

RESEARCH

Open Access



Genetic polymorphism and natural selection of circumsporozoite surface protein in *Plasmodium falciparum* field isolates from Myanmar

Hương Giang Lê^{1,2}, Jung-Mi Kang^{1,2}, Mya Moe³, Hojong Jun⁴, Thi Lam Thái^{1,2}, Jinyoung Lee⁴, Moe Kyaw Myint³, Khin Lin³, Woon-Mok Sohn¹, Ho-Joon Shin⁵, Tong-Soo Kim⁴ and Byoung-Kuk Na^{1,2*}

Abstract

Background: *Plasmodium falciparum* circumsporozoite protein (PfCSP) is one of the most extensively studied malaria vaccine candidates, but the genetic polymorphism of PfCSP within and among the global *P. falciparum* population raises concerns regarding the efficacy of a PfCSP-based vaccine efficacy. In this study, genetic diversity and natural selection of PfCSP in Myanmar as well as global *P. falciparum* were comprehensively analysed.

Methods: Blood samples were collected from 51 *P. falciparum* infected Myanmar patients. Fifty-one full-length PfCSP genes were amplified from the blood samples through a nested polymerase chain reaction, cloned into a TA cloning vector, and then sequenced. Polymorphic characteristics and natural selection of Myanmar PfCSP were analysed using the DNASTAR, MEGA6, and DnaSP programs. Polymorphic diversity and natural selection in publicly available global PfCSP were also analysed.

Results: The N-terminal and C-terminal non-repeat regions of Myanmar PfCSP showed limited genetic variations. A comparative analysis of the two regions in global PfCSP displayed similar patterns of low genetic diversity in global population, but substantial geographic differentiation was also observed. The most notable polymorphisms identified in the N-terminal region of global PfCSP were A98G and 19-amino acid length insertion in global population with different frequencies. Major polymorphic characters in the C-terminal region of Myanmar and global PfCSP were found in the Th2R and Th3R regions, where natural selection and recombination occurred. The central repeat region of Myanmar PfCSP was highly polymorphic, with differing numbers of repetitive repeat sequences NANP and NVDP. The numbers of the NANP repeats varied among global PfCSP, with the highest number of repeats seen in Asian and Oceanian PfCSP. Haplotype network analysis of global PfCSP revealed that global PfCSP clustered into 103 different haplotypes with geographically-separated populations.

Conclusion: Myanmar and global PfCSP displayed genetic diversity. N-terminal and C-terminal non-repeat regions were relatively conserved, but the central repeat region displayed high levels of genetic polymorphism in Myanmar and global PfCSP. The observed geographic pattern of genetic differentiation and the points of evidence for natural selection and recombination suggest that the functional consequences of the polymorphism should be considered for developing a vaccine based on PfCSP.

Keywords: *Plasmodium falciparum*, Circumsporozoite protein, Genetic polymorphism, Natural selection, Myanmar

*Correspondence: bkna@gnu.ac.kr

¹ Department of Parasitology and Tropical Medicine and Institute of Health Sciences, Gyeongsang National University College of Medicine, Jinju 52727, Republic of Korea

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



Background

Malaria, caused by *Plasmodium* spp. infections, is one of the most significant life-threatening infectious diseases to humans worldwide. It accounted for more than 216 million cases and approximately 450,000 deaths across the globe in 2017 [1]. The Greater Mekong Sub-region (GMS) including Myanmar has long been one of the most malarious regions in the world [2]. Among the countries in GMS, Myanmar has the highest malaria burden, accounting for an estimated 77% of malaria cases and approximately 79% of malaria deaths in the GMS [3]. In spite of recent decreases in malaria cases and deaths, malaria is still a major public health concern in Myanmar [4].

To date, there is no licensed vaccine against malaria, though many efforts and studies have been performed in order to develop effective vaccines. Various vaccine constructs based on diverse antigens from sexual and asexual stages of *Plasmodium falciparum* have been investigated. Among these, RTS,S, currently the most advanced malaria vaccine candidate [5, 6], is based on circumsporozoite protein of *P. falciparum* (PfCSP). RTS,S is comprised of a liposome-based adjuvant (AS01) and hepatitis B virus surface antigen (HBsAg) virus-like particles incorporating a portion of the PfCSP genetically fused to HBsAg. PfCSP is a dominant surface protein of sporozoites, and it plays a critical role in the invasion of hepatocytes by sporozoites [7–9]. PfCSP is divided into three distinct regions: a highly variable central repeat region flanked by a conserved N-terminal region and a C-terminal non-repeat region. The central repeat region, which has been recognized as a major target for antibody-mediated neutralization, is rich in Asn-Ala-Asn-Pro (NANP) tandem repeats and also contains a small number of Asn-Val-Asp-Pro (NVDP) motifs [10–12]. The C-terminal non-repeat region includes two polymorphic sub-regions, Th2R and Th3R, where T cell epitopes were identified. These regions show moderate polymorphisms which might have resulted from natural selection by the host immune system [13–15].

Recent genome sequencing studies have demonstrated that *P. falciparum* from different geographic regions have diverse genetic makeup [16, 17], which emphasizes the importance of comprehensive analysis of parasite genetic diversity and population structure in the global *P. falciparum* population. Indeed, most *P. falciparum* vaccine candidate antigens including PfCSP have been found to show various genetic and antigenic polymorphisms in global isolates [18], which can obstruct or reduce the efficacy of vaccines based on PfCSP. Therefore, understanding the genetic nature of vaccine candidate antigens in global *P. falciparum* isolates is critical for designing an effective vaccine. In this study, genetic polymorphism and natural

selection of PfCSP in *P. falciparum* Myanmar isolates were analysed. A comparative analysis of global PfCSP was also performed in order to gain an in-depth understanding of the genetic makeup of PfCSP in the global *P. falciparum* population.

Methods

Blood samples

A total of 51 blood samples used in this study were collected from malaria patients infected with *P. falciparum* in Myanmar in 2015. The patients were selected in field surveys for malaria, which were conducted in towns and villages in Naung Cho, Pyin Oo Lwin, Tha Beik Kyin townships, and Mandalay in Upper Myanmar. Infections were diagnosed through microscopic examination of thin and thick blood smears. Finger-prick blood samples were taken from *P. falciparum* infected symptomatic patients prior to drug treatment and spotted on Whatman 3 MM filter paper (GE Healthcare, Maidstone, UK) for confirmation by polymerase chain reaction (PCR) targeting the 18S ribosomal RNA (rRNA) gene [4, 19], as well as for subsequent molecular analysis. The mean age of patients who donated the blood samples was 32.7 years-old and ranged between 13 and 57 years. Informed consent was obtained from all patients before blood collection. The study protocol was reviewed and approved by either the Ethics committee of the Ministry of Health, Myanmar (97/Ethics 2015) or the Biomedical Research Ethics Review Board of Inha University School of Medicine, Republic of Korea (INHA 15-013).

