scientific reports

Check for updates

OPEN

Genetic diversity and natural selection of *Plasmodium falciparum* Pf41 vaccine candidate in clinical isolates from Senegal

Rokhaya Sané[1,2], Babacar Souleymane Sambe[1], Aissatou Diagne[1,2], Joseph Faye[3], Fatoumata Diene Sarr[3], Serigne Ousmane Mbacké Diaw[1], Ibrahima Sarr[1,2], Arona Sabène Diatta[1], Hélène Ataume Mawounge Diatta[1,2], Papa Mbacké Sembène[2], Inès Vigan-Womas[1], Aissatou Toure-Balde[1,6], Faith Osier[4,5,6] & Makhtar Niang[1,6]✉

The merozoite surface antigen Pf41 was previously identified among the top 10 of new malaria vaccine candidates. Pf41 possesses red blood cell binding regions and conserved domains. We used population genetics approaches to determine the genetic diversity and to identify regions under balancing selection for the potential inclusion of Pf41 as candidate in a multicomponent vaccine. We screened 116 clinical isolates collected from different administrative regions in Senegal for *P. falciparum* positivity, *Pf41* amplification and sequencing. We analyzed Pf41 sequences for polymorphism, natural selection, haplotype prevalence and linkage disequilibrium. Neutrality tests (Tajima's D, FLD, FLF and MEME) were computed using DnaSP v6. and Datamonkey Hyphy. Population Analysis with Reticulate Trees (Popart) version 1.7 software was used to construct haplotypes network showing the distribution of haplotypes per study site. *P. falciparum* positivity from the 116 successfully tested samples was 93.1% of which 73 were successfully sequenced for Pf41. We found a low genetic diversity ($\pi = 0.00144 \pm 0.00022$) and high haplotype diversity ($Hd = 0.765 \pm 0.037$) of *Pf41* sequences that can be attributed to linkage disequilibrium. We identified several substitutions under positive selection and negatively selected codons at inter-species level in the central and 6-Cys domains of Pf41, respectively. The predominant SNP S232R was found fixed by positive selection in Senegalese isolates. The genetic diversity of Pf41 antigen is low in clinical isolates from Senegal. With a central domain under balancing selection and two highly conserved 6-Cys domains under negative selection due to functional constraints, the Pf41 antigen appears as a good vaccine candidate. Further monitoring of allelic variants on larger and diverse sets of samples would justify the rational for functional assays and Pf41 integration in a multicomponent vaccine.

Keywords *Plasmodium falciparum*, Pf41, 6-Cys, Genetic diversity, Natural selection, Vaccine candidate

Abbreviations

Cys	Cystein
DNA	Deoxyribonucleic acid
HABP	High activity binding peptide
LD	Linkage desequilibrium
LRT	Likelihood ratio test
MEME	Mixed effects model of evolution
MEGA	Molecular evolutionary genetics analysis
PCR	Polymerase chain reaction

[1]Pôle Immunophysiopathologie et Maladies Infectieuses, Institut Pasteur de Dakar, 36 Avenue Pasteur, 220 Dakar, Senegal. [2]Département de Biologie Animale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop de Dakar, Dakar, Fann, Sénégal. [3]Pôle Epidémiologie, Recherche Clinique et Sciences des Données, Institut Pasteur de Dakar, 36 Avenue Pasteur, 220 Dakar, Senegal. [4]Centre for Geographic Medicine Research (Coast), Kenya Medical Research Institute-Wellcome Trust Research Programme, Kilifi 80108, Kenya. [5]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK. [6]Aissatou Toure-Balde, Faith Osier and Makhtar Niang Contributed equally to this work. ✉email: makhtar.niang@pasteur.sn

qPCR	Quantitative polymerase chain reaction
Pf41	*Plasmodium falciparum* Merozoite surface protein 41
Pf41	Gene
Rm	Minimum number of recombination events
WHO	World Health Organization

Malaria due to *Plasmodium falciparum* is a parasitic disease widely endemic in tropical and subtropical areas, and responsible for clinical cases and deaths mostly in children under five years of age and pregnant women[1]. Despite integrative preventive and control efforts, an estimated of 249 million clinical cases and 608,000 deaths due to malaria occurred in 2023, with *P. falciparum* being the main causative agent[2]. Clinical malaria symptoms occur during the parasite's blood developmental stage, which is initiated when the human erythrocytes are invaded by merozoites that are released from the liver.

Vaccines targeting proteins of the merozoite involved in the invasion of the erythrocyte are conceptually attractive because merozoites are repeatedly and directly exposed to the human humoral immune system. Moreover, naturally acquired antibodies against these proteins have been shown to confer at least partial immunity[3]. A good vaccine candidate antigen should ideally have a low level of polymorphism and be highly conserved to be effective in various geographical areas and to avoid an allele-specific immune response. Antigens such as the Merozoite Surface Protein-1 and 2 (MSP-1 and MSP-2) and S antigen (heat stable protein located in the parasitophorous vacuole of the mature asexual intraerythrocytic parasite) are highly polymorphic while others such as Circumsporozoite Protein (CSP), Apical Membrane Antigen-1 (AMA-1) and Erythrocyte Binding Antigen-175 (EBA-175) possess functional constraints that limit their degree of polymorphism[4]. Most of malaria vaccines were initially designed without taking into account the genetic diversity of the candidate antigen, resulting in poor protective efficacy. For example, FMP1/AS02A MSP1 or AMA1 candidates had limited vaccine allele representation in naturally circulating parasites, especially in Africa[5]. Characterizing the genetic diversity and signatures of selection in the parasite genome or for specific antigens has been useful in identifying genes under balancing selection and genic regions that are immunogenic for further vaccine development[6]. Exploring the allele frequency spectrum and haplotype diversity of vaccine candidates can inform how natural selection and demographic events shape them across multiple populations[6].

