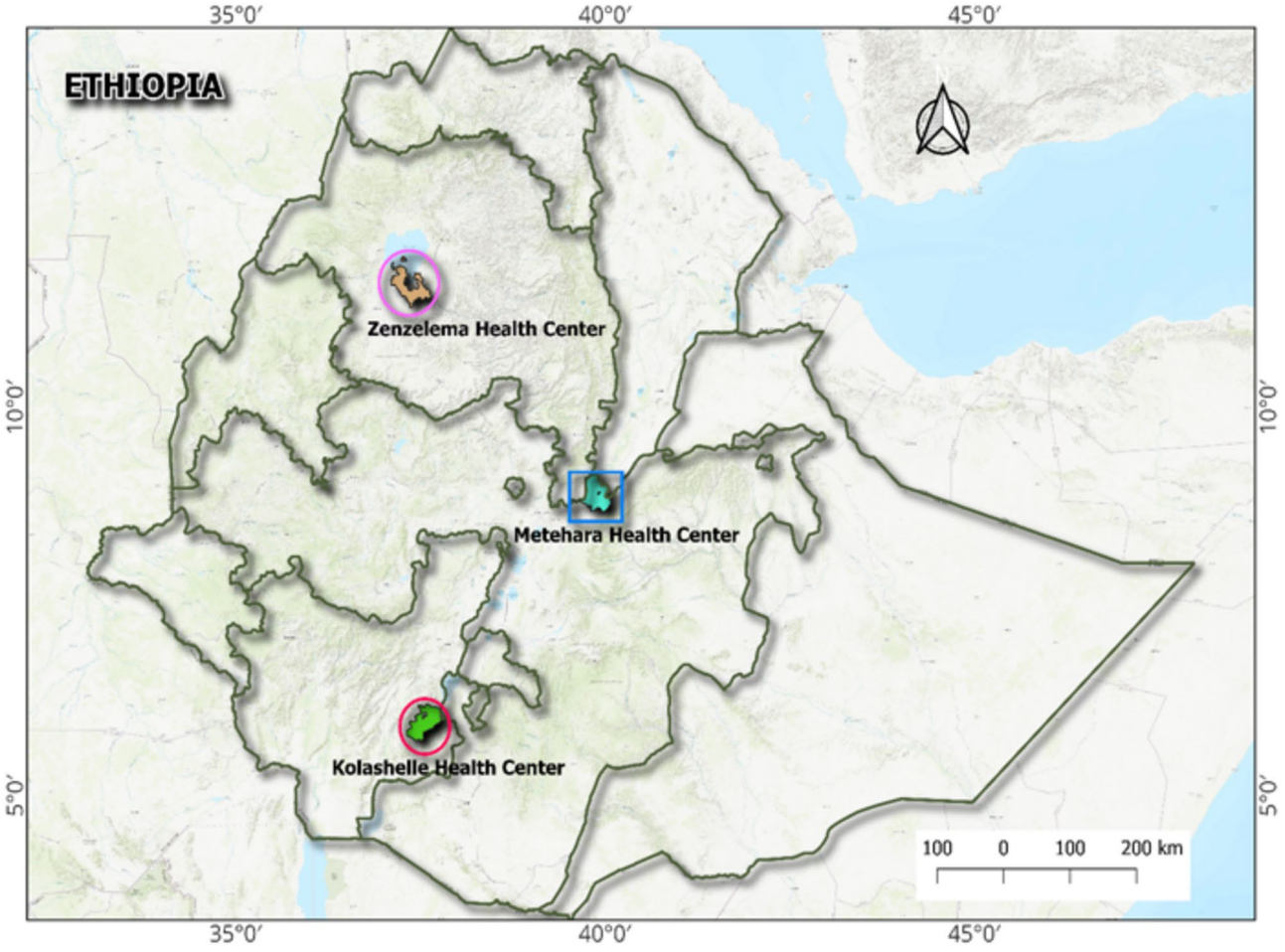

Fig. 1. Map of Ethiopia showing the included health centers (generated using Quantum GIS version 3.36.0 software).

blood samples, confirmed as *P. falciparum* positive by microscopy, were processed for DNA extraction using the G-spin™ Total DNA Extraction Mini Kit. The extracted DNA was then shipped to the Medical Research Council Unit at The Gambia, London School of Hygiene and Tropical Medicine, for further analysis and quantification using *VarATS*[21]. The samples analyzed in this study were part of a previously published investigation on *Pfhrp2/3* genetic diversity in Ethiopia[22].

Amplicon generation and sequencing

One hundred twenty PCR-confirmed *P. falciparum* positive samples were included for *Pfcsp* amplicon sequencing. The full *Pfcsp* gene was amplified via conventional PCR using forward primer CSP-F (5'-TGGGA AACAGGAAAATTGGTAT-3') and reverse primer CSP-R (5' TACGACATTAAACACACTGGAA-3'). Briefly, 5 µl of DNA template was added into a master mix containing 1 × ThermoPol reaction buffer (New England BioLab, USA), 0.3mM dNTP (New England BioLab, USA), 0.5 µl of Taq DNA polymerase (0.1 U/µL), (New England BioLab, USA), 0.5 µl of MgCl2, and 0.5 µM of each forward and reverse primer in a total PCR reaction volume of 25 µl. The amplifications were performed under the following PCR conditions: 5 min at 95 °C initial denaturation, followed by 34 cycles of 95 °C denaturation for 15 s, 54 °C annealing for 30 s, 68 °C extension for 1 min, and a final extension at 72 °C for 2 min. The *P. falciparum* lab strain 3D7 was used as the positive control, while molecular-grade water served as the negative control. The amplified *Pfcsp* products were visualized on a 1.5% agarose gel using a 1 kb DNA ladder for size determination. Sequencing libraries were then prepared from the generated amplicons using the SQK-109 and EXP-NBD196 ligation sequencing kits from Oxford Nanopore Technologies, following the manufacturer's instruction. Sequencing was performed on the GridION platform with a FLO-MIN114 Flow Cell.

