

Assessment of Genetic Diversity of Plasmodium Falciparum Circumsporozoite Protein in Sudan; The RTS,S Leading Malaria Vaccine Candidate

Nouh Saad Mohamed (✉ nouh.s.mahgoub@nileuniversity-edu.com)

Nile University

Hanadi AbdElbagi

Nile University

Ahad R. Elsadig

National University

Abdalla Elssir Ahmed

Nile University

Yassir Osman Mohammed

Tropical Medicine Research Institute

Lubna Taj Elssir

Tropical Medicine Research Institute

Mohammed Ahmed B Elnour

Tropical Medicine Research Institute

Yousif Ali

Federal Ministry of Health

Mohamed S. Ali

El Neelain University

Omnia Fateh Alrahman

Tropical Medicine Research Institute

Emmanuel Edwar Siddig

Nile University

Ayman Ahmed

Institute of Endemic Diseases

Rihab Ali Omer

University of Leipzig

Research Article

Keywords: Malaria, Plasmodium falciparum circumsporozoite protein (PfCSP), Vaccine, RTS,S, Hapd, N-terminal, central repeats, and C-terminal regions

Posted Date: June 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-619335/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The currently used malaria vaccine; the RTS,S, is designed based on the *Plasmodium falciparum* circumsporozoite protein (PfCSP). The PfCSP gene, besides having different polymorphic patterns, can vary between *P. falciparum* isolates due to geographical origin and host immune response. Such aspects are essential when considering the deployment of the RTS,S vaccine in a certain region. Therefore, in this study we assessed the genetic diversity of *P. falciparum* in Sudan based on the PfCSP gene by investigating the diversity at the N-terminal, central repeat, and the C-terminal regions. The results of the N-terminal region showed the presence of 2 different haplotypes with a haplotype diversity (Hapd) of 0.425 ± 0.00727 . The presence of the unique insertion of NNNGDNGREGKDEDKRDGNN was reported. The KLKQP motif was conserved in all the studied isolates. At the central repeat region, 11 haplotypes were seen with a Hapd of 0.779 ± 0.00097 . The analysis of the genetic diversity in the C-terminal region showed the presence of 10 haplotypes with a Hapd of 0.457 ± 0.073 . In this study, the results indicated a high conservation at the PfCSP gene. This may further contribute in understanding the genetic polymorphisms of *P. falciparum* prior to the deployment of the RTS,S vaccine in Sudan.

Introduction

Malaria continues to be a major global health problem, particularly in Africa where *Plasmodium falciparum* is the most prevalent species. In 2019, the WHO estimated about 229 million cases and 409 thousand deaths worldwide, with 95% of malaria occurrence to be reported from Africa [1]. In Sudan, malaria is a serious public health problem and over 75% of the country population at risk of the disease [1]. The global technical strategy for malaria 2016–2030 aims to achieve 90% reduction in malaria incidences by 2030 worldwide [2]. The systematic fight against malaria consist of enhanced surveillance and monitoring, vector control, case-detection and treatment in cooperation with improved diagnosis and treatment capacity [3] yet there are several rapidly developing challenges for this approach mainly including the emerging resistance for drugs and insecticides, and the limited resources and diagnostic capacity in the endemic countries. However, effective vaccine will offer a powerful alternative that might make a breakthrough in this fight [2].

Malaria vaccine has been recommended as one of the pillars for the successful malaria eliminate [2]. In the last two decades, a newly designed vaccine to encounter *P. falciparum* malaria infection was developed based on the *P. falciparum* circumsporozoite protein gene (PfCSP) which secreted by the sporozoite stage [4]. The PfCSP gene is known to be subdivided into 3 regions namely the N-terminal region, the central repeats region, and the C-terminal region. The N-terminal region is located at the 5' end of the PfCSP gene, comprised of around 84 to 104 amino acids (aa) [5]. The N-terminal region is also characterized by the presence of the KLKQP motif which is essential for the parasite entry to the hepatocytes. This along with the Th1R T-cell epitope which induces the antibody response in the human host [6]. For the C-terminal region which present at the 3' end, it encompasses the polymorphic Th2R and Th3R T-cell sub-regions which are also been recognized by the human immune system [5]. Concerning the central repeats region, it is consisted of multiple motifs repeats of 4 aa. The NANP and the NVDP motifs

were mostly seen among the different *P. falciparum* strains with NANP more frequent [7]. However, several other motifs were also reported from different geographical regions [8].

The currently used malaria vaccine, the RTS,S, is a virus-like elements of hepatitis B virus surface antigen and an adjuvant liposome anchored to portion of the PfCSP [9]. The deployment of the RTS,S vaccine took place in a few African countries including Ghana, Malawi, and Kenya[10]. The vaccine showed variable immunization duration course that might last for up to 1 year; hence a booster dose is recommended especially in areas of high malaria multiplicity and intensity [11].