A family of proteins containing the s48/45 domain has recently attracted attention due to their unique molecular structure[7]. This domain presents a β sandwich representation and generally contains six to fourteen conserved cysteins (6-cys)[8]. Members of the 6-cys family protein are associated with diverse and critical roles in *Plasmodium* life cycle[9]. In fact, some members reside on the surface of the infected erythrocyte during the parasite's life cycle stage and others on the surface of infected liver cells and gametocytes sexual stages, and have been shown to play important roles in hepatocyte growth and fertilization, respectively[9]. However, 6-cys proteins associated with the blood-stage forms of the parasite have no known function[10]. Among them, the merozoite's surface protein PFD0240c (Pf41) has been shown to be recognized by sera of individuals with previous clinical malaria attacks[3] and which may therefore be involved in the host immune response through its interaction with the glycophosphatidylinositol (GPI)-anchored Pf12[11]. Indeed, Pf41 is not GPI-anchored and antibodies generated to the relatively long spacer region between its two 6-cys domains indicated surface expression by fluorescent microscopy[12]. Pf41 contains 4 High Activity Binding Peptides (HABP): the first two are located in the central region of the protein outside the two Cys domains, while others HABPs are located towards the C-terminal region within the cystein-rich domain II[13]. To date, the genetic diversity of *Pf41* has not been investigated, but such information is essential since a high level of Pf41 polymorphism would comprise its use as a potential canidate vaccine that would be effective in various geographical areas and that avoid an allele-specific immune response.

This study used an in-silico analysis based on population genetics approaches to investigate the genetic diversity of the *P. falciparum* Pf41 vaccine candidate in *P. falciparum* field isolates from Senegal.

Results

Demographics of the study patients

A total of 116 patients were recruited from four study sites of which 108 (59.26% females and 40.74% males) tested positive for *P. falciparum* infections by PCR (Fig. 1 and Table 1). The 108 *P. falciparum* positive patients were distributed as follows: 35, 22, 26 and 25 from Dielmo, Kedougou, Kolda and Tambacounda respectively. Among them, 34.26% were < 5 years, 35.18% were 5–15 years, and 30.56% were > 15 years and the overall median age was 10 years (Table 1). The selection procedure for *P. falciparum* positivity, *Pf41* amplification and sequencing is summarized in Fig. 1A.

Pf41 characterization

Of the 108 *P. falciparum* positive samples, 101 were amplified for *Pf41* gene, among which 73 were successfully sequenced and selected for genetic analyses. (Fig. 1A). The latter comprised 29 samples from Central Senegal (Dielmo) and 44 from Southern Senegal (Kedougou, Kolda and Tambacounda). The complete sequence length of 1043 bp (61–1104 bp) was obtained from all 73 samples, corresponding to 21–368 aa position, including the first 6-cystein domain (21–117aa), the second 6-cystein domain (242–368 aa) and the central domain of PF41 (117–242aa), relative to the Pf3D7_0404900 reference (Fig. 1B). The full-length 1137 bp nucleotide sequence of *Pf41* was amplified by PCR using primer F (1–25 bp) and primer Rc (1117–1137 nt (Fig. 1C), and an illustration of *Pf41* amplified product is shown in Fig. 1D. Using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), the Pf41 sequence queried against known sequences had 100% identity with the *P. falciparum* Pf3D7_0404900 gene in the NCBI database.

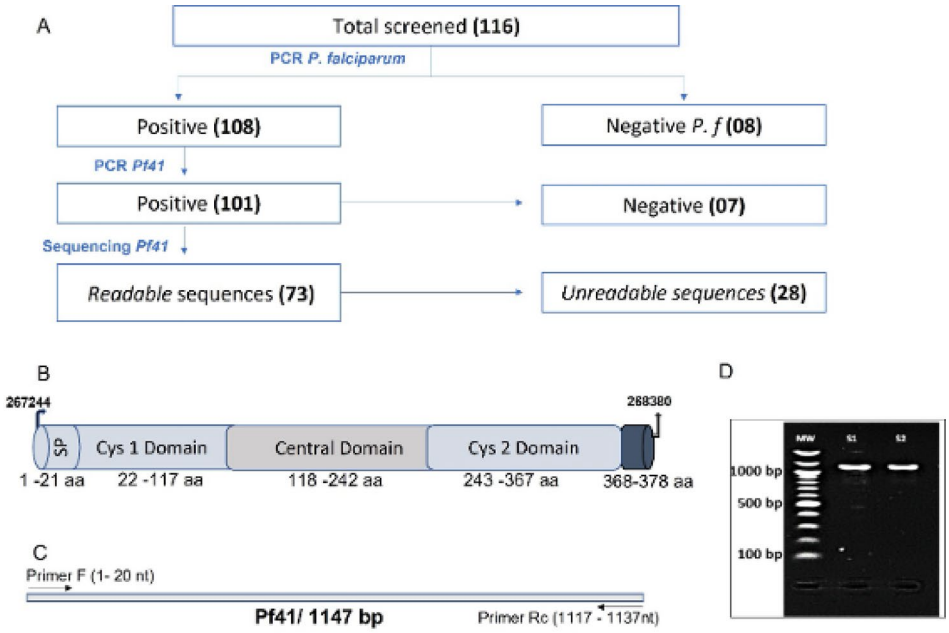

Fig. 1. Samples screening process and amplification of *Pf41* gene. (**A**), flow chart of the *P. falciparum* positive sample selection to Pf41 sequencing. (**B**) Schematic representation of *Pf41* domain organization (**C**) targeted Pf41 sequence for PCR amplification using primers F (1–20 nt) and Rc (1117–1137 nt) and (**D**) agarose gel image showing positive amplification of the 1137 bp fragment of P41 gene of *P. falciparum*. Culumns MW, S1 and S2 on the gel image indicate Molecular Weight and positive amplification of samples 1 and 2 at 1137 bp, respectively.

	Dielmo 2000	Kedougou 2020	Kolda 2020	Tambacounda 2020	Global
Ages	N = 35	N = 22	N = 26	N = 25	N = 108
Minimum	0.5	7	3	2	0.5
Maximum	11	47	83	55	83
Median	4	16	26	10	10
Mean	4.07	18.23	28.5	12.4	14.76
SD	2.71	10.19	20.89	11.87	15.6
Sex					
F	54.29	45.45	42.31	16	59.26
M	45.71	54.55	57.69	84	40.74
Sex ratio	0.84	1.2	1.36	5.25	0.69
Age group					
0–5 ans	80	0	3.85	32	34.26
5–15 ans	20	45.45	38.46	44	35.18
+ 15 ans	0	54.55	57.69	24	30.56

Table 1. Demographical profile of study population.