Bioinformatics processing and analysis

Base-calling and barcoding was performed by Guppy v. 7.0.9 within MinKNOW v23.07.5. Base calling and demultiplexing parameters were set with minimum barcoding quality of 60, a minimum read quality score was set to 9, and no filter applied for minimum read length. Quality filtered reads were aligned to the reference genomic sequence for *Pfcsp* (NC_000521.4) using Minimap2 (version 2.21-r1071). The output was then sorted and converted into a binary alignment map (BAM) file using Samtools (version 1.8). Additionally, the sequence reads were mapped to the full *P. falciparum* 3D7 reference genome (PlasmoDB PF3D7_0304600) and manually inspected using the Integrative Genomics Viewer (IGV) tool (version 2.3.94) to confirm the presence of reads at the expected coding regions on the chromosomes.

FASTQ files that successfully aligned to the target gene were subsequently used as input for de novo assembly. A phylogenetic tree was constructed to explore the relationship between Ethiopian *Pfcsp* sequences and reference sequences. Maximum likelihood methods were used to construct the tree in MEGA7 software[23]. Nucleotide diversity parameters, including the number of haplotypes (H), haplotype diversity (Hd), and the number of segregating sites (S), were calculated using DnaSP software (v6.12.03)[24]. To examine the genetic relationships among sequences from the study sites and global *Pfcsp* sequences, a haplotype network was constructed using PopArt 1.7 software[25]. Additionally, genetic differentiation between parasite populations was assessed by estimating gene flow using Wright's Fixation Index (FST) and heterozygosity (HS)[24].

Result

From the 120 samples collected across Kola Shele, Metehara and Zenzelema health centers, 96 produced *Pfcsp* amplicons suitable for sequencing. A library of 96 *Pfcsp* amplicons was prepared and sequenced using Oxford Nanopore Technology (ONT) to assess their genetic diversity. Of these, 85 full length *Pfcsp* sequences passed quality filtering steps, allowing for the generation of consensus sequences. Sequencing analysis showed that the full-length of examined isolates of *Pfcsp* sequences were 1066 bp in size and comprised 11 haplotypes with a haplotype diversity of 0.761. In addition, the analysis of the three blocks of *Pfcsp* nucleotide sequences, two haplotypes were identified in the N-terminal region, while ten haplotypes were detected in the central and C-terminal regions (Table 1).

Genetic differentiation and population structure

Genetic diversity and differentiation of the *Pfcsp* populations across the three health centers Zenzelema, Metehara, and Kola Shele were analyzed using Hs for within population diversity and Fst for population differentiation. The study found that all three population pairs show low within-population genetic diversity, with Hs values ranging from 0.00478 to 0.01067. In addition, Fst values around zero across all the investigated health centers indicate that these populations are not genetically differentiated and are likely sharing genetic material, possibly

Region	H	Hd	S	Nucleotide diversity (π)	Jukes & Cantor (π)
Full Sequence	11	0.761	60	0.01955	0.01992
N-Terminal	2	0.410	1	0.00287	0.00288
Central Repeat	10	0.756	44	0.02680	0.02756
C-Terminal	10	0.760	15	0.01732	0.01763

Table 1. Haplotype and nucleotide diversity across different regions of the *Pfcsp* gene in Ethiopian isolate. H: number of unique haplotypes, Hd: haplotype diversity, S: number of segregating sites.

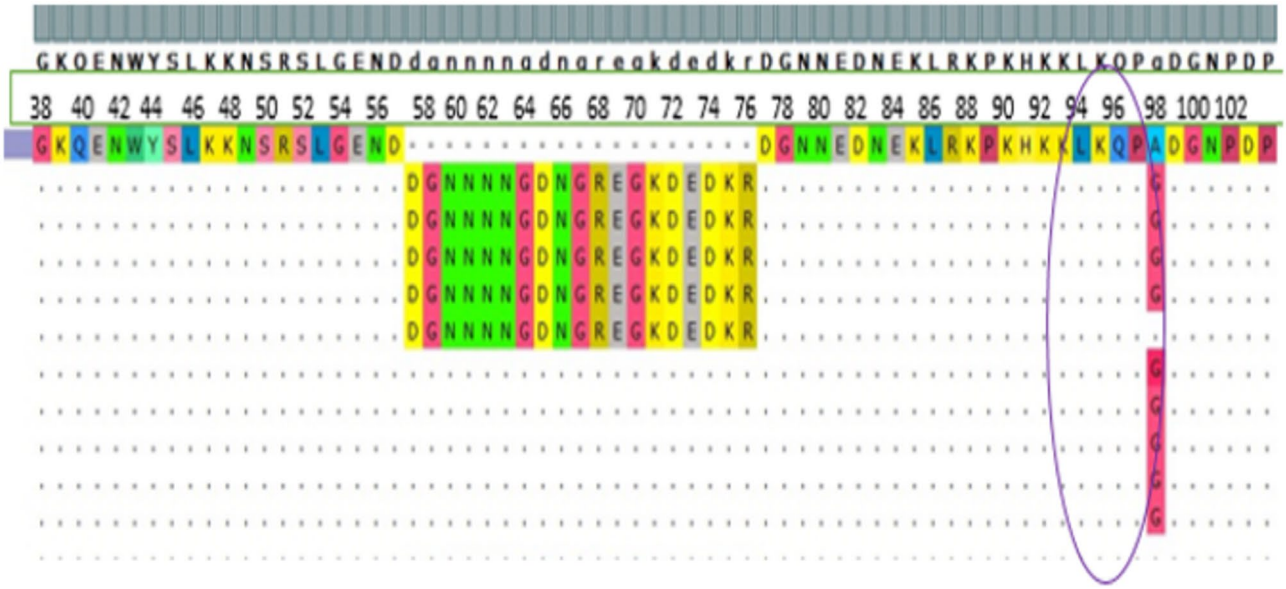

Fig. 2. Genetic diversity in the N-terminal region of *Pfcsp* among Ethiopian isolates. The sequences displayed are representative of the total sequences analyzed to illustrate any observed polymorphisms. Dots indicate the insertion of the amino acid segment (DGNNNNGDNGREGKDEDKR) and denote identical amino acid residues to those in the reference strain (3D7 NC-000521). The Th1R epitope motif (KLKQP), highlighted by a purple circle, plays a role in the invasion of sporozoite in the mosquito salivary gland and their adherence to hepatocytes was present in all isolate.