In Sudan malaria is serious public health problem and under certain circumstances it develops to an epidemic level [12]. The national malaria control program is challenged by the emergence and rapid development of both, drugs and insecticide resistance, urging the need for alternative tools for the disease control such as an effective vaccine [13, 14]. It occurs with variable severity onset and syndrome throughout the year, however, in some parts of the country malaria incidence is sporadic signifying the genetic diversity of the circulating *P. falciparum* and consequently promote the need to identify and characterize parasite population that's circulate in each specific region [13]. Therefore, this study is aimed at assessing the genetic diversity and population structure of the *P. falciparum* parasite in Khartoum state, Sudan, based on the PfCSP gene in order to assess congruency with different PfCSP reference strains together with different Sudanese isolates [15]. Moreover such data help in assess opportunity for the RTS,S vaccine successful deployment and evaluate effectiveness of the vaccine and aid in developing Sudan RTS,S vaccine [10].

Results

Genotyping and assessment of *P. falciparum* multiple clonal infections

The study included a total of 261 febrile patients consisted of 143 (54.8%) male and 118 (45.2%) females. All study participants have reported fever > 37.5 C. Based on the WHO classification, all patients were classified as mild uncomplicated malaria infection. The microscopic examinations revealed that a total 196 (75.1%) patients were positive for *P. falciparum* infection whiles the remaining 65 (24.9%) were harboring co-infection of *P. falciparum*/*P. vivax*. Molecular analysis using the semi-nested PCR confirmed the presence of 23.9% (47/196) *P. falciparum*/*P. vivax* co-infections. Confirmed cases of mono-infections of *P. falciparum* were further amplified for the msp1 gene to exclude multiple allelic infections. Out of the 149 *P. falciparum* mono-infections, the msp1 revealed the presence of 83 (55.7%) multiple allelic infection which were excluded from the amplification process of the PfCSP gene. The selected 66 (44.3%) *P. falciparum* single allelic, mono-infections isolates were amplified for the PfCSP gene. PCR amplification produced a product sizes that vary in length from 1 kb to 1.2 kb.

Phylogenetic network of the Sudanese PfCSP and PfCSP reference sequences

In comparison with the different PfCSP reference strains, the Sudanese isolates showed high similarity with the African isolates. The phylogenetic analysis showed that most of the sequences clustered

together forming a major 2 different clades consisting of the majority of the study isolates. the first clade consisted of 24 isolates of them 18 isolates were clustered together with a bootstrap value of 100% similarity and the other 6 isolates clustered together with a bootstrap value of 63. For the second major clade, 18 isolates clustered together with a bootstrap value of 100% and showing slight similarity with the Wellcome and the HB3 reference strains. The remaining isolates were distributed forming 2 different smaller clades with a high similarity with the 3D7, MAD20, RO33, Dd2, 7G8, and the NF54 reference strains (Fig. 1).

Genetic diversity in the N-terminal region of the Sudanese and global PfCSP

The analysis of the N-terminal region showed the presence of 2 different haplotypes; Hap01 included the 3D7 reference strain plus 19 isolates, and Hap02 included 47 isolates. The haplotype diversity score was 0.425 ± 0.00727 . The low haplotype diversity among the Sudanese isolates indicates that the N-terminal region is highly conserved. The 2 haplotypes were observed due to the amino acid polymorphism at A98G. The amino acid insertion "61NNGDNGREGKDEDKRDGNN80" was detected in a total of 47 isolates representing Hap02. Whereas the KLKQP motif was conserved in all the isolates (Fig. 2).

Genetic diversity in the central repeat region of the Sudanese PfCSP

Based on the NANP and NVDP repeats of the central repeat region, a total of 11 haplotypes were seen with a Hapd of 0.779 ± 0.00097 . Hap01 and Hap02 were the most frequent haplotypes; 24 isolates and 18 isolates, respectively, followed by haplotypes Hap03, Hap05, and Hap06 which consisted of 9, 7, and 2 isolates, correspondingly. Whereas the remaining haplotypes consisted of a single isolate per each. The numbers of NANP repeats at each haplotype were ranging from 33 to 37 repeats whereas NVDP repeats were ranging from 3 to 6 repeats. Accordingly, the total number of repeats per haplotype was ranging from 37 to 40 repeats (Fig. 3).