Genetic diversity

The alignment of the 73 *Pf41*—sequences revealed 1017 invariable sites and 22 variables sites, of which 13 were synonymous substitutions (positions: 62–71–105–382–411–813–963–1002–1059–1063–1089–1095 and 1099) and 13 were non-synonymous (positions: 71–85–99–234–694–695–1024–1061–1062–1072–1091–1092 and 1096). All the observed base substitutions on the 22 mutated loci resulted into 13 non-synonymous changes: K21R,E–S22R–K24T–K29E–L33F–K78N–S232R, H–Y342N–Q354P, H–N355H–T358P–Q364R, H and E367K (numeration based on the 3D7_Pf41 reference sequence) (Table 2). The alignment of PF41 proteins sequences from *P. falciparum* Senegalese isolates and reference strain (Pf41_3D7) revealed that cystein residues characterizing the 6-Cys proteins at positions C25; C42; C56; C64; C111; C113 (Cys-1-domain); C245; C270; C284; C297; C346 and C348 (Cys-2-domain) are well conserved (Supplementary Fig. S1).

The nucleotide diversity (Table 3) in studied Pf41 segment was very low ($\pi = 0.00144 \pm 0.00022$) with highest peaks coming between the middle intersection of the segment (612 bp–792 bp) corresponding to Pf41central

Mutated Loci	Number of characters at site	*Pf41* Domains	Amino acid changes	Syn/Non Syn
61	2	SP	K21R, E	Non Syn
62	1	SP	–	Syn
64	1	Domain I	S22R	Non Syn
71	1	Domain I	–	Syn
72	1	Domain I	K24T	Non Syn
85	1	Domain I	K29E	Non Syn
99	1	Domain I	L33F	Non Syn
105	1	Domain I	–	Syn
234	1	Domain I	K78N	Non Syn
382	1	Central domain	–	Syn
411	1	Central domain	–	Syn
694	2	Central domain	R232S	Non Syn
695	2	Central domain	R232H	Non Syn
813	1	Domain II	–	Syn
963	1	Domain II	–	Syn
1002	1	Domain II	–	Syn
1024	1	Domain II	Y342N	Non Syn
1059	1	Domain II		Syn
1061		Domain II	Q354P	Non Syn
1062	1	Domain II	N355H	Non Syn
1063	1	Domain II	–	Syn
1072	1	Domain II	T358P	Non Syn
1089	1	Domain II	–	Syn
1091	2	Domain II	Q364R	Non Syn
1092	1	Domain II	Q364H	Non Syn
1095	1	Domain II	–	Syn
1096	1	Domain II	E367K	Non Syn
1099	1	Domain II	–	Syn

Table 2. Polymorphism index of *Pf41* gene (positions, allele frequencies per mutation loci and amino acid changes).

	Genetic diversity					Neutrality tests			
	S	H	Hd	π	k	Tajima's D	Fu and Li's D	Fu and Li's F	Z-test
Domain I (S22–T117)	6	7	0.228	0.001	0.2892	− 1.8087*	− 2.5234*	− 2.576*	–
Central domain (T117–I242)	4	5	0.559	0.00164	0.6142	− 0.5267	− 1.3585	− 1.2279	–
Domain II (I242–E368)	13	11	0.343	0.00157	0.5872	− 2.1959**	− 1.2861	− 1.8171	–
Pf41 gene (3 domains)	23	22	0.765	0.00144	1.4906	− 2.0969*	− 2.3194	− 2.5443*	− 0.23 ($P = 0.82$)

Table 3. Genetic diversity indices and neutrality tests based on *Pf41* sequences. n, number of samples; range, interval covering a defined domain of *Pf41* gene; π, nucleotide diversity; S, number of polymorphic sites; k, Average number of nucleotide differences; H, number of haplotypes; Hd, haplotype diversity; Fu and Li's D; Fu and Li's F and Tajima's D, neutrality tests used to detect selection in *Pf41* sequences; Statistical significance: *$P < 0.05$; Statistical significance: **$P < 0.02$.

domain. Sliding window analysis revealed that nucleotide diversity (π) was 0.005 in this region (Fig. 2A, top panel). The analysis of each domain of *Pf41* gene given the π values is reported in Table 3. The haplotype diversity across the three domains of *Pf41* gene was high (Hd = 0.765 ± 0.037). Analysis of the Hd values for the *Pf41* gene revealed that the highest haplotype diversity was found at the central domain although the greatest number of haplotypes was found in the Cys-2 domain). The Hd value for each fragment was respectively 0.228 ± 0.00131 for Cys-1 domain, 0.559 ± 0.00053 for central domain and 0.343 ± 0.00068 for Cys-2 domain (Table 3).