Genomic DNA extraction and amplification of PfCSP

Genomic DNA was extracted from dried blood spots using the QIAamp DNA Blood Kit (Quiagen, Hilden, Germany) following the manufacturer's protocol. The full-length region encoding PfCSP was amplified through a nested PCR method. The primers for the first round PCR were 5'-ATG ATG AGA AAA TTA GCT ATT TTA TCT GTT-3' and 5'-CTA ATT AAG GAA CAA GAA GGA TAA TAC CAT-3'. The primers used for nested PCR were 5'-AGA AAA TTA GCT ATT TTA TCT GTT TCT-3' and 5'-ACA AGA AGG ATA ATA CCA TTA TTA ATC-3'. Ex *Taq* DNA polymerase (Takara, Otsu, Japan) with proof-reading activity was used in all PCR amplification steps to minimize the nucleotide mis-incorporation. The following thermal cycling conditions were used for both amplifications: 94 °C for 5 min; and 30 cycles of 94 °C for 1 min, 52 °C for 1 min, and 72 °C for 1 min 30 s, followed by a final extension at 72 °C for 10 min. The PCR product was resolved on a 1.2% agarose gel, purified from gel, then ligated into the T&A cloning vector (Real Biotech Corporation, Banqiao City, Taiwan). Each ligation mixture was transformed

into *Escherichia coli* DH5 α competent cells and positive clones, with appropriate inserts screened by colony PCR. The nucleotide sequences of the cloned PfCSP were analysed through automatic DNA sequencing with M13 forward and M13 reverse primers by the Sanger method. Plasmids from at least two independent clones from each transformation mixture were sequenced in both directions so as to verify the sequence accuracy. The nucleotide sequences reported in this study have been deposited in the GenBank database under the accession numbers MF350670–MF350720.

Sequence polymorphism analysis

The nucleotide and deduced amino acid sequences of PfCSP were analysed using EditSeq and SeqMan in the DNASTAR package (DNASTAR, Madison, WI, USA). The PfCSP sequence of the laboratory-adapted *P. falciparum* strain 3D7 (XM_001351086) was included in the alignment for comparison as a reference sequence. The values of segregating sites (S), the average number of pair-wise nucleotide differences (K), haplotype diversity (H_d), and nucleotide diversity (π) were calculated using DnaSP version 5.10.00 [20]. The π was also calculated on a sliding window plot of 10 bases with a step size of 5 bp in order to estimate the stepwise diversity across the sequences. In order to test the null hypothesis of neutrality of PfCSP, the rates of synonymous (d_S) and non-synonymous (d_N) substitutions were estimated and were compared using the Z-test ($P < 0.05$) in MEGA6 program [21] using Nei and Gojobori's method [22] with the Jukes and Cantor (JC) correction of 1000 bootstrap replications. Tajima's D test [23], Fu and Li's D and F statistics analysis [24] were performed using DnaSP ver. 5.10.00 [20] in order to evaluate the neutral theory of natural selection. The recombination parameter (R), which included the effective population size and probability of recombination between adjacent nucleotides per generation, and the minimum number of recombination events (R_m) were analysed using DnaSP ver. 5.10.00 [20].

Genetic diversities of PfCSP among global *Plasmodium falciparum* isolates

The genetic diversities of PfCSP among global *P. falciparum* isolates were analysed. The PfCSP sequences used in this study were from Thailand, Philippines, Vietnam, India, Iran, Papua New Guinea, Vanuatu, Solomon Islands, Kenya, Cameroon, Ghana, Tanzania, Senegal, Gambia, Brazil, and Venezuela (Additional file 1: Table S1). These sequences cover full-length or partial portions of PfCSP. Genetic polymorphism and tests of neutrality were calculated for each population using DnaSP ver. 5.10.00 [20] and MEGA6 [21] as described above. A logo plot was constructed for each PfCSP

population in order to analyse polymorphic patterns of the C-terminal non-repeat region in global PfCSP using the WebLogo program (<https://weblogo.berkeley.edu/logo.cgi>). In order to investigate the genetic relationships among global PfCSP haplotypes, the haplotype network for 817 full-length sequences of PfCSP from Myanmar and other countries listed above was analysed using NETWORK version 5.0.0.3 with the Median joining algorithm [25].