Diversity patterns in the C-terminal region of the Sudanese PfCSP

The analysis of the genetic diversity in the C-terminal region of the Sudanese PfCSP gene showed the presence of 10 different haplotypes with a haplotype diversity of 0.457 ± 0.073 . Hap01 was the most frequent haplotype; 49 isolates. Hap02 was reported among 5 isolates, Hap07 was reported among 6 isolates. the remaining haplotypes were only represented by one isolate each (Fig. 4). Non-synonymous substitutions were seen in different amino acids in the C-terminal region. D294N, K298N, and N301S, were seen in Hap03, Hap07, and all the haplotypes except for Hap01 and Hap06, respectively. Several non-synonymous amino acids changes were also seen at the Th2R and the Th3R T-cell epitope regions including T317K, E317K, Q318E, K321N, I322K, T322K, R322K, K324Q, I327L, G352N, S354P, R355K, N356D, Q357E, and E361A (Fig. 4). The analysis of the C-terminal nucleotide diversity was 0.00929 ± 0.00150 , and an average number of nucleotide differences of 3.51199 was observed. Neutrality testing indices to investigate the null hypothesis of natural selection using Fu and Li's D^* , Fu and Li's F^* , and Tajima's D tests were -0.40167 , -0.45855 , and -0.35629 . All tests were statistically not significant, $P > 0.10$.

Haplotype network analysis and diversity patterns in the C-terminal region of the Sudanese and global PfCSP

Based on the C-terminal region, the analysis of the haplotype network of the Sudanese isolates and the global published C-terminal region of the PfCSP the number of haplotypes detected was 69 different haplotypes with a haplotype diversity of 0.8651 ± 0.008 . Hap11 was the most frequent haplotype, followed by Hap43, Hap12. The haplotype that includes the 3D7 reference strain was considered Hap01. Hap01 was including most of the studied Sudanese isolates; 48 isolates out of 66 isolates. The remaining isolates were distributed into Hap02 (5 isolates), Hap03 (1), Hap04 (1), Hap05 (1), Hap06 (1), Hap07 (6), Hap08 (1) Hap09 (1), and Hap10 (1) (Fig. 5).

Discussion

In Khartoum State, *P. falciparum* is accounted for up to 90% of malaria infections, while 10% are acknowledged to be caused by *P. vivax*. Investigating the genetic diversity of the Sudanese *P. falciparum* is extremely important to further warrant the deployment of the RTS,S malaria vaccine [15]. The aim of this study was to investigate the genetic diversity of PfCSP gene in Sudan. In this study, the investigated PfCSP gene showed a high conservation at the N-terminal region when compared to the 3D7 reference strain. Although, a higher conservation is reported in this study in which only 2 haplotypes were detected compared to previous studies which could be attributed to the difference in location from which samples were isolated and also the sample size included [15]. Also, the presence of the other unique polymorphisms can be certain to each specific location depending on the circulating *P. falciparum* strain. The (NNGDNGREGKDEDKRDGNN) insertion which was not reported previously could be mainly due to the small sample size been investigated. Meanwhile, this insertion has been detected in several other regions such as Myanmar [7]. Although, the functional role of this insertion when it's present or absent is limited investigated and not fully understood yet, it's presence could be explained in terms of different parasite strains, geographical or environmental adaptation of the parasite, or it might be present as a form of parasite adaptation to the host immune response which is mainly targeted at the Th1R T-cell epitope present at the N-terminal region [6]. Accordingly, the Th1R T-cell epitope at the N-terminal region is showing relatively high polymorphisms, this can be explained as a role of parasite adaptation to host immune system. This was previously investigated by Conway et al when a monoclonal antibody anchored to the linear T-cell epitope "81EDNEKLRKPKH91" which is located before the KLKQP motif which is responsible for sporozoite entry to the hepatocytes [16]. The monoclonal antibody effectively neutralized the infectivity of the sporozoite in-vivo [16]. Also, when Bongfen et al., demonstrate the development of polypeptides that flank the N-terminal region played a significant role in the initiation of inhibitory antibodies production to prevent sporozoite invasion to the hepatocytes [17]. Thus, supporting the hypothesis that the N-terminal region can play an important role as a vaccine candidate which is aided by the high conservation rate [16].

This study has detected a high haplotype diversity at the central repeat region, this in turn has highlighted the genetic diversity of this region is mainly sustained by balancing natural selection due to the effect of

human immune response [16]. Interestingly, as the difference in the NANP and NVDP repeats controls the difference in the PfCSP gene length, the numbers of NANP repeats detected among this studied isolates were lower than previously reported by Le et al., et al. [7] and Zeeshan et al. [8] where higher numbers of NANP repeats was detected as well as several unique repeat motifs including NANS, NTNP, NVVP, NAHP, NAKP, NVNP, NAIP, NANL, NPNP, NVAD, NADP, SANP, and KANP. However, this variation in the central repeat region requires more research to understand the effect of these variation in terms of parasite genetic signature and host immune response. Previously, the central repeat region has been found to affect the stability of the PfCSP gene, which is suggested to be a copy-number dependent and further growths by the increase in the number of repeats [18]. The Sudanese isolates showed a relatively low numbers of repeats that ranged from 37 to 40 repeats similar to the African and South American PfCSP; 36 to 37 repeats, while considerably lower than the Asian PfCSP; 40 to 43 repeats [7]. This indicates that the difference in the numbers of the tetrapeptide repeats at the central repeat region could be attributed to geographical distribution, host (human and/or mosquito vectors) population genetic differences, or the different selection pressure imposed by the different case management protocols/drugs in use that resulted in form of co-evolution response to maintain parasite stability at specific origin. Also, the central repeat region is important in the RTS,S malaria vaccine. The RTS,S consisted of 19 tetrapeptide repeats of the NANP repeat [19]. However, still no reports indicated that the different repeats number can or may affect the effectiveness of the RTS,S vaccine.