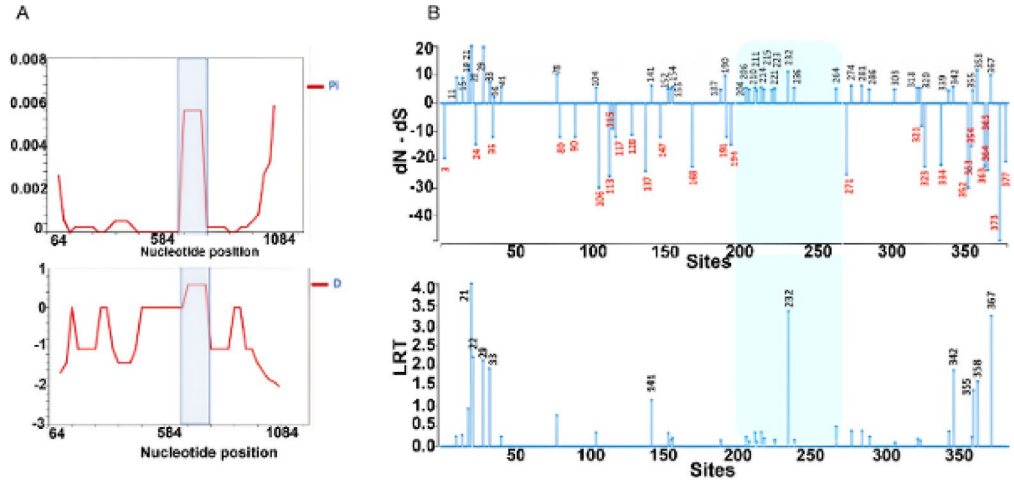

Fig. 2. Trend of polymorphism and neutrality along *Pf41* gene in Senegalese isolates. (**A**) nucleotide diversity (Pi, top panel) and Tajima's D (D, bottom panel). The curves were generated by sliding window with 100 sites and 25 sites for step size. (**B**) Single-Likelihood Ancestor Counting (SLAC) site graph to identify positive and negative codon (top panel) and Likelihood test statistic for episodic diversification; $\omega > 1$ (bottom panel).

Natural selection

All tests for neutrality by Fu and Lis' F and Tajima's D were negative across the 3 domains of *Pf41* gene. The estimated Tajima's D was − 2.0969 ($p < 0.05$) across Pf41 gene (Table 3) thus displaying purifying selection. Other neutrality tests such as Fu and Li's D* and F* gave negative values but only the FLF was statistically ($p < 0.05$) significant (Table 3). However, the graphical representation of Tajima's D across *Pf41* gene showed that the region with the highest Pi values at nucleotides 612 to 792 (second part of Pf41_central domain) had positive Tajima's D (Fig. 2A, botton panel).

MEME method estimating d_N—d_S for each site (codon) was then performed to identify whether individual codons in *Pf41* gene were under selection. In total 39 codons were found to be under positive selection and 27 codons were under negative selection (Fig. 2B, top panel). The region with signatures of balancing selection (nucleotide 612 to 792) has only codons which were under positive pressure. In the first segment of the central domain of Pf41 (351 to 582) as well as in the two 6_cysteins domains, there were both codons under positive and negative selection. The Likelihood ratio test statistic (LRT) for episodic diversification (i.e. $p > 0$) gave the significant value for codons under positive selection (Fig. 2B, bottom panel). Thus, in the Cys-1 domain, codons 21–22–29 and 33 were under positive selection while in the Cys-2-domain, codons 342–355–358 and 367 were under positive pressure. Codons 232 and 141 located in the central domain of Pf41 were the only sites subject to significant positive selection.

Haplotype distribution of Pf41 and linkage disequilibrium

The TSC network shows haplotype distribution between the sites (Fig. 3A). The analysis revealed the presence of 23 haplotypes detected based on the sequences analysis of 73 Senegalese isolates. Amino acid mutations at positions 21–22–24–29–33–78–232–342–354–355–358 and 367 of Pf41 were used to create a 3D7 reference haplotype of KSKKLKSYQNTQE. Among these haplotypes, H_1 (KSKKLK**R**YQNTQE) and H_2 (KSKKLK**H**YQNTQE) were the most widely distributed and were found in all study sites (Fig. 3B), and had the highest frequencies of 38.35%, and 31.08%, respectively.

The source of genetic variation was tested for recombination using the minimum recombination events as implemented in DnaSP. Globally, the minimum numbers of recombination events between adjacent polymorphic sites (Rm) were 5 in the total length of the *Pf41* gene (Supplementary Table S1). Recombinations were detected between nucleotides sites 71 and 85; 85 and 694; 694 and 1059; 1059 and 1072; 1072 and 1095. Position 694 was found linked with position 85 and 1059 located respectively in cys-1 and cys-2 domains, despite the nucleotide distance.

Discussion

An effective vaccine against *P. falciparum* malaria still remains a research priority despite the recent official approbation of two malaria vaccines RTS, S and R21 by WHO[14].

An antigen to be considered as a potential vaccine candidate should optimally possess a low polymorphism to guarantee its efficacity across different geographical locations and avoid allele-specific immune responses[15]. Importantly, members of the 6-Cys s48/45 family of proteins are found on the surface of *P. falciparum* in every stage, and several of these antigens have been investigated as vaccine targets[16]. Although 6-Cys domain proteins were identified more than 20 years ago, their functions at asexual blood stages are yet to be defined[17]. Pf41 has been shown to have high RBC binding activity peptides[13], making it a promising malaria vaccine candidate. In this study, we carried nucleotide diversity analysis and neutrality test population genetic analyses to investigate

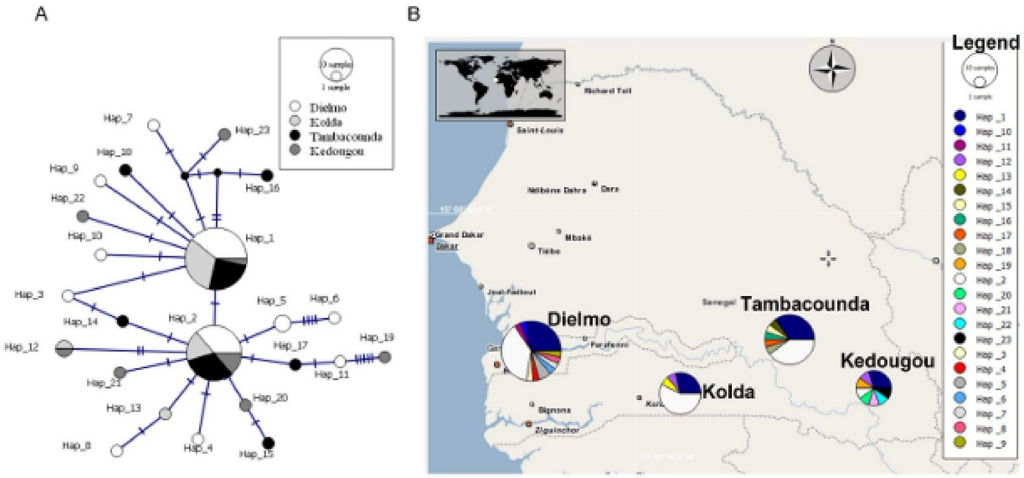

Fig. 3. TCS Haplotype network visualizing the evolutionary relationships between different haplotypes of the *Pf41* gene in 73 clinical isolates from Senegal (**A**). Each haplotype is visualized in a map interface implemented with Marble library (http://marble.kde.org) and represented by a circle, the size of which is proportional to its overall frequency in the dataset. Black lines on the branches indicate the number of mutational changes between two different haplotypes. Haplotype colors correspond to the sampling area (annotated in the legend). Geographical distribution of haplotypes in sampled sites is represented in (**B**).

the genetic diversity as well as the frequency of known haplotypes and immunogenic regions for further consideration in a vaccine selection study.