Haplotypes	No of isolate (%)	No of repeats		
		NANP	NVDP	Total
Haplotype_1	34 (40)	34	5	39
Haplotype_2	1 (1.18)	38	4	42
Haplotype_3	5 (5.9)	36	4	40
Haplotype_4	19 (22.3)	37	5	42
Haplotype_5	16 (18.8)	38	4	42
Haplotype_6	2 (2.35)	38	4	42
Haplotype_7	2 (2.35)	35	6	41
Haplotype_8	1 (1.18)	35	6	41
Haplotype_9	1 (1.18)	36	4	40
Haplotype_10	4 (4.7)	36	4	40

Table 2. Frequency distribution of repeat counts for *Pfcsp* central repeat region haplotypes in Ethiopian isolate.

due to gene flow among them despite the isolates were collected from health centers located far apart, potentially separated by distances of thousands of miles (Table S1).

Genetic diversity in the N terminal region

Genetic polymorphism in the N-terminal region of the *Pfcsp* gene was assessed using the reference sequence (NC-000521). The finding revealed that the N-terminal region of Ethiopian *Pfcsp* gene was highly conserved across all the samples. Two distinct haplotypes were identified (Table 1). Haplotype 1 contained an insertion of a 19 amino acid segment at position ^{58}DGNNNNGDNGREGKDEDKR^{76}, whereas haplotype 2 lacked this insertion. However, both haplotype exhibited an A98G amino acid substitution. More importantly, the KLKQP amino acid motif was conserved in all isolates studied regardless of the haplotypes (Fig. 2).

Genetic diversity in the central repeat region

The central repeat region of *Pfcsp* in Ethiopian isolates consists of conserved tetra peptide amino acid motifs, NANP and NVDP, with no novel variants observed among all isolates analyzed. A total of 10 haplotypes were identified, with haplotype diversity of 0.756 (Table 1). Haplotype 1 was the most common, found in 40% of isolates (34/85) and followed by Haplotype 4 and Haplotype 5, occurring in 22.3% (19/85) and 18.8% (16/85) of isolates, respectively. In contrast, Haplotypes 2, 8, and 9 were each identified in only a single isolate (Table 2).

The number of NANP repeats in each haplotype ranged from 34 to 38, while NVDP repeats ranged from 4 to 6 contributing to the polymorphism in the central repeat region. Consequently, the total number of NANP and NVDP repeats per haplotype varied from 39 to 42, where the majority 40% of isolate contained 39 tetra-peptide amino acid repeat (Table 2). Further, Fig. 3a and b illustrates the nucleotide alignment of the 10 haplotypes for the central region and structural patterns of NANP and NVDP tetra-peptide motifs in the identified haplotypes of the central repeat region. A similar NANP and NVDP arrangement across all 85 analyzed *Pfcsp* sequences is shown in Figure S1.

Genetic diversity in the C-terminal region

Analysis of genetic diversity in the C-terminal region of the Ethiopian *Pfcsp* gene revealed 10 distinct haplotypes (Table 1). Haplotype 1 was the most common, found in 34 isolates, followed by Haplotype 4 and Haplotype 5, which appeared in 19 and 15 isolates, respectively. Haplotype 3, 6, 7, and 10 were observed in 5, 2, 3, and 4 isolates, respectively, while the remaining haplotypes (2, 8, and 9) were each found in only one isolate.

In the C-terminal regions Th2R and Th3R epitope showed polymorphism compared to the rest of amino acid sequence in the region. Several non-synonymous mutations were identified in these epitopes compared to 3D7 (NC-000521) and CSP-NF54 sequences that are presented in RTS, S and R21M vaccine. Surprisingly, our analysis showed that none of the Ethiopian *Pfcsp* haplotypes matched the vaccine haplotype in the Th2R (^{311}PSDKHIKEYLNKIQNSL^{327}) and Th3R (^{352}NKPKDELDYANDI^{364}) epitopes. Each haplotype contained at least one amino acid substitution relative to the reference strains, highlighting the polymorphic nature of all identified haplotypes. Through examination of each haplotypes showed that amino acid substitutions at 10 positions in the Th2R epitope (K314Q, K317E/T, E318K/Q, N321K, K322R/T, Q324K, and L327I) and at 6 positions in the Th3R epitope (N352G, P354S, D356N, E357Q, D359N, and A361E) (Fig. 4).