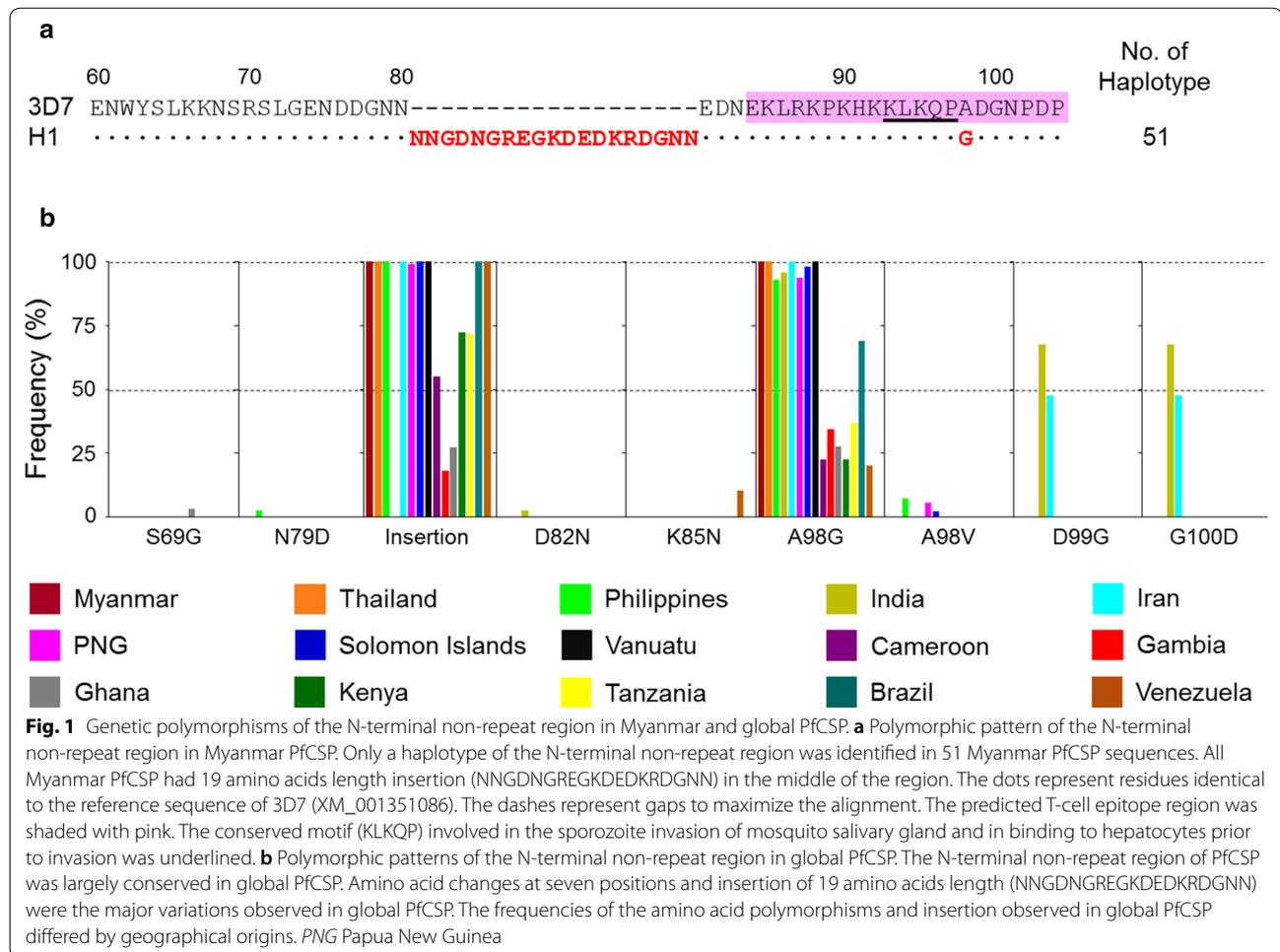
Results

Amplification of Myanmar PfCSP

Fifty-one full-length PfCSP were successfully amplified from the 51 blood samples analysed in this study. As expected, size variations were observed in the amplified PfCSP. The approximate sizes of amplified products varied from 0.9 to 1.3 kb, which was mainly caused by differences in the number of tandem repeats in the central repeat region.

Genetic polymorphisms in the N-terminal region of Myanmar and global PfCSP

The N-terminal non-repeat region was highly conserved in Myanmar PfCSP. Only a haplotype with a 57 bp (encoding 19 amino acids of NNGDNGREGKDEDKRDGNN) insertion in the middle portion of the region compared to the 3D7 reference sequence (XM_001351086) was identified (Fig. 1a). A comparative analysis of the N-terminal non-repeat region in global PfCSP also showed that the region is relatively well-conserved in global PfCSP. The 19 amino acids length insertion was the major variation observed in global PfCSP, but the frequency of this insertion in global PfCSP differed according to geographical origin. Asian PfCSP, that is, from Myanmar, Thailand, Philippines, and Iran with the exception of India, showed the 19 amino acids insertion in the region. Oceanian and South American PfCSP, that is, from PNG, Solomon Islands, Vanuatu, Brazil, and Venezuela, also showed a high frequency of the insertion, ranging from 98.9 to 100%. Meanwhile, the frequency was lower in African PfCSP, specifically in Cameroon (55.0%), Gambia (18.0%), Ghana (27.0%), Kenya (72.2%), and Tanzania (71.6%) (Fig. 1b). Amino acid polymorphisms were also found at seven positions (S69G, N79D, D82N, K85N, A98G/V, D99G and G100D) in global PfCSP. A98G was commonly identified in global PfCSP, but its frequency varied among global PfCSP. The overall frequency was high (up to 92%) in Asian and Oceanian PfCSP, but low (less than 37%) in African and South American PfCSP, with the exception of Brazil (69.0%). Meanwhile, S69G, N79D, D82N, K85N, and A98V showed uneven geographic distributions and very low frequencies. D99G and G100D were identified only in Indian and Iranian PfCSP. Most amino



acid polymorphisms identified in the N-terminal region of PfCSP were located in the predicted T cell epitope region, ⁸⁴EKLRKPKHKKLLKQPADGNPDP¹⁰⁴. However, the conserved motif (KLKQP) that was involved in sporozoite invasion of mosquito salivary gland and in binding to hepatocytes prior to invasion [26] was well-conserved in all global PfCSP.

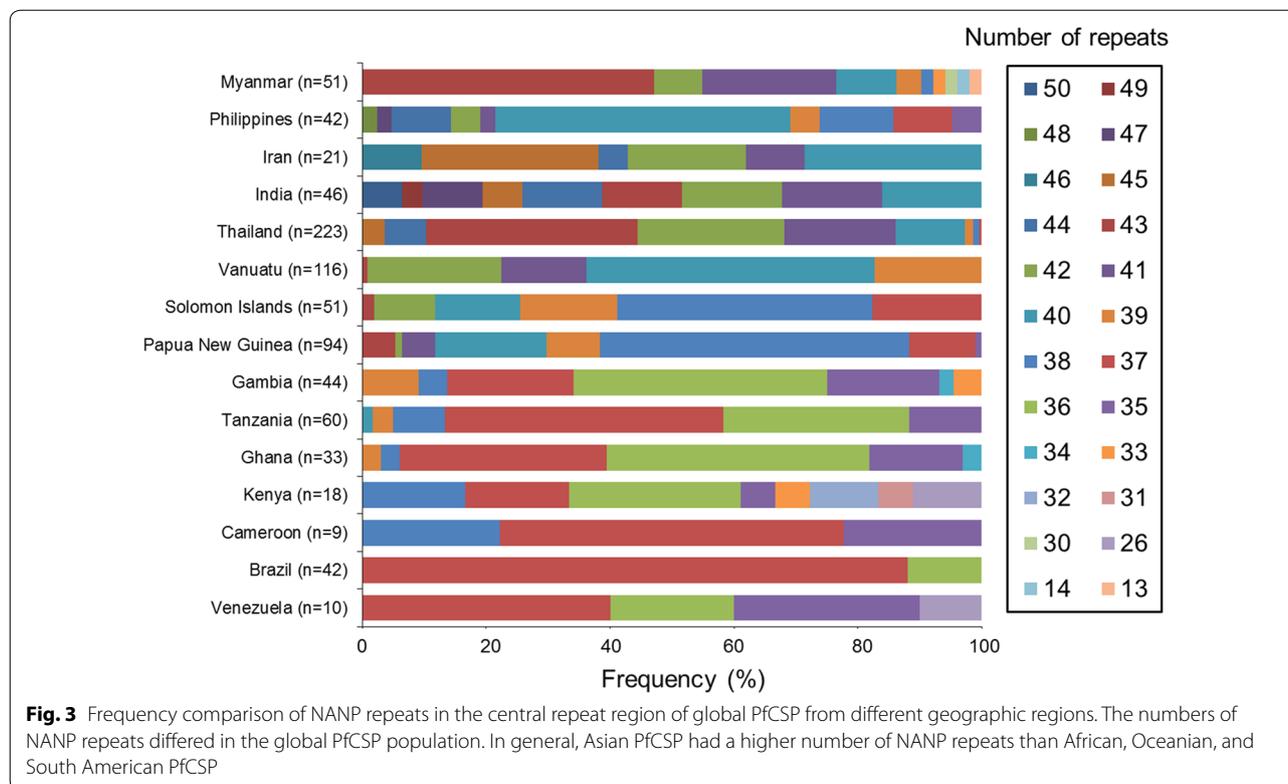
Genetic polymorphisms in the central repeat region of Myanmar PfCSP