Investigating the genetic diversity of the C-terminal region of the Sudanese PfCSP gene showed a relatively low genetic diversity expressed by the presence of only 10 haplotypes, with Hap01 to be the most frequent; 49 isolates, and it has two insertions; the “KHIEKEYLNKIQNSL” in the Th2R and “NKPKDEL DYANDI” in the Th3R regions T-cell epitopes. This haplotype has been also prevalent among the African PfCSP isolates, this is further explained by the haplotype network where most of the Sudanese isolates clustered with the Ghanaian isolates. The low haplotype diversity detected among the Sudanese PfCSP isolates indicates a low level of genetic diversity. And, since the RTS,S malaria vaccine is composed of the C-terminal T-cell epitope, this region can be very important in terms of designing regional specific vaccines based on identifying the circulating *P. falciparum* strains in that specific region [19].

The analysis of the global C-terminal region haplotype network indicated a unique clustering of the Asian isolates in a relatively separated haplotypes from the African and the South American isolates. This has been also discussed previously by Le et al., where African isolates presented with branched and complex patterns of haplotypes [7]. This also supports the suggestion that African PfCSP is more diverse than the PfCSP of the other geographical regions.

Since the deployment of the RTS,S malaria vaccine is mainly controlled by understanding and identifying the genetic diversity of the circulating *P. falciparum* strains in a certain geographical region [10]. RTS,S vaccine trials showed a significant effect in reducing malaria incidence in many African countries including Gambia, Mozambique, Ghana, Tanzania, Gabon, and Kenya [20–22]. Therefore, this study findings are showing the higher prevalence of a haplotype which is more similar to the 3D7 reference

strain and also worth mentioning that the current RTS,S is primarily constructed from the PfCSP of the 3D7 strain, the deployment of the RTS,S malaria vaccine clinical trials can be effective in reducing malaria incidence in Sudan.

In conclusion, this study is the first study to investigate the genetic diversity of *P. falciparum* in Sudan based on studying the complete PfCSP gene. This study was limited to investigating the genetic diversity in specific region in Sudan; Khartoum. Therefore, further studies are warranted to explore the genetic diversity of the PfCSP in the whole country since Sudan is known with its large geographical space that has different environmental conditions and endemicity situations that might influence the development of different adaptive evolutionary traits among the parasite population increasing its genetic diversity. The combination of these factors might impact the effectiveness of the RTS,S vaccine if been deployed in Sudan or areas with similar scenario, that why a more comprehensive study with a larger sample size that collected from different regions is highly recommended. The PfCSP gene is considered a core element of the RTS,S malaria vaccine that currently been used, nevertheless, the genetic diversity in the PfCSP gene among the different regions *P. falciparum* strains may affect the efficacy of the RTS,S vaccine. In this study, the analysis of the genetic diversity of the PfCSP population structure indicated that the PfCSP gene in Sudan is highly conserved with few genetic variations reported at the central repeat region and the C-terminal region. Meanwhile, the limited diversity reported at the N-terminal region postulates the usefulness of this region in the development of new vaccine candidate.

Methods

Study sites and *P. falciparum* isolates

P. falciparum isolates were collected from different health centers in Khartoum State between January and December 2019. During the study period, a total of 261 febrile patients were recruited. Malaria diagnosis was made by expert microscopists using Giemsa stained thick and thin blood films. Prior to samples collection, patients or legal guardians of children less than 18 years old were offered a detailed description of the study, and if they have voluntary accept to enroll in our study, they were requested to sign an inform consent to include their samples and information in the study. Demographic and clinical data were collected. The Clinical phenotyping of malaria infection were made according to the WHO guidelines [23]. The collected blood samples were preserved at 4 °C until molecular examination. All methods followed for DNA extraction, Molecular detection and clonal infection determination of *Plasmodium* species, amplification and sequences analysis of the PfCSP gene were performed in accordance with the relevant guidelines and regulations.