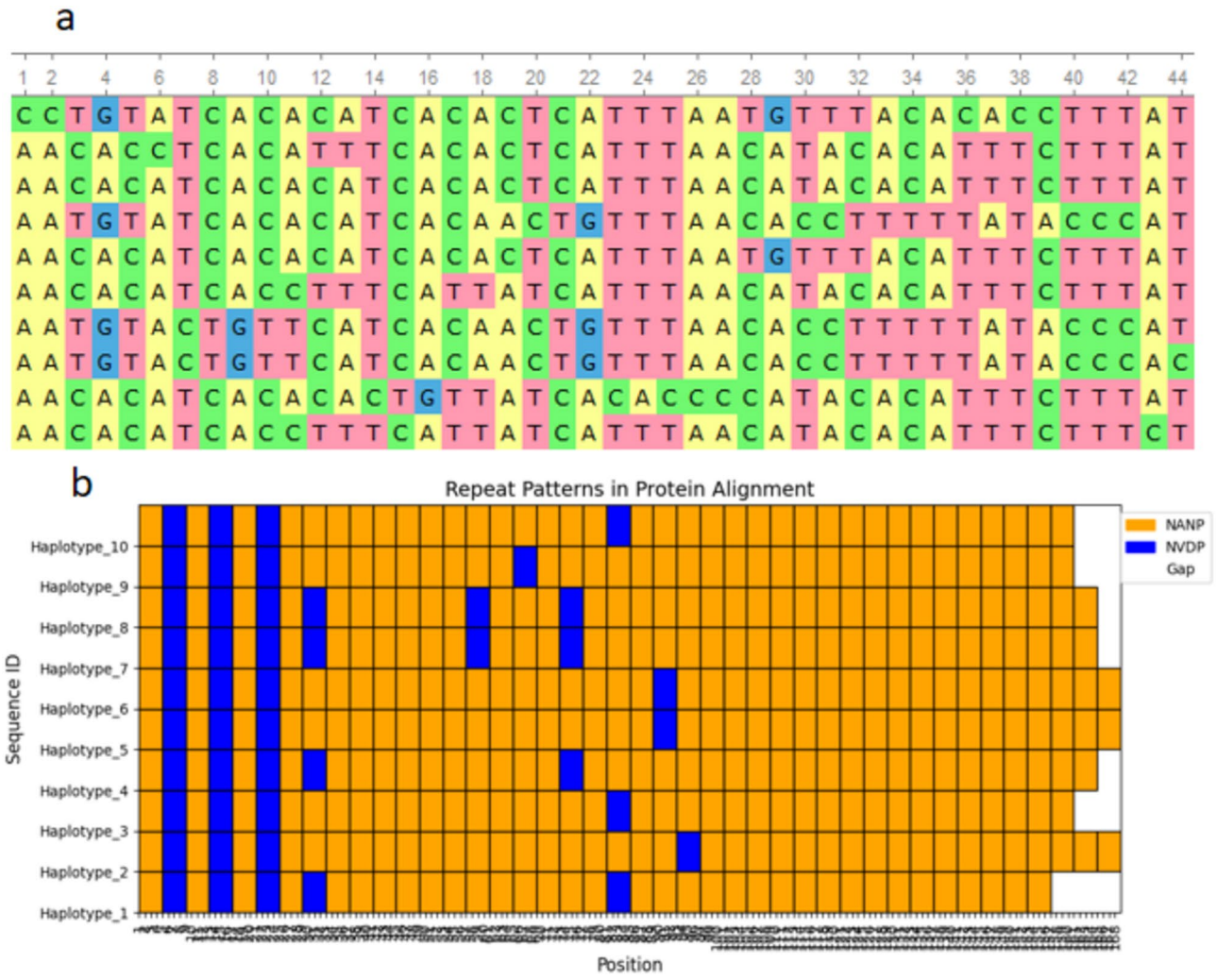

Fig. 3. (**a**) Nucleotide alignment of the 10 haplotypes in the central region. (**b**) Polymorphism of central repeat region in Ethiopian isolates compared to the reference sequence 3D7 (NC-000521). Haplotype 1 to Haplotype 10 denote the structural arrangements of identified haplotypes, with color distinctions for the tetra-peptide motifs: orange represents NANP, and blue represents NVDP.

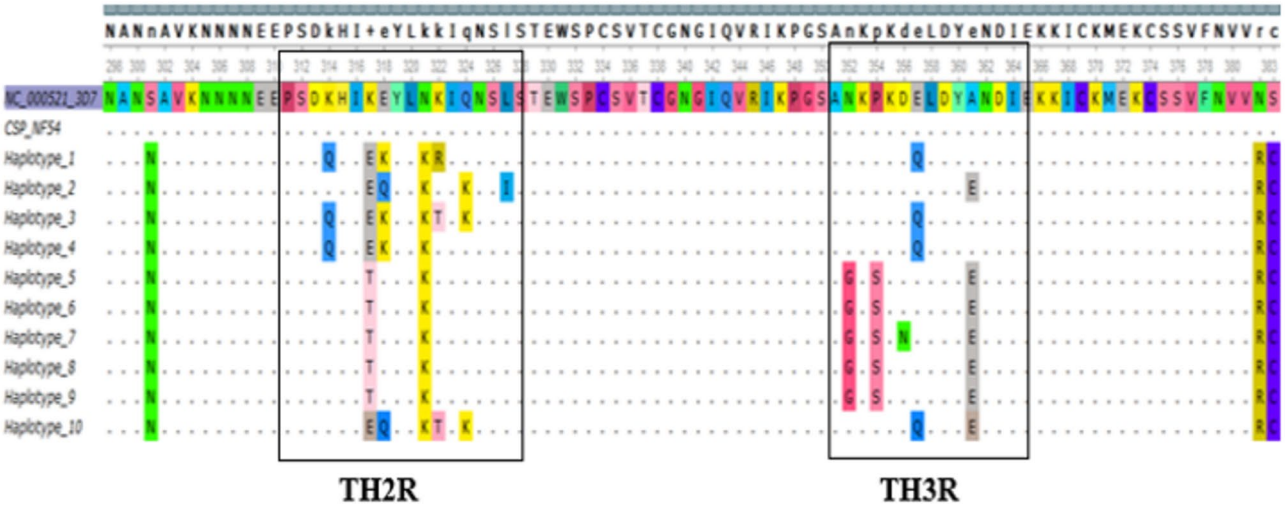

Fig. 4. Polymorphism of C-terminal region of the Ethiopian *Pfcsp* gene. The comparison is based on the *P. falciparum* reference sequence 3D7 (NC-000521) and CSP-NF54. Dots represents identical amino acids to those in the reference sequences. The black rectangle boxes on the left side and right side represents the Th2R and Th3R regions, respectively.