A total of 14 unique haplotypes of Myanmar PfCSP were identified in amino acid levels (Fig. 2a). Two novel repeat allotypes, which encode NTNP and NANS motifs, were identified in two haplotypes (H3 and H9) of Myanmar PfCSP. Each haplotype of Myanmar PfCSP had different numbers of previously-known tetrapeptide repeats, NANP and NVDP motifs, ranging from 17 to 48. These different numbers of repeats resulted in size polymorphisms in the central repeat region among Myanmar PfCSP. Most Myanmar PfCSP had numbers of

tetrapeptide repeats between 44 and 47 with a frequency of 76.4% (Fig. 2b).

Polymorphic patterns of NANP repeats in global PfCSP

The numbers of NANP repeats in PfCSP populations from different geographical regions including Philippines, Iran, India, Thailand, Vanuatu, Papua New Guinea, Solomon Islands, Gambia, Tanzania, Ghana, Kenya, Cameroon, Brazil, and Venezuela (Additional file 1: Table S1) were analysed and compared with those of Myanmar PfCSP. Comparative analysis of global PfCSP revealed that the numbers of NANP repeats in global PfCSP differed by geographical origins (Fig. 3). Asian PfCSP had a high number of NANP repeats ranging from 40 to 43. Meanwhile, 36 and 37 NANP repeats were mainly observed in African and South American PfCSP. PfCSP from two Oceanian countries, Papua New Guinea and Solomon Islands, had 38 NANP repeats with a high frequency, whereas a higher number of repeats (40 repeats) was predominant in Vanuatu PfCSP.



Nucleotide diversity and natural selection of the C-terminal non-repeat region in Myanmar PfCSP

The nucleotide diversity and genetic differentiation were analysed in the C-terminal non-repeat region in Myanmar PfCSP. The average number of nucleotide differences (K) in this region was 0.92. The overall haplotype diversity (H_d) and nucleotide diversity (π) in this region were estimated to be 0.329 ± 0.072 and 0.004 ± 0.003 , respectively. In order to investigate whether natural selection has contributed to the diversity in C-terminal non-repeat region in Myanmar PfCSP, the value of $dN-dS$ for this region was analysed. The estimated value of $dN-dS$ was found to be 0.005 ± 0.003 (Table 1), suggesting that this region may be influenced by a positive natural selection. Tajima's D test was also performed to further analyse the natural selection in the C-terminal non-repeat region in Myanmar PfCSP. Tajima's D value was found to be 0.764 ($P > 0.1$) (Table 1). The F_u and F_L 's D and F values were also positive at 0.889 ($P > 0.1$) and 0.992 ($P > 0.1$), respectively.

Nucleotide diversity, natural selection, and recombination of the C-terminal non-repeat region among global PfCSP

Genetic diversity in the C-terminal non-repeat region among global PfCSP was analysed in order to assess the extent of genetic diversity between the populations

(Table 1). The K value in African PfCSP (21.76) was higher than that in South American (3.74), Asian (2.73), and Oceanian (1.09) PfCSP. The greatest nucleotide diversity was observed in African PfCSP ($\pi = 0.105 \pm 0.021$) followed by South American PfCSP ($\pi = 0.018 \pm 0.002$), Asian PfCSP ($\pi = 0.013 \pm 0.001$), and Oceanian PfCSP ($\pi = 0.005 \pm 0.0002$). The $dN-dS$ values for all global PfCSP were estimated to be positive, suggesting that positive natural selection may occur in the C-terminal non-repeat region of global PfCSP, but this trend was not statistically significant. Negative values of Tajima's D were identified in the C-terminal non-repeat region of South American PfCSP (-0.188 , $P > 0.1$), African PfCSP (-0.831 , $P > 0.1$), and Asian PfCSP (-1.032 , $P > 0.1$), indicating that they were under purifying selection. Meanwhile, the C-terminal non-repeat region of Oceanian PfCSP (0.626, $P > 0.1$) showed positive Tajima's D values, suggesting the effects of balance selection on the population. Sliding window plot analysis (window length of 10 bp and step size of 5 bp) in global PfCSP showed that global PfCSP shared highly similar patterns of nucleotide diversity across the region. Nucleotide diversity peaked at the Th2R and Th3R T-cell epitopes, although the values of π were slightly different within and between PfCSP populations according to geographical origin (Fig. 5). Interestingly, only a single major peak of nucleotide