Molecular detection of *Plasmodium* species

The total genomic DNA of the parasite was extracted from patients' blood samples using the QIAamp DNA Blood Mini Kit (Qiagen Inc. Germany). In order to confirm the microscopic results, all DNA samples were examined by the semi nested polymerase chain reaction (PCR) using the following primers, UNR: 5' GAC GGT ATC TGA TCG TCT T 3', PLF: 5' AGT GTG TAT CAA TCG AGT TT 3', and HUF: 5' GAG CCG CCT

GGA TAC CG 3' for the first PCR reaction and PLF primer and falciparum-R: 5' AGT TCC CCT AGA ATA GTT ACA 3', vivax-R: 5' AGG ACT TCC AAG CCG AAG 3', ovale-R: 5' GCA TAA GGA ATG CAA AGA ACA G 3', and malariae-R: 5' GCC CTC CAA TTG CCT TCT 3', for the second PCR reaction to confirm the presence of *P. falciparum* mono-infection. PCR reaction was made using i-Taq maxime PCR premix kit (iNtRON Biotechnology, South Korea) according to the manufacturer instructions. Primers concentrations and thermocycling conditions were adjusted according to Rubio et al., 1999 [24].

Molecular assessment of *P. falciparum* multiple clonal infections

In order to identify single clonal infection of the confirmed *P. falciparum* cases, the Merozoite Surface Protein 1 (*msp1*) gene was amplified to exclude multiclonal infections. Nested PCR was done using the following primers, outer *msp1*-f: 5' CTA GAA GCT TTA GAA GAT GCA GTA TTG '3 and outer *msp1*-r: 5' CTT AAA TAA TAT TCT AAT TCA AGT GGA TCA '3. Following the first outer reaction clonal primers were used including MAD20-f: 5' AAA TGA AGA AGA AAT TAC TAC AAA AGG TGC '3, MAD20-r: 5' GCT TGC ATC AGC TGG AGG GCT TGC ACC AGA '3, RO33-f: 5' TAA AGG ATG GAG CAA ATA CTC AAG TTG TTG '3, RO33-r: 5' CAT CTG AAG GAT TTG CAG CAC CTG GAG ATC '3, K1-f: 5' AAA TGA AGG AAC AAG TGG AAC AGC TGT TAC '3, and K1-r: 5' ATC TGA AGG ATT TGT ACG TCT TGA ATT ACC '3. The primers concentrations per each reaction and the PCR amplification conditions used were according to Ntoumi et al. [25].

Amplification of the PfCSP gene

After the amplification of the *msp1* gene, *P. falciparum* were sorted out to exclude multiclonal infection. Genomic DNA of single clonal infections were further used for the amplification of the PfCSP gene. PfCSP gene amplification was done as previously published by Zeeshan et al. [8], using the following primers, outer PfCSP-f: 5' TTA GCT ATT TTA TCT GTT TCT TC '3 and outer PfCSP-r: 5' TAA GGA ACA AGA AGG ATA ATA CC '3, for the first PCR reaction. The second PCR reaction was made using the primers, nested nPfCSP-f: 5' GAA ATG AAT TAT TAT GGG AAA CAG '3 and nPfCSP-r: 5' GAA GGA TAA TAC CAT TAT TAA TCC '3. PCR amplicons of the first run were used as a template for the second PCR. The amplified PfCSP products were visualized using agarose gel electrophoresis in comparison to a 100 base pair DNA ladder. Amplified PfCSP PCR products were further purified from gel using gel extraction kits and instructions were followed according to the manufacturer's guidelines (New England Biolabs Inc, New England). Each amplicon was sent in replicate for sequencing by the Sanger dideoxy sequencing method using the primer SeqPfCSP: 5' TGG GTC ATT TGG CAT ATT GTG '3 using ABI3500 sequencer (Applied Biosystems SeqStudio, 3500 series) provided by Macrogen Inc. (The Netherlands).

PfCSP Sequences analysis

The obtained sequences of the PfCSP amplicons were checked for the correctness of sequences reads and nucleotides base calling errors by aligning of the forward and the complementary reverse sequence using GENTle software (v1.9.4) [26]. Sequences were also trimmed at the primers sequences reads to reduce sequencing mismatching that occurs at the beginning of the sequencing process. Using the NCBI

GenBank database, all worldwide previously published PfCSP sequences were included in the analysis of the comparison with the Sudanese sequences similarity and divergence (<https://www.ncbi.nlm.nih.gov/nuccore>). Local Sudanese sequences comparison and divergence were made on the N-terminal, C-terminal and the NANP repeat regions of the PfCSP gene. Reference sequences of the PfCSP gene including the 3D7 (XM_001351086) [27], NF54 (M22982.1) [28], 7G8 (AB121015.1), HB3 (AB121018.1), Dd2 (AB121017.1), MAD20 (AB121020.1), and RO33 (AB121021.1) [29], and the Wellcome strain (M15505.1) [30]. Using MEGA7 software, the nucleotide substitution model with the least Bayesian Information Criterion (BIC) score was considered as a best fit model to conduct the phylogenetic tree [31]. Hasegawa-Kishino-Yano with Gamma distribution model (HKY+G+I) was used for creating the tree [32]. Using the software DnaSP (v5.10) [33], nucleotides diversity parameters including the numbers of haplotypes (Hap), haplotypes diversity (Hapd), the average number of nucleotide differences between two sequences (p), and the numbers of segregating sites (S) were obtained. The haplotype networks were constructed using the online tcsBU software [34]. Natural selection theory was also tested on the PfCSP using Fuand Li's D, Fuand Li's F statistics, and Tajima's D test using DnaSP [35, 36].