Population	*N*	H	Hd	S	Nucleotide diversity (π)	Jukes Cantor (π)
All_ Site	10	10	0.760	15	0.01732	0.01763
Metehara	28	6	0.767	13	0.01695	0.01726
Kolla Shele	26	7	0.782	14	0.01802	0.01833
Zenzelema	31	7	0.757	14	0.01755	0.01787

Table 3. Genetic diversity of the *Pfcsp* C-terminal region among Ethiopian isolate across the study site. N: number of sequences, H: number of unique haplotypes, Hd; haplotype diversity, S; number of segregating sites.

Moreover, the study found that genetic polymorphism in C- terminal region was consistent among all isolate included from the three health centres, with only slight differences. Genetic diversity indices demonstrated that moderate nucleotide diversity ranging ~ 0.017 to 0.0182 was observed among isolates. Accordingly, isolate from Kolla Shele health center showed relatively highest nucleotide diversity (0.0182), while the lowest nucleotide diversity was observed in isolate from Metehara health center (~ 0.017). Additionally, haplotype diversity was uniformly high across all study population, ranging from 0.76 to 0.78. Besides, the number of segregating site was almost comparable across the study sites (Table 3).

Haplotype network analysis

Haplotype network was constructed using sequences from C-terminal region to assess the genetic relationship among isolates from the three health centers. The analysis showed that four haplotypes were shared among isolates from all centers, while two haplotypes were shared between two health centers indicating the genetic connectivity among Ethiopian isolate. Additionally, four singleton haplotype were identified, each unique to isolate from specific health centers suggesting local differentiation (Fig. 5a).

Furthermore, sequences from Sudan, Ghana, and Thailand were retrieved from NCBI and analyzed alongside our sequences to construct a haplotype network. Interestingly, the network showed that Ethiopian *Pfcsp* haplotypes share a common haplotype with those from Ghana, Sudan, and Thailand, represented by a large central circle indicating genetic overlap (Fig. 5b).

Phylogenetic analysis of Ethiopian *Pfcsp* isolates

The phylogenetic tree displayed that, although few samples clustered by study site, samples from all the three health centers are notably mixed within individual clades. Likewise, the tree further showed that *Pfcsp* gene sequences from local isolates showed similarity to *Pfcsp* reference strains obtained from global databases (Fig. 6).

Discussion

The two WHO-approved malaria vaccines, RTS, S/ AS01 and R21/Matrix that are based on the *Pfcsp* gene, offer hope to fight against malaria. These vaccines support malaria elimination strategies by serving as additional tools to combat the disease and address the growing concern of rising cases, particularly among children in malaria-endemic African countries, which account for 95% of reported cases[1]. Genetic polymorphism in the *Pfcsp* gene has been reported as factor limiting vaccine efficacy, as both vaccines were developed based on a single allelic

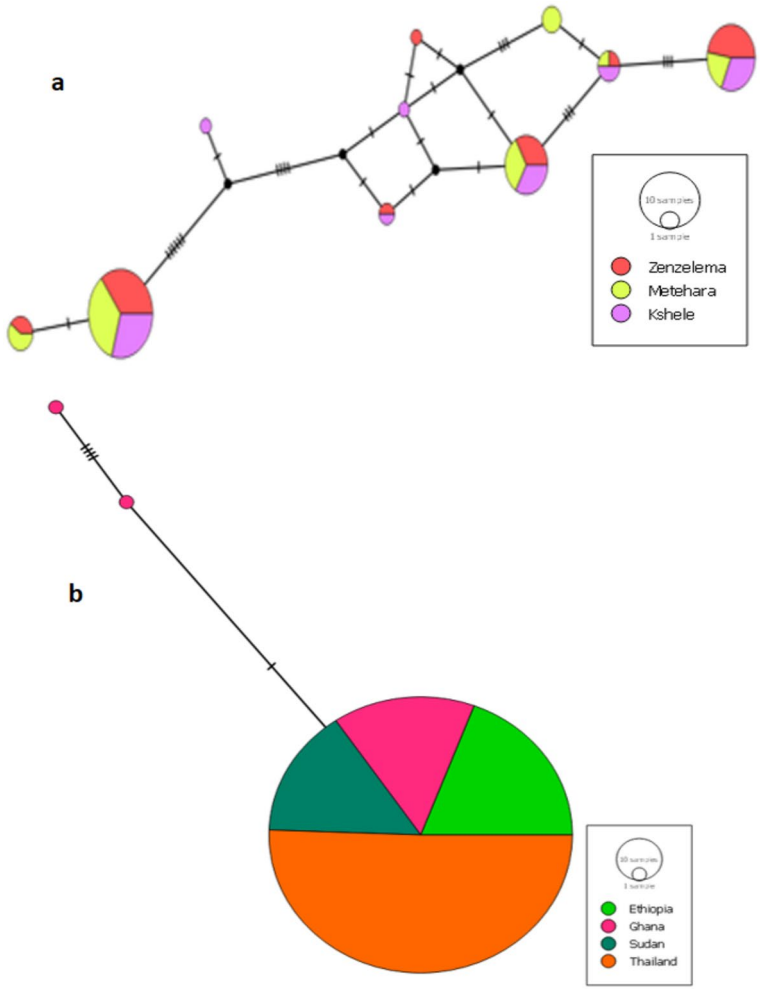

Fig. 5. (**a**) Haplotype network analysis of Ethiopian *Pfcsp* C-terminal region. The networks showed the 10 haplotypes identified in the present study. The branch length indicate the divergence and the size of each node indicates the frequency of each node. The color of each node represents the three health centers. (**b**) Haplotype network showing the genetic relationships of *Pfcsp* sequences from Ethiopia, Ghana, Sudan, and Thailand. The central pie chart, composed four colors representing each country, illustrates shared haplotypes across the region.

form of the antigen[17–19]. Investigating the genetic diversity of *P. falciparum* in Ethiopia is a key to generate data that support the future rollout of malaria vaccine in the country.