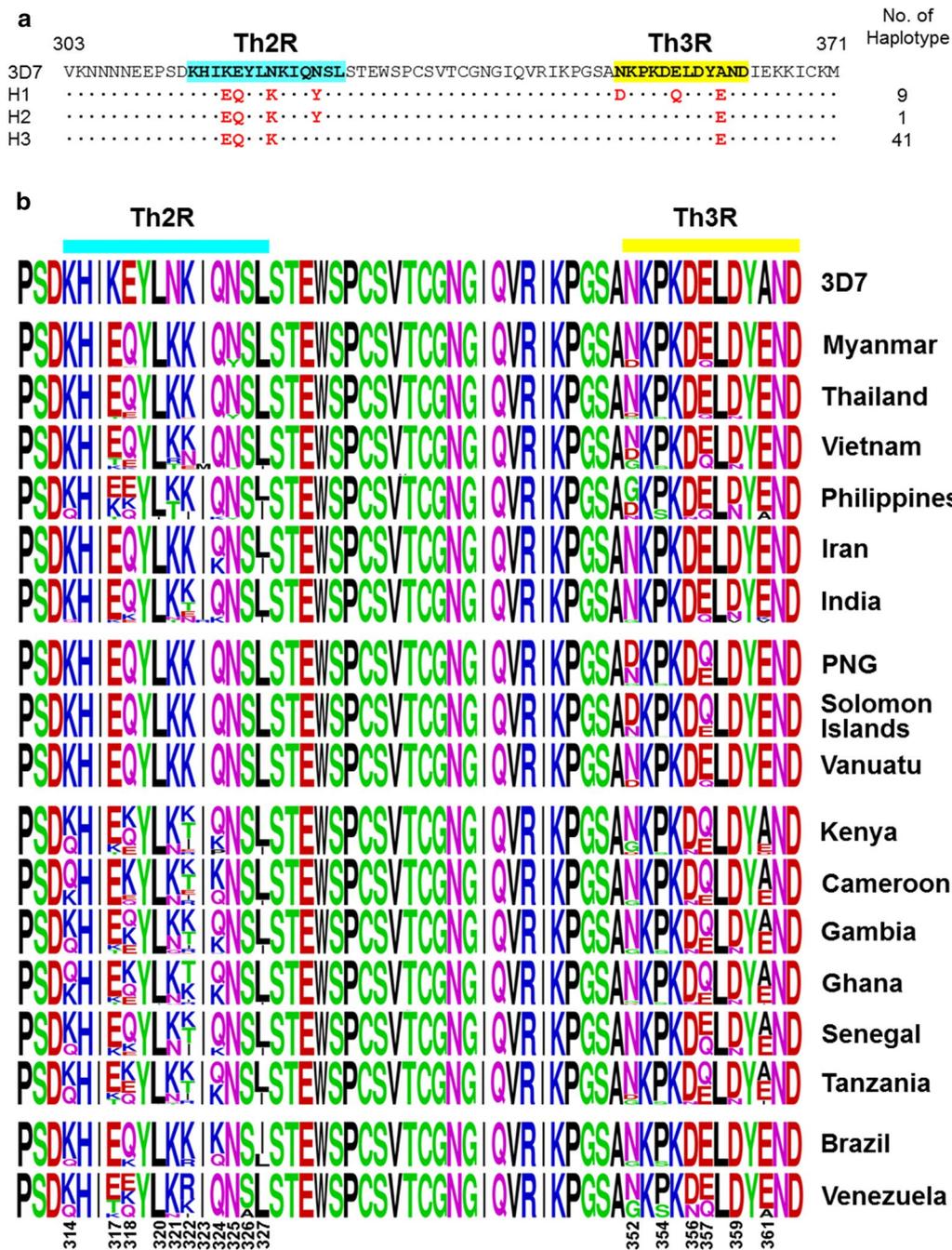


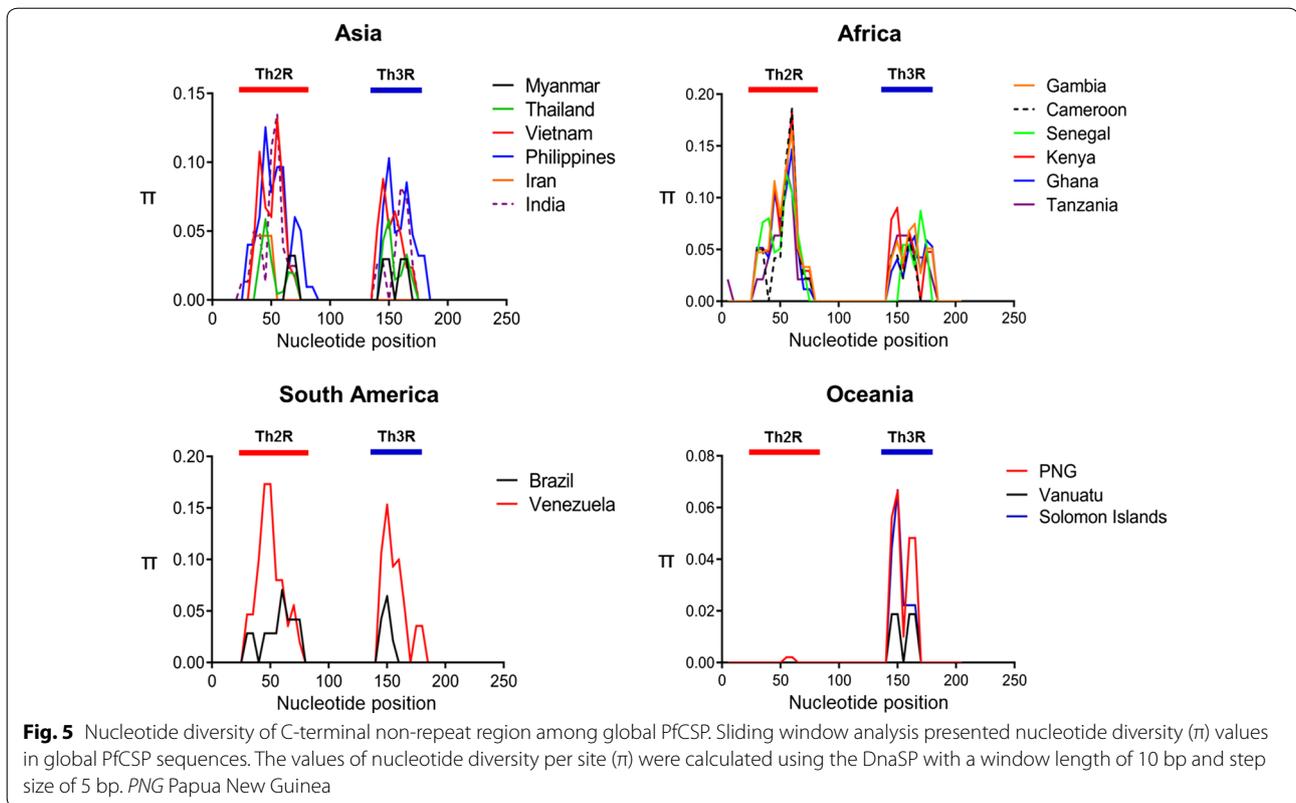
Fig. 4 Polymorphic patterns of the C-terminal non-repeat region in Myanmar and global PfCSP. **a** Amino acid polymorphisms in the C-terminal non-repeat region of Myanmar PfCSP. The three different haplotypes (H1–H3) of the C-terminal non-repeat region were identified in Myanmar PfCSP. Th2R and Th3R regions are shaded with sky blue and yellow, respectively. Amino acid changes at particular amino acid positions are indicated as red. The number of haplotypes is the number of PfCSP sequences belonging to each haplotype. The dots represent residues identical to the reference sequence of 3D7 (XM_001351086). **b** Comparative analysis of polymorphic patterns of the C-terminal non-repeat region in global PfCSP. The pattern of amino acid changes differed by country or geographic continent, but most amino acid changes were concentrated in the Th2R and Th3R regions in global PfCSP. A logo plot was constructed for each PfCSP population using the WebLogo program. PNG Papua New Guinea