Declarations

Ethics approval and consent to participate

Written consents were obtained prior to data and samples collection from all the study participants following being well informed about the study and potential risk and volunteering to participate, while written consent was taken from parents or legal guardians in case of children. Ethical approval for conducting the study was obtained from the National University Research Ethics Committee, National University - Sudan (NU-REC13).

Data Availability

All datasets generated for this study are included in the manuscript. Sequences generated in this study are available at the NCBI GenBank database under the accession numbers, MW554526-MW554591.

Acknowledgments

We are grateful to all individuals who agreed to participate in this study. NSM recipient of a partial fund from the Royal Society of Tropical Medicine and Hygiene to conduct the study (RSTMH Small Grant round 2019).

Author Contributions

NSM, HA, and AA: Study designing. HA, AEA, ARE, MSA, and EES: Performed experiments. NSM, LTE, OFA, YOM, AA, and RAO: Data analysis and manuscript preparation: Field work support: NSM, HA, AM, LTE, YA, and MABE. Manuscript review and data analysis: NSM, EES, AA, and RAO.

Additional Information

Competing interests

The authors declare no competing interests.

References

1. Organization WH: **World malaria report 2020: 20 years of global progress and challenges**. 2020.
2. Organization WH: **Global technical strategy for malaria 2016-2030**: World Health Organization, 2015.
3. Shretta R, Liu J, Cotter C, Cohen J, Dolenz C, Makomva K, Newby G, Ménard D, Phillips A, Tatarsky A: **Malaria elimination and eradication**. 2017.
4. Stoute JA, Slaoui M, Heppner DG, Momin P, Kester KE, Desmons P, Wellde BT, Garçon N, Krzych U, Marchand M: **A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria**. *New England Journal of Medicine* 1997, **336**(2):86-91.
5. Calvo-Calle JM, De Oliveira G, Clavijo P, Maracic M, Tam J, Lu Y, Nardin E, Nussenzweig R, Cochrane A: **Immunogenicity of multiple antigen peptides containing B and non-repeat T cell epitopes of the circumsporozoite protein of *Plasmodium falciparum***. *The Journal of Immunology* 1993, **150**(4):1403-1412.
6. Enea V, Ellis J, Zavala F, Arnot DE, Asavanich A, Masuda A, Quakyi I, Nussenzweig RS: **DNA cloning of *Plasmodium falciparum* circumsporozoite gene: amino acid sequence of repetitive epitope**. *Science* 1984, **225**(4662):628-630.
7. Lê HG, Kang J-M, Moe M, Jun H, Thái TL, Lee J, Myint MK, Lin K, Sohn W-M, Shin H-J: **Genetic polymorphism and natural selection of circumsporozoite surface protein in *Plasmodium falciparum* field isolates from Myanmar**. *Malaria journal* 2018, **17**(1):1-14.
8. Zeeshan M, Alam MT, Vinayak S, Bora H, Tyagi RK, Alam MS, Choudhary V, Mitra P, Lumb V, Bharti PK: **Genetic variation in the *Plasmodium falciparum* circumsporozoite protein in India and its relevance to RTS, S malaria vaccine**. *PloS one* 2012, **7**(8):e43430.
9. Rutgers T, Gordon D, Gathoye AM, Hollingdale M, Hockmeyer W, Rosenberg M, Wilde MD: **Hepatitis B Surface Antigen as Carrier Matrix for the Repetitive Epitope of the Circumsporozoite Protein of *Plasmodium Falciparum***. *Nature Biotechnology* 1988, **6**(9):1065-1070.
10. Alonso PL, Sacarlal J, Aponte JJ, Leach A, Macete E, Milman J, Mandomando I, Spiessens B, Guinovart C, Espasa M: **Efficacy of the RTS, S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial**. *The Lancet* 2004, **364**(9443):1411-1420.
11. Penny MA, Galactionova K, Tarantino M, Tanner M, Smith TA: **The public health impact of malaria vaccine RTS, S in malaria endemic Africa: country-specific predictions using 18 month follow-up Phase III data and simulation models**. *BMC medicine* 2015, **13**(1):1-20.