The present study showed, although the clinical isolates were collected from three malaria-endemic sites in Ethiopia with distinct ecological settings and located far apart, the *Pfcsp* gene diversity indicated no genetic differentiation (FST ~ 0) among these populations. This findings concur with earlier reports that documented absence of genetic differentiation among isolates collected from two health centers in Ghana[26] and five health centers in Tanzania[8], despite being separated by approximately 800 km. Besides, the result from the haplotype network and phylogenetic tree analysis confirm and further clarify the genetic mixing among the study isolates and also revealed the Ethiopian *Pfcsp* sequences share similarities from those obtained from other part of the world. The findings suggest the ongoing gene flow among parasites that could be due to continuous human migration between different areas, which facilitates genetic mixing[8,27].

The study exhibited the N-terminal region is highly conserved revealing low genetic polymorphism that align with findings from isolates in other endemic countries[12,27,28]. Additionally, the study determined that the KLKQP motif remained well conserved across Ethiopian isolates, similar to Kenyan[12], Sudan[29] and Indian[30] isolates. Conversely, all isolates exhibited the non-synonymous A98G mutation, and the majority of isolates had a 19-amino-acid insertion in the middle of N-terminal region. The study finding is consistent with previous reports[27,29,31]. The N-terminal region of *Pfcsp* is crucial for sporozoite invasion into hepatocytes[32,33]. The occurrence of non-synonymous mutations and amino acid insertions in this region might indicate parasite adaptation mechanisms to evade the human immune response[29]. These changes could result in structural modifications that impact sporozoite attachment and invasion of hepatocytes[12,30]. The low genetic polymorphism, combined with the functional significance of the N-terminal region, highlights its potential as a key component of *Pfcsp* based vaccines[34].

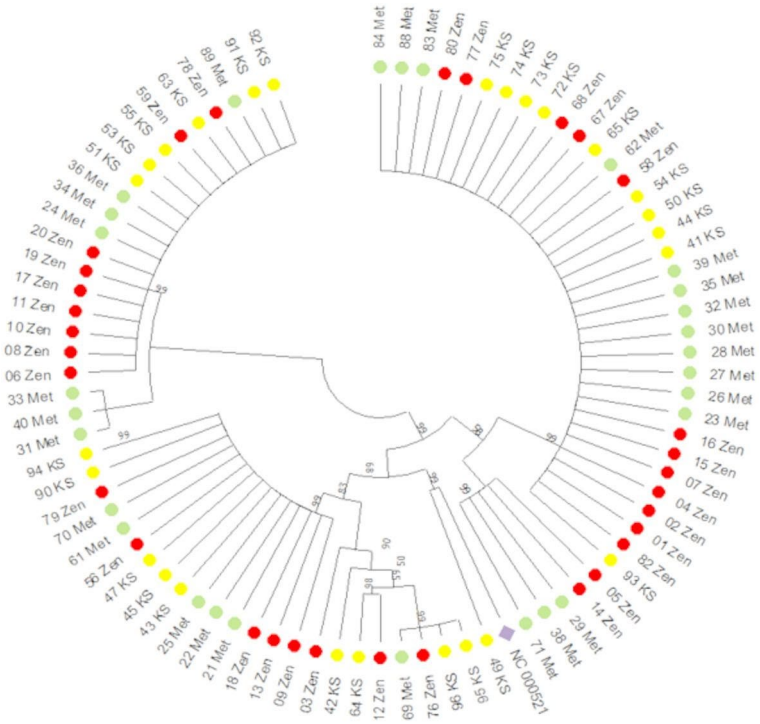

Fig. 6. Phylogenetic tree of the Ethiopian *Pfcsp* sequences along with reference sequences (NC- 000521). Red circles correspond to sequences from Zenzelema ($n = 31$), green circles correspond sequences from Metehara ($n = 28$), and yellow circles correspond sequences from Kolla Shele ($n = 26$). Purple diamond shape correspond sequences from the NCBI (*Pfcsp* reference).

The current study revealed that the number of tetra-peptide amino acid motifs NANP and NVDP in the central repeat region varied among isolates, ranging from 39 to 42. This finding is very similar with previous reports from African isolates[12,27,29,31]. In contrast, isolates with higher number of NANP and NVDP repeats than the present study were reported from India[30] and Myanmar[35]. Interestingly, unlike results from other countries, no novel tetra-peptide motifs were observed in Ethiopian isolates. However, several unique repeat motifs, including MANP, NANS, IVDP, NTNP, NVVP, NAHP, NAKP, NATP, NVNP, NAIP, NANL, NPNP, NVAD, NADP, SANP, and KANP, have been reported in various endemic countries[12,27,30,35]. The variation observed in the number of tetra-peptide motifs in the central repeat region might be attributed to geographical differences, parasite genetic diversity, or varying in selection pressures[29]. It has been far documented that the tetra-peptide amino acid repeats of the central repeat region is an important component of the immuno-dominant B-cell epitopes targeted by the neutralizing antibodies[12]. Nevertheless, this variation can contribute to polymorphism in the length and structure of the gene[12,29], which could affect stability of the gene[36]. Further, the central repeat region is an important component of both the RTS, S and R21-Matrix malaria vaccines, both vaccines include 19 NANP repeats that are important for inducing antibody response[8,37,38]. However, no studies have yet reported whether variations in the number of repeats might influence the effectiveness of malaria vaccines. Hence, the polymorphism in this region requires further study.