Table 1 Genetic polymorphism and tests of neutrality in C-terminal region of global PfCSP

Country	n	K	S	Total No. of mutations	H	Hd ± SD	π ± SD	dN-dS	Tajimas D ^{P-value}	Fu and Li's D ^{P-value}	Fu and Li's F ^{P-value}
<i>Asia</i>											
Myanmar	51	0.92	3	3	3	0.329 ± 0.072	0.004 ± 0.003	0.005 ± 0.003	0.764 ^a	0.889 ^a	0.992 ^a
Vietnam	23	4.39	14	16	23	1.000 ± 0.013	0.019 ± 0.001	0.024 ± 0.008	0.050 ^a	0.862 ^a	0.722 ^a
Thailand	223	1.74	12	13	8	0.485 ± 0.039	0.009 ± 0.001	0.011 ± 0.004	-0.483 ^a	1.475 ^d	0.886 ^a
Philippines	42	5.79	18	19	9	0.821 ± 0.041	0.028 ± 0.001	0.033 ± 0.009	1.019 ^a	1.649 ^b	1.701 ^c
Iran	21	0.93	2	2	2	0.467 ± 0.075	0.004 ± 0.001	0.006 ± 0.004	1.564 ^a	0.858 ^a	1.205 ^a
India	15	3.71	18	19	13	0.981 ± 0.031	0.019 ± 0.004	0.023 ± 0.007	-1.486 ^a	-1.173 ^a	-1.448 ^a
Total	375	2.73	25	29	38	0.653 ± 0.027	0.013 ± 0.001	0.016 ± 0.005	-1.032 ^a	-0.235 ^a	-0.698 ^a
<i>Oceania</i>											
PNG	94	1.17	5	5	4	0.532 ± 0.034	0.006 ± 0.001	0.007 ± 0.004	0.419 ^a	-0.012 ^a	0.151 ^a
Vanuatu	116	0.55	2	2	2	0.276 ± 0.046	0.003 ± 0.001	0.003 ± 0.002	0.706 ^a	0.674 ^a	0.801 ^a
Solomon Islands	51	0.83	4	4	3	0.397 ± 0.066	0.004 ± 0.0007	0.005 ± 0.003	-0.1426 ^a	-1.209 ^a	-1.029 ^a
Total	261	1.09	5	5	4	0.522 ± 0.012	0.005 ± 0.0002	0.007 ± 0.004	0.626 ^a	-0.214 ^a	0.087 ^a
<i>Africa</i>											
Kenya	18	5.83	17	20	12	0.922 ± 0.051	0.028 ± 0.003	0.035 ± 0.011	0.010 ^a	0.295 ^a	0.247 ^a
Cameroon	9	4.56	12	15	7	0.944 ± 0.07	0.022 ± 0.005	0.028 ± 0.009	-0.842 ^a	-0.689 ^a	-0.813 ^a
Gambia	44	5.83	18	21	20	0.944 ± 0.019	0.028 ± 0.002	0.036 ± 0.009	0.677 ^a	1.006 ^a	1.068 ^a
Ghana	33	5.02	18	22	19	0.932 ± 0.028	0.024 ± 0.003	0.031 ± 0.009	-0.257 ^a	0.474 ^a	0.278 ^a
Tanzania	60	6.45	22	26	30	0.969 ± 0.008	0.031 ± 0.001	0.039 ± 0.011	0.499 ^a	0.852 ^a	0.863 ^a
Senegal	11	4.54	10	12	9	0.964 ± 0.051	0.019 ± 0.002	0.024 ± 0.007	0.475 ^a	1.149 ^a	1.106 ^a
Total	175	21.76	150	169	58	0.968 ± 0.004	0.105 ± 0.021	0.169 ± 0.055	-0.831 ^a	2.648 ^b	1.205 ^a
<i>South America</i>											
Venezuela	10	6.44	14	16	7	0.911 ± 0.077	0.031 ± 0.003	0.034 ± 0.014	0.645 ^a	0.799 ^a	0.857 ^a
Brazil	42	2.33	8	8	3	0.459 ± 0.08	0.011 ± 0.002	0.014 ± 0.005	0.721 ^a	1.309 ^d	1.317 ^a
Total	52	3.74	16	18	9	0.630 ± 0.065	0.018 ± 0.002	0.022 ± 0.007	-0.188 ^a	0.431 ^a	0.257 ^a

PNG Papua New Guinea, n number of sequences analysed, S number of segregating sites, K average number of nucleotide differences, H number of haplotypes, Hd haplotype diversity, π observed average pairwise nucleotide diversity, dN rate of non-synonymous mutations, dS rate of synonymous mutations, dS standard deviation

^a P > 0.1; ^b P < 0.02; ^c P < 0.05; ^d 0.05 < P < 0.1



diversity was identified at the Th3R epitopes in Oceanian PfCSP. Recombination events were predicted in Myanmar and global PfCSP. The recombination parameters expected in PfCSP-differed by geographical population, but a possible recombination event has been suggested in global PfCSP (Table 2). High R_m values were predicted for African PfCSP, while lower levels of R_m were identified in PfCSP from other geographical areas.