This is, to the best of our knowledge, the first investigations of the genetic diversity of Pf41 vaccine candidate in clinical isolates from patients in Senegal. We analyzed 73 Pf41 sequences representing 72.27% of the 101 Pf41-positive samples, and demonstrated a low value of nucleotide diversity in *Pf41* sequences from *P. falciparum* isolates in Senegal ($\pi = 0.00144 \pm 0.00022$), which was lower than that found for its homologous gene *Pv41* in *P. vivax* isolates from Colombia ($\pi = 0.0037 \pm 0.0006$)[18] and the blood stage vaccine candidate PfRH5 (0.0046 ± 0.0001)[19] in Nigeria. The lack of sufficient Sanger-sequenced Pf41 data is a limitation that affects our study since it precludes from a larger comparison. In addition, the limited number of samples tested and analyzed in this study, might account for the observed results.

The study has revealed a high degree of conservation of the 6-cystein motifs of *Pf41* gene in clinical isolates from Senegal. Evolutionary biology studies have shown that, amino acids that are highly conserved (i.e. their positions evolve at a slower rate than other positions) are more likely to be of biological relevance for ligand binding[20]. The presence of multiple positive selection sites in these cystein-rich regions could raise questions about how these regions can still be considered conserved. However, it has been observed that certain codons within the 6-cystein domains may be subject to positive selection, while the codons responsible for the characteristic cystein residues of this family do not exhibit such selection. Furthermore, the analysis indicates that, when considered collectively, the two 6-Cystein domains are under negative selection. The study by Kryazhimskiy et al.[21] has discussed how both positive and negative selection act on amino acid preferences in conserved protein domains, highlighting the balance between preserving essential functions and adapting to environmental pressures[22].

The protein Pf41 could be considered as a good candidate vaccine since it has a low genetic diversity, and conserved and functional regions, though this should be confirmed by functional studies.

The negative and significant Tajima's D and Fu and Li's F* values found across the Pf41 sequences from Senegal suggest a purifying or positive selection, or a population expansion that resulted in an excess of rare alleles. In this study, the high number of segregating sites, including high number of haplotypes with low frequencies in the *Pf41* gene, is indicative of a purifying selection. In spite of the overall average Tajima's D value of *Pf41* gene being negative, Tajima's D value of its central domain was positive, and was confirmed by the codon analysis which has shown that there were positively selected sites in Pf41 central domain, notably between amino acids 204 and 264 (nucleotides 612 to 792). The most frequent non-synonymous alleles (S232R and S232H) overlap with the Tajima'D and nucleotide diversity peaks for the Pf41 vaccine candidate. Regions with positive Tajima's D are evidence of balancing selection probably indicating an immune selection and the presence of putative immune epitopes[5,23]. Missense mutations that are found within these regions would potentially contribute to the generation of antigen variants. In fact, a single amino acid substitution may induce drastic structural alterations, which compromise the protein stability or perturb the binding interfaces to the point of impairing the protein function[24], these missenses may be particular.

It has been shown that the PF12 and PF41 proteins form an inverted heteroduplex on the parasite membrane[10] and that the central region of Pf41 is actually involved in the formation of this complex. Since the central domain of Pf41 appears to be important for binding to other antigens, it would be relevant to study the function of mutations detected in that domain (S232R and S232H). Furthermore, mutations at codon 232 could confer a selective advantage as they seem to be fixed in the population progressively.

Many haplotypes of Pf41 were circulating but only two (KSKKLK**R**YQNTQE and KSKKLK**H**YQNTQE) were frequent, while others were at low frequencies. Immune selection favours a low-medium frequency of distinct haplotypes and thus increased probability of newly infecting parasites that carry antigenically distinct haplotypes to those previously encountered by the host[6]. The presence of many haplotypes with low frequency, together with the linkage disequilibrium (LD) observed between codons (71 and 85); (85 and 694) (694 and 1059); (1059 and 1072); (1072 and 1095) of Pf41 gene, may suggest that an interaction exists between single nucleotide polymorphisms and might pose challenge to the efficacy of a Pf41 formulated vaccine. In fact, LD can significantly influence the efficacy of a Pf41-based vaccine by affecting genetic diversity, allele-specific immune responses, and the emergence of escape variants. For example, LD can lead to the co-inheritance of alleles that are linked together, which may include both protective and non-protective variants. Similarly, the study of the prevalence of haplotypes is relevant since it enables to tackle the neglected problem of antigenic diversity in the design of malaria vaccine by including the most prevalent haplotype(s), or a diversity-covering vaccine that includes at least one representative haplotype from each of the defined subgroups of haplotypes[6]. Larger sets of *P. falciparum* positive samples from various geographical areas are required to substantially document the extent of the genetic diversity of Pf41, using population genetics data already generated and freely available.

This study contributes to the understanding of the genetic diversity and natural selection of *Pf41* gene towards its potential inclusion in a multicomponent malaria vaccine. *Pf41* showed a limited polymorphism and conserved domains, which were under negative selection, allowing the physicochemical properties of these domains to be maintained. The study also found in the central domain a dimorphic SNP which has been fixed by positive selection and which may be immunologically relevant. These findings suggest that Pf41 protein could be a promising vaccine candidate. The extent of genetic diversity of *P. falciparum* is largely associated with transmission intensity and geographic origin suggesting that in silico populations genetic studies including sequence data obtained across multiple time period and multiple geographical areas can be used to identify the most relevant haplotypes for inclusion in a multivalent vaccine.