In line with reports from other malaria-endemic countries, this study also identified extensive polymorphisms in the C-terminal region among Ethiopian isolates[12,27,30]. The polymorphisms are more frequently observed in the Th2R and Th3R regions that serves as epitopes for CD4 + and CD8 + T cell responses, respectively[12,26]. Notably, the Th2R region exhibits greater polymorphism than the Th3R region as documented by[12,16,26,30]. Besides, our findings revealed amino acid polymorphism in the Th2R and Th3R regions, ranging from one to six that enable the parasite to evade host immune pressure. This might reduce vaccine efficiency, agreeing with previous studies done elsewhere[26,39]. More importantly, none of the Ethiopian isolates matched the Th2R and Th3R haplotypes of the 3D7/ NF54 vaccine strain signifying the need for designing regional specific vaccines. The findings align with various studies conducted in Africa and globally that have reported a low level of matching between clinical isolates and the vaccine reference strains[8,26,30,39]. Similarly, comparative studies on the genetic diversity of the 3D7 *Pfcsp* gene revealed that the vaccine strains shared only a 0.2–5.0% match with the global *Pfcsp* gene pool[38]. The absence of the vaccine haplotype in Ethiopian *Pfcsp* strains may reduce the efficacy of the RTS, S/AS01 and R21 vaccines, as prior studies have demonstrated a positive correlation between vaccine efficacy and haplotype matching[17–19].

Study limitations

While our study offers key information regarding to the genetic diversity of *csp* gene from *P. falciparum* clinical isolates in Ethiopia for the first time, there are few important limitations. First, the scope of the study, focus

on only three health centers in malaria endemic region limit the finding representativeness for understanding the full genetic diversity across the country. Hence, continuous monitoring with figurative number would enhance the ability to capture the broader genetic diversity. Second, the findings are based on gene sequence observations to identify any polymorphisms. However, further research is needed to assess the correlation of this polymorphism on vaccine efficacy, as well as their impact on parasite fitness and host biology.

Conclusions

The current study revealed a substantial level of genetic diversity in the *Pfcsp* gene in Ethiopia. However, the lack of genetic differentiation across the study sites highlighted the ongoing gene flow among parasite populations. In addition, the study also showed the N-terminal region and central repeat region of the Ethiopian *Pfcsp* gene are relatively conserved, whereas the C-terminal region particularly Th2R and Th3R exhibit high levels of polymorphism. The polymorphism in Th2R and Th3R is seen in all identified haplotypes, resulting in none of the Ethiopian *Pfcsp* sequences matching the vaccine haplotypes. Lastly, this study is the first to examine the genetic diversity of *P. falciparum* in Ethiopia, addressing knowledge gaps by providing critical insights to enhance *Pfcsp*-based vaccines.

Data availability

All the analyzed data have been incorporated into the manuscript and supplementary material. Additional information can be obtained from corresponding authors based on reasonable request. The nucleotide sequences obtained from this study have been deposited in GenBank (accession numbers PQ801375-PQ80145). Alternatively, the FASTQ files generated for this study have been uploaded to Zenodo and can be accessed up on reasonable request through the provided link (https://zenodo.org/records/15130571).

Received: 2 January 2025; Accepted: 25 April 2025
Published online: 29 April 2025

References

1. WHO. *Addressing Inequity in the Global Malaria Response* (World malaria report, 2024).
2. WHO. *Global Technical Strategy for Malaria 2016–2030* (World Health Organization, 2016).
3. Duffy, P. E. & Gorres, P. J. Malaria vaccines since 2000: Progress, priorities, products. *Npj Vaccines* **5**, 48 (2020).
4. WHO. WHO recommends R21/Matrix-M vaccine for malaria prevention in updated advice on immunization, (2023). https://www.who.int/news/item/02-10-2023-who-recommends-r21-matrix-m-vaccine-for-malaria-prevention-in-updated-advice-on-immunization
5. Penny, M. A. et al. Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: A systematic comparison of predictions from four mathematical models. *Lancet* **387**, 367–375 (2016).
6. Millet, P. Current status and prospects of malaria vaccines. *J. Travel Med.* **2**, 96–98 (1995).
7. Hill, A. V. Vaccines against malaria. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **366**(1579), 2806–2814 (2011).
8. Lyimo, B. M. et al. Genetic polymorphism and evidence of signatures of selection in the *Plasmodium falciparum* circumsporozoite protein gene in Tanzanian regions with different malaria endemicity. *Malar. J.* **23**, 139 (2024).
9. Zavala, F. RTS,S: The first malaria vaccine. *J. Clin. Invest.* **132**, 1 (2022).
10. Aderinto, N. et al. A perspective on Oxford's R21/Matrix-M™ malaria vaccine and the future of global eradication efforts. *Malar. J.* **23**, 16 (2024).
11. RTS & S Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial. *Lancet* **386** (9988), 31–45 (2015).
12. Maina, M. et al. Genetic variation of the *Plasmodium falciparum* circumsporozoite protein in parasite isolates from Homabay County in Kenya. *Front. Parasitol.* **3**, 1346017 (2024).
13. El-Moamly, A. A. & El-Sweify, M. A. Malaria vaccines: The 60-year journey of hope and final success—lessons learned and future prospects. *Trop. Med. Health.* **51**, 29 (2023).
14. Laurens, M. B. & RTS S/AS01 vaccine (Mosquirix™): An overview. *Hum. Vaccin Immunother.* **16**, 480–489 (2020).
15. Zavala, F., Cochrane, A. H., Nardin, E. H., Nussenzweig, R. S. & Nussenzweig, V. Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. *J. Exp. Med.* **157**, 1947–1957 (1983).
16. Dieng, C. C. et al. Genetic variations of *Plasmodium falciparum* circumsporozoite protein and the impact on interactions with human immunoproteins and malaria vaccine efficacy. *Infect. Genet. Evol.* **110**, 105418 (2023).
17. Arora, N. L. & Pannu, A. K. Evaluating the effectiveness of the who's RTS,S/AS01 vaccine in the journey towards malaria eradication. *Risk Manag Healthc. Policy*. **14**, 1033–1039 (2021).
18. Didierlaurent, A. M. et al. Enhancement of adaptive immunity by the human vaccine adjuvant AS01 depends on activated dendritic cells. *J. Immunol.* **193** (4), 1920–1930 (2014).
19. Neafsey, D. E. et al. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. *N Engl. J. Med.* **373**, 2025–2037 (2015).
20. World Health Organization (31. Disease Outbreak News; Malaria in Ethiopia. October (2024). Available at: http://www.who.int/emergencies/disease-outbreak-news/item/2024-DON542 Accessed 01 Nov 2024.
21. Hofmann, N. et al. Ultra-sensitive detection of *Plasmodium falciparum* by amplification of multi-copy subtelomeric targets. *PLoS Med.* **12**, e1001788 (2015).
22. Mandefro, A. et al. Novel *Plasmodium falciparum* histidine-rich protein 2/3 repeat type in Ethiopian malaria infection: Does this affect performance of HRP2-based malaria RDT? *Malar. J.* **23**, 262 (2024).
23. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
24. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
25. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
26. Amegashie, E. A. et al. Population genetic analysis of the *Plasmodium falciparum* circumsporozoite protein in two distinct ecological regions in Ghana. *Malar. J.* **19**, 437 (2020).
27. He, Z. Q. et al. Genetic polymorphism of circumsporozoite protein of *Plasmodium falciparum* among Chinese migrant workers returning from Africa to Henan Province. *Malar. J.* **21**, 248 (2022).
28. Mohamed, N. S. et al. Genetic polymorphism of the N-terminal region in circumsporozoite surface protein of *Plasmodium falciparum* field isolates from Sudan. *Malar. J.* **18**, 1–10 (2019).