12. Mohamed NS, Ali Y, Muneer MS, Siddig EE, Sibley CH, Ahmed A: **Malaria epidemic in humanitarian crisis settings the case of South Kordofan state, Sudan.** *J Infect Dev Ctries* (1972-2680 (Electronic)).
13. Ageep TB, Damiens D, Alsharif B, Ahmed A, Salih EH, Ahmed FT, Diabaté A, Lees RS, Gilles JR, El Sayed BB: **Participation of irradiated Anopheles arabiensis males in swarms following field release in Sudan.** *Malaria journal* 2014, **13**(1):1-11.
14. Mohamed NS, Abdelbagi H, Osman HA, Ahmed AE, Yousif AM, Edris YB, Osman EY, Elsadig AR, Siddig EE, Mustafa M *et al*: **A snapshot of Plasmodium falciparum malaria drug resistance markers in Sudan: a pilot study.** *BMC Research Notes* 2020, **13**(1):512.
15. Mohamed NS, Albsheer MMA, Abdelbagi H, Siddig EE, Mohamed MA, Ahmed AE, Omer RA, Muneer MS, Ahmed A, Osman HA: **Genetic polymorphism of the N-terminal region in circumsporozoite surface protein of Plasmodium falciparum field isolates from Sudan.** *Malaria journal* 2019, **18**(1):1-10.
16. Conway DJ: **Natural selection on polymorphic malaria antigens and the search for a vaccine.** *Parasitology today* 1997, **13**(1):26-29.
17. Bongfen SE, Ntsama PM, Offner S, Smith T, Felger I, Tanner M, Alonso P, Nebie I, Romero JF, Silvie O: **The N-terminal domain of Plasmodium falciparum circumsporozoite protein represents a target of protective immunity.** *Vaccine* 2009, **27**(2):328-335.
18. Escalante AA, Grebert HM, Isea R, Goldman IF, Basco L, Magris M, Biswas S, Kariuki S, Lal AA: **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.** *Molecular biochemical parasitology* 2002, **125**(1-2):83-90.
19. Garçon N, Heppner DG, Cohen J: **Development of RTS, S/AS02: a purified subunit-based malaria vaccine candidate formulated with a novel adjuvant.** *Expert review of vaccines* 2003, **2**(2):231-238.
20. Bojang KA, Milligan PJ, Pinder M, Vigneron L, Allouche A, Kester KE, Ballou WR, Conway DJ, Reece WH, Gothard P: **Efficacy of RTS, S/AS02 malaria vaccine against Plasmodium falciparum infection in semi-immune adult men in The Gambia: a randomised trial.** *The Lancet* 2001, **358**(9297):1927-1934.
21. Aponte JJ, Aide P, Renom M, Mandomando I, Bassat Q, Sacarlal J, Manaca MN, Lafuente S, Barbosa A, Leach A: **Safety of the RTS, S/AS02D candidate malaria vaccine in infants living in a highly endemic area of Mozambique: a double blind randomised controlled phase I/IIb trial.** *The Lancet* 2007, **370**(9598):1543-1551.
22. Asante KP, Abdulla S, Agnandji S, Lyimo J, Vekemans J, Soulanoudjingar S, Owusu R, Shomari M, Leach A, Jongert E: **Safety and efficacy of the RTS, S/AS01E candidate malaria vaccine given with expanded-programme-on-immunisation vaccines: 19 month follow-up of a randomised, open-label, phase 2 trial.** *The Lancet infectious diseases* 2011, **11**(10):741-749.
23. Organization WH: **Guidelines for the treatment of malaria:** World Health Organization, 2015.
24. Rubio J, Benito A, Berzosa P, Roche J, Puente S, Subirats M, Lopez-Velez R, Garcia L, Alvar J: **Usefulness of seminested multiplex PCR in surveillance of imported malaria in Spain.** *Journal of*

Clinical Microbiology 1999, **37**(10):3260-3264.

25. Ntoumi F, Ngoundou-Landji J, Lekoulou F, Luty A, Deloron P, Ringwald P: **Site-based study on polymorphism of Plasmodium falciparum MSP-1 and MSP-2 genes in isolates from two villages in Central Africa.** *Parassitologia* 2000, **42**(3-4):197-203.
26. Manske M: **GENTle, a free multi-purpose molecular biology tool.** Universität zu Köln, 2006.
27. Hall N, Pain A, Berriman M, Churcher C, Harris B, Harris D, Mungall K, Bowman S, Atkin R, Baker S: **Sequence of Plasmodium falciparum chromosomes 1, 3–9 and 13.** *Nature* 2002, **419**(6906):527-531.
28. Caspers P, Gentz R, Matile H, Pink JR, Sinigaglia F: **The circumsporozoite protein gene from NF54, a Plasmodium falciparum isolate used in malaria vaccine trials.** *Molecular biochemical parasitology* 1989, **35**(2):185-189.
29. Tanabe K, Sakihama N, Hattori T, Ranford-Cartwright L, Goldman I, Escalante AA, Lal AA: **Genetic distance in housekeeping genes between Plasmodium falciparum and Plasmodium reichenowi and within P. falciparum.** *Journal of molecular evolution* 2004, **59**(5):687-694.
30. Lockyer MJ, Schwarz RT: **Strain variation in the circumsporozoite protein gene of Plasmodium falciparum.** *Molecular biochemical parasitology* 1987, **22**(1):101-108.
31. Kumar S, Stecher G, Tamura K: **MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.** *Molecular biology evolution* 2016, **33**(7):1870-1874.
32. Kishino H, Hasegawa M: **Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea.** *Journal of molecular evolution* 1989, **29**(2):170-179.
33. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**(11):1451-1452.
34. Múrias dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M: **tcsBU: a tool to extend TCS network layout and visualization.** *Bioinformatics* 2016, **32**(4):627-628.
35. Fu Y-X, Li W-H: **Statistical tests of neutrality of mutations.** *Genetics* 1993, **133**(3):693-709.
36. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**(3):585-595.