Material and methods

Study sites

The samples used in this study originated from two geographical areas in Senegal: the village of Dielmo located in the center-west of the country, and the three regions of Kedougou, Kolda and Tambacounda situated in the South and Southeastern parts of the country. The village of Dielmo (13° 43′ 22.2″ N, 16° 24′ 40.1″ W) has been extensively described in previous articles[25,26], and has been since June 1990, the site of a long-term investigation of host-parasite relationships and the mechanisms of protective immunity among the population. The village is located in the Sudan Savannah region in central Senegal, 280 km southeast of Dakar the capital. In the vicinity of Dielmo, the presence of a small river (Nema) which flows all year round, ensures the almost constant presence of vectors[27]. Over the years, the epidemiology of malaria has undergone significant changes, to the extent that the elimination of the disease is being considered in this area[28–30].

The regions of Kedougou (12° 56′ 05″ N, 12° 17′ 47″ W), Kolda (12° 94′ 93″ N, 14° 47′ 23″ W) and Tambacounda (13° 56′ 19″ N, 13° 17′ 40″ W) are located in southeast Senegal[31]. Malaria transmission is very active in these regions and National Malaria Control Program (NMCP) reported that the three regions accounted together for 83.3% of the 479,261 confirmed clinical malaria cases in 2023[32].

Study population and samples

In Dielmo, patients were recruited in 2000 among febrile (axillary temperature > 37.5 °C) individuals who consulted at the health post for fever-related illness as part of a routine surveillance of malaria[25]. The blood samples used in this study were withdrawn from an archived collection of frozen specimens obtained in 2000, corresponding to a period of high malaria transmission (EIR = 353.8)[29]. In the Southern regions, the study included febrile patients aged 6 months or older presenting with a febrile episode and/or one or more symptoms suggestive of malaria (headache, nausea, dizziness, chills, fatigue, …), and who consulted between October 2020 and January 2021 at the regional hospitals in Kolda and Tambacounda, and health posts and center in Kedougou. In both sites, informed consents and ascents were obtained respectively from adult patients and parents or legal guardians of underaged children.

A total of 116 patients were screened for *P. falciparum* carriage of which, 40 originated from Dielmo in 2000 and 76 from the southeast regions (Tambacounda and Kedougou, n = 25 each, and, Kolda n = 26) between October 2020 and January 2021. From each enrolled patient, venous blood was collected on an EDTA tube for molecular diagnosis of *Plasmodium* infection and the characterization of *Plasmodium* species. Finally, only specimens with singly-infected by *P. falciparum* (108 patients) were considered for further analysis. Demographic and clinical data of patients such as age, sex and disease severity were recorded in a dedicated register.

DNA extraction and *Plasmodium* species identification

Two hundred microliters of blood sample were used for the genomic DNA (gDNA) extraction of Plasmodium parasites using Qiagen DNA mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Eluted DNA samples were stored at − 20 °C until use.

The presence of *Plasmodium* parasites was first investigated by a "screening real-time quantitative PCR" (qPCR) as previously described[33] with genus-specific primers targeting the *Plasmodium cytochrome b* gene. In a second assay, a species-specific nested PCR was used on qPCR positive DNA samples to differentiate between the four major *Plasmodium* species i.e. *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale* according to the protocol described by Snounou et al.[34]. The gDNA from confirmed *P. falciparum*, *P. malariae*, *P. ovale* and *P. vivax* infected blood samples served for positive controls in all amplifications; sterile water and gDNA from uninfected

blood samples served for negative controls to ensure lack of contamination. Nested PCR results were scored as categorical variables (presence vs absence of amplification).

Amplification and sequencing of *Pf41 gene*

Only *P. falciparum* mono-infected samples (108 samples) were used for the amplification and genotyping of *Pf41* gene. The primers sequences *Pf41*-F (5′-ATGAAGGGTGTTATTTTTTG-3′) and *Pf41*-Rc (5′-TCATGAC GTCTTAAAAGAGT-3′) were designed by using the Primer 3 software and used to PCR-amplify the complete coding sequence of *Pf41*[35]. The gDNA was amplified in a total volume of 25 µL containing 12.5 µL of GoTaq Green Master Mix 2X (Promega), 1.5 µl of each of the forward and reverse primer at 0.1 µM, 7.5 µl of H2O and 2 µL of DNA template. The Thermal Cycler SimpliAmp™ (Applied Biosystems™) was used with the following PCR thermal conditions: the initial denaturation was performed at 94 °C for 1 min; followed by 25 cycles of denaturation at 94 °C for 45 s, annealing at 57 °C for 1 min and extension at 72 °C for 2 min; and with a final extension at 72 °C for 5 min. Amplicons were individually analyzed on 1% agarose gel electrophoresis in parallel to a 1 kb base pairs molecular ladder, stained with ethidium bromide and visualized under UV using the E-Box CX5 (Vilber) gel imaging. In order to confirm the sequence identity of the amplified DNA, the PCR products were Sanger-sequenced at Genewiz from both 3′ and 5′ directions (2 × coverage).

Data analysis

Genetic diversity of Pf41 sequences

Pf41 nucleotides sequences were edited and assembled using BioEdit software[36]. The 3D7 reference sequences for Pf41 (Pf3D7_0404900) was downloaded from NCBI database (https://www.ncbi.nlm.nih.gov) and used to detect single nucleotide polymorphisms (SNPs). DnaSP Version 6.12.03[37] was used to compute genetic diversity indices such as nucleotide diversity (π), mean pairwise differences (k), polymorphic sites (S), and haplotype diversity (Hd). The source of genetic variation was tested using DnaSP by measuring Linkage Disequilibrium (LD), probability of recombination between adjacent nucleotides per generation and minimum number of recombination events (Rm). Haplotype network was constructed using the Templeton, Crandall and Sing (TCS) model[38] in the PopArt program[39] while the geographical positions of sequences were displayed in the map interface implemented with Marble library (http://marble.kde.org), whose appearance can be customised using several pre-defined themes.