29. Mohamed, N. S. et al. Assessment of genetic diversity of *Plasmodium falciparum* circumsporozoite protein in Sudan: The RTS,S leading malaria vaccine candidate. *Malar. J.* **20**, 436 (2021).
30. Zeeshan, M. et al. Genetic variation in the *Plasmodium falciparum* circumsporozoite protein in India and its relevance to RTS, S malaria vaccine. *PLoS One*. **7**, e43430 (2012).
31. Huang, H. Y. et al. Genetic polymorphism of *Plasmodium falciparum* circumsporozoite protein on Bioko Island, Equatorial Guinea and global comparative analysis. *Malar. J.* **19**, 1–13 (2020).
32. Dundas, K., Shears, M. J., Sinnis, P. & Wright, G. J. Important extracellular interactions between *Plasmodium* sporozoites and host cells required for infection. *Trends Parasitol.* **35**, 129–139 (2019).
33. Ancsin, J. B. & Kisilevsky, R. A binding site for highly sulfated Heparan sulfate is identified in the N terminus of the circumsporozoite protein: Significance for malarial sporozoite attachment to hepatocytes. *J. Biol. Chem.* **279**, 21824–21832 (2004).
34. Bongfen, S. E. et al. The N-terminal domain of *Plasmodium falciparum* circumsporozoite protein represents a target of protective immunity. *Vaccine* **27**, 328–335 (2009).
35. Le, H. G. et al. Genetic polymorphism and natural selection of circumsporozoite surface protein in *Plasmodium falciparum* field isolates from Myanmar. *Malar. J.* **17**, 361 (2018).
36. Escalante, A. A. et al. A study of genetic diversity in the gene encoding the circumsporozoite protein (*CSP*) of *Plasmodium falciparum* from different transmission areas - XVI. Asembo Bay cohort project. *Mol. Biochem. Parasitol.* **125**, 83–90 (2002).
37. Garçon, N., Heppner, D. G. & Cohen, J. Development of RTS,S/AS02: A purified subunit-based malaria vaccine candidate formulated with a novel adjuvant. *Expert Rev. Vaccines* **2**, 231–238 (2003).
38. Collins, K. et al. Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci. Rep.* **7**, 46621 (2017).
39. Pringle, J. C. et al. High *Plasmodium falciparum* genetic diversity and Temporal stability despite control efforts in high transmission settings along the international border between Zambia and the Democratic Republic of the congo. *Malar. J.* **18**, 400 (2019).

Acknowledgements

Authors express sincere gratitude to the Medical Research Council Unit the Gambia, London School of Hygiene and Tropical Medicine, for providing access to their genomic laboratory, and PAMGEN H3Africa project for financial support. Additionally, we appreciate study participants for their voluntary participation. We extend our special thanks to Mr. Bacha Mokonnen for his contribution in this study.

Author contributions

L.G., A.M., E.O. and A.A.N. conceptualize and designed the study. A.M. collected samples. F.C., A.M., E.O. and M.K. conducted laboratory analysis. A.M.K. performed bioinformatics analysis. A.M. drafted the original manuscript. A.M.K., A.A.N., L.G. and E.O. reviewed the manuscript. All authors read and approved the final manuscript.

Funding

Addis Ababa University Post graduate studies and PAMGEN H3 Africa project (H3A/002/18) partly funded for this research. The funders had no role in the study design, data collection, reporting of this study.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board (IRB) of the Aklilu Lemma Institute of Pathobiology, Addis Ababa University (Ref No ALIPB/ IRB77/2014/22). Blood samples were collected following written informed consent and/or assent from parents or guardian for children. All methods were carried out in accordance with the institution's relevant guidelines and regulations.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-00140-0.

Correspondence and requests for materials should be addressed to A.M. or L.G.