Figures

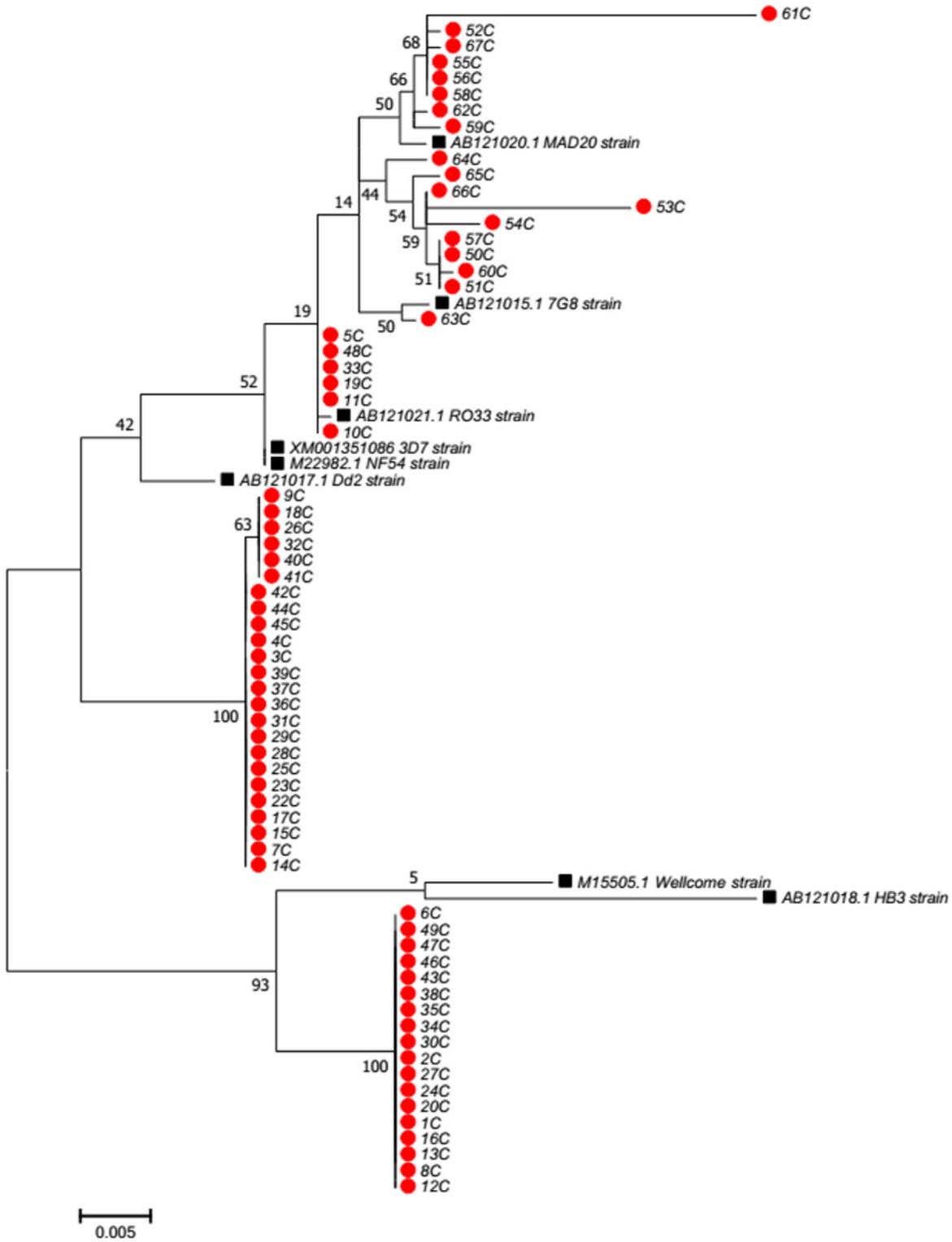


Figure 1

Phylogenetic network of the Sudanese PfCSP and PfCSP reference sequences.