Haplotype network analysis of PfCSP among global *P. falciparum* isolates

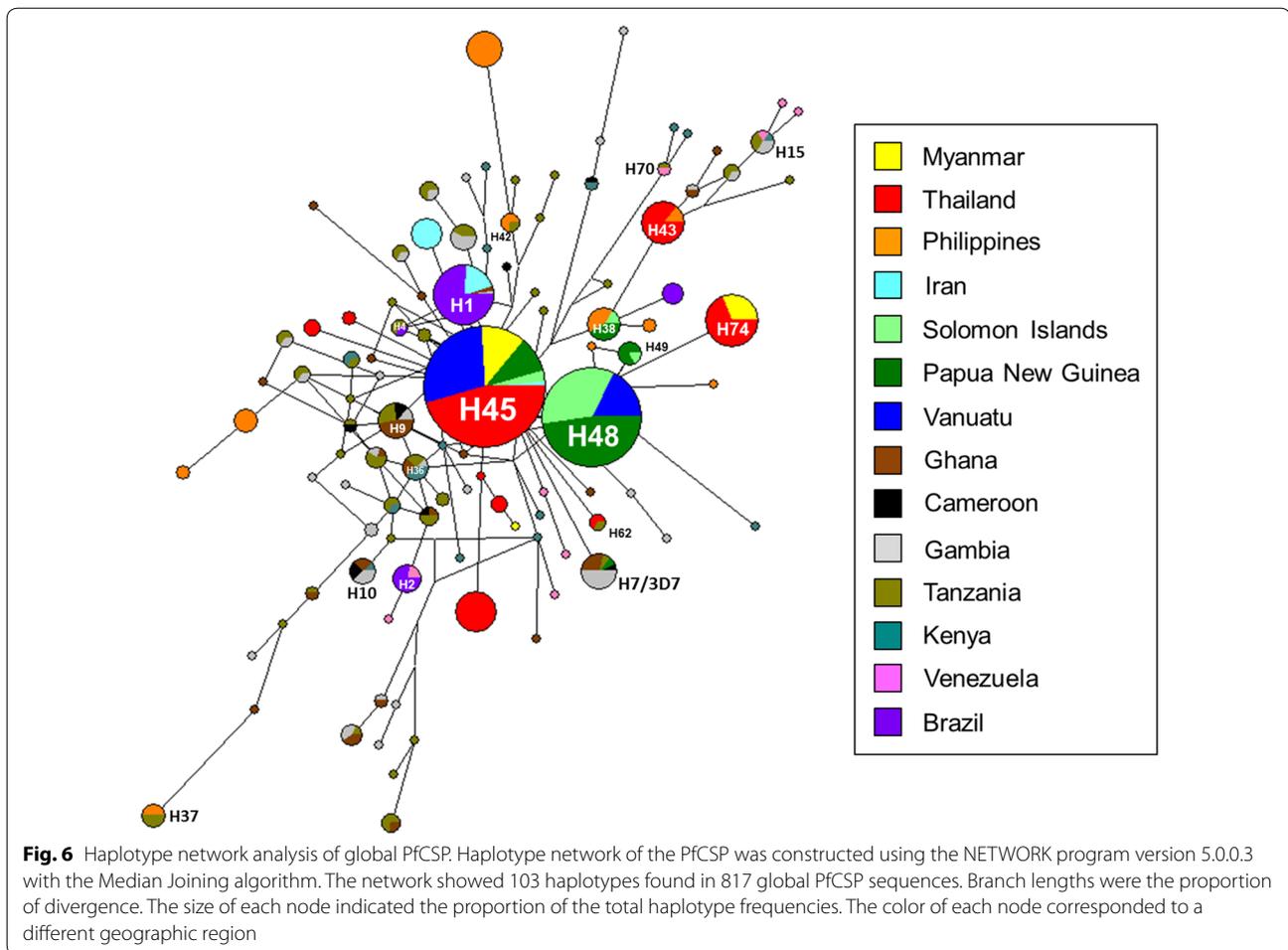
A haplotype network was constructed in order to analyse the relationships between and among PfCSP from global *P. falciparum* isolates. Indian, Vietnamese, and Senegal PfCSP sequences were precluded, as these sequences were not cover full-length sequences. A total of 103 distinct haplotypes were identified in the 817 global PfCSP sequences analysed (Fig. 6). Forty-eight haplotypes (46.6%) were shared by PfCSP sequences from at least two different countries, while 55 haplotypes with a frequency of 53.4% were limited to singleton. The haplotype prevalence ranged from 0.12% to 41.86%. Haplotype 45 (H45) was the most prevalent haplotype with a frequency of 41.86% and was occupied by Asian and Oceanian PfCSP. The two haplotypes of H48 and H49 were shared only with Oceanian PfCSP. Meanwhile, the two haplotypes of

Table 2 Recombination events in global PfCSP

Country	Ra	Rb	Rm
Myanmar	0.000	0.001	0
Vietnam	0.504	116	3
Thailand	0.000	0.001	2
Philippines	0.082	16.9	1
India	0.000	0.001	0
Iran	0.043	8.2	2
PNG	0.024	4.9	1
Solomon Islands	0.000	0.001	0
Vanuatu	0.000	0.001	0
Ghana	0.071	19.5	6
Tanzania	0.171	35.1	6
Cameroon	0.077	15.9	2
Senegal	0.172	41.2	2
Kenya	0.110	22.7	6
Gambia	0.116	23.9	5
Brazil	0.000	0.001	0
Venezuela	0.1403	28.9	2

The R and R_m were estimated excluding the sites containing alignment gaps or those segregating for three nucleotides

R_a recombination parameter between adjacent sites, R_b recombination parameter for entire gene, R_m minimum number of recombination events between adjacent sites, PNG Papua New Guinea



H43 and H74 were composed only of Asian PfCSP. Only nine haplotypes (H1, H4, H7, H15, H37, H42, H45, H62, and H70) were made by inter-continental PfCSP. No haplotypes were identified to be mixed haplotypes shared among populations from all four continents. Most singletons were mainly identified among African PfCSP, suggesting that high genetic polymorphisms occurred in the African population. Haplotypes from Venezuela PfCSP were closely linked to African PfCSP, suggesting genetic closeness between the two populations. The haplotype identical to the 3D7 sequence (XM_001351086) was haplotype 7 (H7), which occupied 1.71% of frequency and was mainly shared by African PfCSP.

Discussion

Diverse kinds of *P. falciparum* antigens have been extensively studied as candidate antigens for a malaria vaccine. However, genetic and antigenic variations in vaccine candidates in the global *P. falciparum* population have been immense challenges in developing an effective malaria vaccine and to certify the efficacy of the vaccine.

Therefore, understanding the genetic nature and antigenic variation of vaccine candidate antigens among global *P. falciparum* populations is important since this can provide potential rejoinders on the effects of genetic diversity in the global population for vaccine efficacy and valuable information for designing optimal vaccine formulation [27]. PfCSP is a leading candidate for a malaria vaccine and recent Phase III RTS,S vaccine trials resulted in significant reduction rates in clinical malaria [5, 6]. However, the PfCSP antigen formulated in RTS,S is a single variant, and therefore the impact of natural genetic variation in the global PfCSP population on vaccine efficacy remains unclear. In this study, the genetic polymorphism and natural selection in the Myanmar PfCSP and global PfCSP populations were comprehensively analysed.