Natural selection

The rates of synonymous substitutions per synonymous site (d_S) and the rate of nonsynonymous substitutions per nonsynonymous site (d_N) were determined from the average values of sequence differences in all pairwise comparisons of each taxon and the standard error was computed from 1000 bootstrap pseudo-replicates implemented in the MEGA 11.0 program[40]. Tajima's D, Fu and Li's F and Fu and Li's D tests were applied to *Pf41* gene sequences using DnaSP v6.12.03 software. The pattern of immune selection was determined using the graphical representation of Tajima's D test in 100 bp sliding window with a 25 bp step-wise increase. To identify gene-wide evidence for episodic positive selection, MEME analysis was conducted using an evolutionary model based on the LRT algorithm where codon-specific positive selection was admitted at p-value < 0.05, and the inference of the lineages in which diversifying selection occurred at a given codon was performed using an empirical Bayes approach. Branches with sites in which the Bayes factor was greater than 1 were considered as targets of episodic diversifying selection. The alignment used for these analyses was submitted to MAFFT program[41] following by a manual checking in BioEdit. Nucleotide sequences were aligned based on the translated amino acid sequences to maintain the reading frame as required for diversifying selection analyses.

Data availability

Sequence data that support the findings of this study have been deposited in the Genbank repository with the primary accession number PV167071-PV167143, and accessible from the link https://www.ncbi.nlm.nih.gov/genbank/update.html or at the addresses info@ncbi.nlm.nih.gov and gb-admin@ncbi.nlm.nih.gov.

Received: 30 August 2024; Accepted: 30 April 2025
Published online: 13 May 2025

References

1. WHO. *World Malaria Report*. 372 (2021). https://www.who.int/publications/i/item/9789240064898
2. World Health Organization. *World Malaria Report 2024: Addressing Inequity in the Global Malaria Response*. (World Health Organization, 2024). https://iris.who.int/handle/10665/379751
3. Crosnier, C. et al. A library of functional recombinant cell-surface and secreted *P. falciparum* merozoite proteins. *Mol. Cell Proteom. MCP.* **12**(12), 3976–3986. https://doi.org/10.1074/mcp.O113.028357 (2013).
4. Mahajan, R. C., Farooq, U., Dubey, M. L. & Malla, N. Genetic polymorphism in *Plasmodium falciparum* vaccine candidate antigens. *Indian J Pathol Microbiol.* **48**(4), 429–438 (2005).
5. Ajibola, O. et al. In silico characterisation of putative *Plasmodium falciparum* vaccine candidates in African malaria populations. *Sci Rep.* **11**(1), 1–13 (2021).
6. Barry, A. E., Schultz, L., Buckee, C. O. & Reeder, J. C. Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, *Plasmodium falciparum*. *PLoS ONE* **4**(12), e8497. https://doi.org/10.1371/journal.pone.0008497 (2009).
7. Shen, F. et al. Genetic diversity and immunogenicity analysis of 6-cysteine protein family members in *Plasmodium ovale curtisi* importess from Africa to China: P12, P38 and P41. *Gene Rep.* **19**, 100657. https://doi.org/10.1016/j.genrep.2020.100657 (2020).
8. Arredondo, S. A. & Kappe, S. H. I. The s48/45 six-cysteine proteins: Mediators of interaction throughout the Plasmodium life cycle. *Int J Parasitol.* **47**(7), 409–423. https://doi.org/10.1016/j.ijpara.2016.10.002 (2017).
9. Dietrich, M. H. et al. Structure of the Pf12 and Pf41 heterodimeric complex of *Plasmodium falciparum* 6-cysteine proteins. *FEMS Microbes.* **3**, xtac005. https://doi.org/10.1093/femsmc/xtac005 (2022).

10. Taechalertpaisarn, T. et al. Biochemical and functional analysis of two *Plasmodium falciparum* blood-stage 6-Cys Proteins: P12 and P41. *PLoS ONE* **7**(7), e41937 (2012).
11. Parker, M. L., Peng, F. & Boulanger, M. J. The Structure of *Plasmodium falciparum* blood-stage 6-Cys protein Pf41 reveals an unexpected intra-domain insertion required for Pf12 coordination. *PLoS ONE* **10**(9), e0139407. https://doi.org/10.1371/journal.pone.0139407 (2015).
12. Sanders, P. R. et al. Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. *J. Biol. Chem.* **280**(48), 40169–40176. https://doi.org/10.1074/jbc.M509631200 (2005).
13. García, J. et al. Identification of conserved erythrocyte binding regions in members of the *Plasmodium falciparum* Cys6 lipid raft-associated protein family. *Vaccine* **27**(30), 3953–3962. https://doi.org/10.1016/j.vaccine.2009.04.039 (2009).
14. World Health Organization. *World Malaria Report 2023. Geneva. License: CC BY-NC-SA 3.0 IGO.*; 2023.
15. Ouattara, A. et al. An in silico analysis of malaria pre-erythrocytic-stage antigens interpreting worldwide genetic data to suggest vaccine candidate variants and epitopes. *Microorganisms* **10**(6), 1090 (2022).
16. Tonkin, M. L. et al. Structural and biochemical characterization of *Plasmodium falciparum* 12 (Pf12) reveals a unique interdomain organization and the potential for an antiparallel arrangement with Pf41*. *J. Biol. Chem.* **288**(18), 12805–12817. https://doi.org/10.1074/jbc.M113.455667 (2013).
17. Paul, G. et al. A novel Pfs38 protein complex on the surface of *Plasmodium falciparum* blood-stage merozoites. *Malar. J.* **16**(1), 79. https://doi.org/10.1186/s12936-017-1716-0 (2017).
18. Forero-Rodríguez, J., Garzón-Ospina, D. & Patarroyo, M. A. Low genetic diversity in the locus encoding the Plasmodium vivax P41 protein in Colombia's parasite population. *Malar. J.* **13**(1), 388. https://doi.org/10.1186/1475-2875-13-388 (2014).
19. Ajibaye, O. et al. Genetic polymorphisms in malaria vaccine candidate *Plasmodium falciparum* reticulocyte-binding protein homologue-5 among populations in Lagos, Nigeria. *Malar J.* **19**(1), 6. https://doi.org/10.1186/s12936-019-3096-0 (2020).
20. Ben Chorin, A. et al. ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci.* **29**(1), 258–267. https://doi.org/10.1002/pro.3779 (2020).
21. Kryazhimskiy, S., Tkačik, G. & Plotkin, J. B. The dynamics of adaptation on correlated fitness landscapes. *Proc. Natl. Acad. Sci.* **106**(44), 18638–18643 (2009).
22. Moses, A. M. & Durbin, R. Inferring selection on amino acid preference in protein domains. *Mol. Biol. Evol.* **26**(3), 527–536. https://doi.org/10.1093/molbev/msn286 (2009).
23. Xu, S. J. et al. Genetic diversity and natural selection of rif gene (PF3D7_1254800) in the *Plasmodium falciparum* global populations. *Mol. Biochem. Parasitol.* **254**, 111558. https://doi.org/10.1016/j.molbiopara.2023.111558 (2023).
24. Ancien, F., Pucci, F., Godfroid, M. & Rooman, M. Prediction and interpretation of deleterious coding variants in terms of protein structural stability. *Sci. Rep.* **8**(1), 4480. https://doi.org/10.1038/s41598-018-22531-2 (2018).
25. Trape, J. et al. The dielmo project: A longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community living in a holoendemic area of senegal. *Am. J. Trop. Med. Hyg.* **51**(2), 123–137. https://doi.org/10.4269/ajtmh.1994.51.123 (1994).
26. Rogier, C., Cissé, B., Ly, A. B., Trape, J. F. & Tall, A. *Plasmodium falciparum* clinical malaria in Dielmo, a holoendemic area in Senegal: No influence of acquired immunity on initial symptomatology and severity of malaria attacks. *Am. J. Trop. Med. Hyg.* **60**(3), 410–420. https://doi.org/10.4269/ajtmh.1999.60.410 (1999).
27. Fontenille, D. et al. High annual and seasonal variations in malaria transmission by anophelines and vector species composition in dielmo, a holoendemic area in senegal. *Am. J. Trop. Med. Hyg.* **56**(3), 247–253. https://doi.org/10.4269/ajtmh.1997.56.247 (1997).
28. Wotodjo, A. et al. Malaria in Dielmo, a Senegal village: Is its elimination possible after seven years of implementation of long-lasting insecticide-treated nets? Sirima S, ed. *PLoS ONE* **12**(7), e0179528. https://doi.org/10.1371/journal.pone.0179528 (2017).
29. Trape, J. et al. The rise and fall of malaria in a west African rural community, Dielmo, Senegal, from 1990 to 2012: a 22 year longitudinal study. *Lancet Infect Dis.* **14**(6), 476–488. https://doi.org/10.1016/S1473-3099(14)70712-1 (2014).
30. Niang, M. et al. Temporal analysis of IgG antibody responses to *Plasmodium falciparum* antigens in relation to changing malaria epidemiology in a West African setting. *Malar. J.* **16**(1), 283. https://doi.org/10.1186/s12936-017-1928-3 (2017).
31. ANSD. *Agence National de La Statistique et de La Démographie.*; 2022. https://www.ansd.sn/
32. PNL. *PLAN STRATEGIQUE NATIONAL DE LUTTE CONTRE LE PALUDISME AU SENEGAL 2021–2025.*; 2020:116. https://senegal-cocreation.com/wp-content/uploads/2021/02/PSN_PNLP_Senegal_Version-finale_-Fevrier-2021.pdf
33. Niang, M. et al. Fine-scale spatiotemporal mapping of asymptomatic and clinical *Plasmodium falciparum* infections: Epidemiological evidence for targeted malaria elimination interventions. *Clin. Infect. Dis.* **73**(12), 2175–2183. https://doi.org/10.1093/cid/ciab161 (2021).
34. Snounou, G. et al. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol. Biochem. Parasitol.* **61**(2), 315–320. https://doi.org/10.1016/0166-6851(93)90077-B (1993).
35. Untergasser, A. et al. Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**(15), e115. https://doi.org/10.1093/nar/gks596 (2012).
36. Hall, T. A. BioEdit software, version 5.0. 9. *N C State Univ Raleigh NC.* Published online 1999.
37. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**(12), 3299–3302. https://doi.org/10.1093/molbev/msx248 (2017).
38. Clement, M., Posada, D. & Crandall, K. A. TCS: A computer program to estimate gene genealogies. *Mol. Ecol.* **9**(10), 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x (2000).
39. Leigh, J. W. & Bryant, D. popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410 (2015).
40. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**(7), 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).
41. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**(4), 772–780. https://doi.org/10.1093/molbev/mst010 (2013).

Acknowledgements

We would like to express our deepest gratitude to the villagers of Dielmo for their participation to the study. We also thank all the patients who participated in the study and the recruitment teams at the healthcare facilities in Kedougou, Kolda and Tambacounda regions.

Author contributions

RS, ATB, FO and MN conceived and coordinated the study with the contribution of BSS, PMS and IVW. BSS, AD, JF, FDS and MN performed the field recruitment and samples collection. AD, SOMD, IS, ASD, and HAMD performed molecular biology assays . AD, OS, HAMD RS, MN, and BSS manage the database and analyzed the data. RS BSS, and MN drafted and revised the manuscript. All authors read the manuscript and approved the final version.

Funding
Fundimg was provided by Institut Pasteur de Dakar and European and Developing Countries Clinical Trials Partnership.

Declarations

Competing interests
The authors declare no competing interests.

Ethics statement
This study was reviewed and approved by the National Ethical Committee for Research in Health Senegal under the reference 00000126/MSAS/CNERS/SP, and all methods were performed in accordance with the relevant guidelines and regualtions. Written informed consent was obtained for all adults participants, while ascent was obtained from parents or legal guardians of underaged children.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-00784-y.

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.