Myanmar PfCSP had a largely well-conserved N-terminal region, which coincided with PfCSP populations from other geographical areas [18, 27–30]. A few amino acid polymorphisms were identified in global PfCSP populations, but A98G was the only commonly identified amino

acid change in global PfCSP N-terminal region, although its frequency differed by country. The most important polymorphic characteristic identified in Myanmar and global PfCSP was a 19 amino acid length insertion (NNG-DNGREGKDEDKRDGNN) in the middle of this region. This insertion was identified in all global PfCSP enrolled in this study with the only exception being Indian PfCSP, but the frequency of this insertion varied with PfCSP populations from different geographical regions. The N-terminal region of PfCSP is known to play an essential role in the invasion process of sporozoites to hepatic cells by mediating or facilitating the interaction between sporozoites and host cells [30–32]. A monoclonal antibody that binds to a linear epitope, ⁸¹EDNEKLRKPKH⁹¹, in the N-terminal region of PfCSP effectively neutralizes sporozoite infectivity in vivo, suggesting a critical role for this epitope in sporozoite infectivity to hepatocytes [33]. The functional significance of the 19 amino acids insertion in PfCSP N-terminal region is currently unclear. However, considering that this insertion is essentially located in the front of the ⁸¹EDNEKLRKPKH⁹¹ linear epitope, and that global PfCSP, except for the Indian population, had the insertion, a study aimed at understanding the role and evolutionary implication of this insertion is warranted. Most amino acid polymorphisms identified in the N-terminal region of global PfCSP was located in the predicted T-cell epitope region (⁸⁴EKLRLKPKHKLLKQPADGNPDP¹⁰⁴), indicating that this region is under host immune responses. The N-terminal region of PfCSP has been largely neglected as a potential vaccine target in spite of being a target of inhibitory antibodies and protective T cell responses. The functional importance of the N-terminal region in protective immunity has been demonstrated. Polypeptides flanking the PfCSP N-terminal region evoked the production of inhibitory antibodies for hepatocyte invasion by sporozoites, and these polypeptides are likely to render partial protective immunity in people residing in malaria-endemic regions [34]. A recent study also suggested that most of the effective antibodies that potentially inhibit malaria infection bind not only to the repeat region, but also to a portion of N-terminal junction of PfCSP [35]. These collectively highlighted the potential of the N-terminal region of PfCSP as a part of PfCSP-based vaccine constructs for malaria vaccine formulation. The low genetic polymorphic nature in the N-terminal region of global PfCSP also supports the notion that the region can be an attractive component of PfCSP-based vaccine.

The central repeat region of PfCSP has been recognized to play crucial roles in sporozoite formation and development [36]. It has been postulated that the genetic diversity of this region may be maintained by balancing selection, mainly affected by host's immune responses

[33]. Differing numbers of tetrapeptide repeats have been identified as an important source of genetic polymorphism in PfCSP. As expected, high levels of genetic polymorphisms due to different numbers of repeats were identified in the central repeat region of Myanmar PfCSP, which resulted in 14 different haplotypes. Interestingly, two novel repeats, NANS and NTNP, were identified in two haplotypes of Myanmar PfCSP, although their frequencies were low. Numerous variant forms of repeats including NVVP, NAKP, NAHP, NAIP, NVNP, NANL, NVAD, NPNP, NADP, KANP, and SANP have been reported in global PfCSP [18], but the effect of these variations is still not clearly understood. The number of repeats in the central repeat region is known to affect PfCSP stability. The stability of the type- β turn structure increases with the number of repeats [27]. Myanmar PfCSP had a high number of tetrapeptide repeats in the central repeat region, as 86.3% of Myanmar PfCSP had a number of repeats ranging from 40 to 43. Comparative analysis of the number of NANP repeats in Myanmar PfCSP and global PfCSP suggested a differing distribution of the repeats according to geographical origin, with the highest in Asian PfCSP (40–43) and the lowest in African and South American PfCSP (36–37). These suggested that PfCSP may have evolved separately, probably by evolutionary force in order to maintain or enhance protein stability, or to evade host immune response, in different geographical origin *P. falciparum* populations, resulting in the differing number of tetrapeptide repeats in the global PfCSP population. The RTS,S, the current malaria vaccine, is composed of 19 NANP tetrapeptide repeats and C-terminal T cell-epitope that are linked to the Hepatitis B surface antigen [37]. To date, there has been no direct evidence indicating that different numbers of repeats can affect the efficiency of RTS,S. However, considering that highly heterogeneous numbers of repeats are maintained in the natural PfCSP population, studies evaluating the effects of polymorphic nature in the central repeat region to RTS,S vaccine efficacy are necessary.

The C-terminal non-repeat region of Myanmar PfCSP displayed limited diversity with only three differing haplotypes among 51 Myanmar PfCSP sequences, coinciding with the previous reports on PfCSP from different geographical origins [29]. Haplotype 3, which had KHIEQYLKKIQNSL and NKPKDEL DYEND in the Th2R and Th3R regions, was the most prevalent haplotype found in Myanmar PfCSP. This allelic variant was also detected at a high frequency in Asian PfCSP populations [18, 28, 38, 39]. The overall values for haplotype diversity (H) and nucleotide diversity (π) for PfCSP C-terminal region were higher in African PfCSP than in PfCSP from other continents, indicating that African PfCSP had higher level of genetic diversity. Comparative

sliding window plot analysis of π in the C-terminal region of global PfCSP revealed similar patterns of nucleotide diversity across the region. Asian PfCSP, African PfCSP, and South American PfCSP displayed relatively similar patterns of π with two peaks at the Th2R and Th3R regions, suggesting that the genetic variations were mainly concentrated at these regions. However, differences were also found between or among PfCSP from different geographical areas. A greater π value was identified at the Th2R region than the Th3R region in Asian, African, and South American PfCSP. Meanwhile, Oceanian PfCSP revealed only a major peak of π value at the Th3R region. Polymorphisms in the Th3R region have been demonstrated as being associated with HLA binding and cytotoxic T cell reactivity [40, 41], thus these polymorphisms may assist parasites in escaping the host immune pressure. Natural selection analysis of global PfCSP C-terminal region suggests that this region is likely to be under natural selection which may maintain or generate genetic diversity in the global PfCSP population. The $dN-dS$ values for Myanmar PfCSP and global PfCSP were positive, implying that balancing selection might act in this region. The values of Tajima's D and Fu and Li's D and F revealed complicated patterns that were distinct between or among global PfCSP. These results suggested that global PfCSP was under a complicated influence of natural selection, in which either positive selection or purifying selection might have occurred in the population, depending on the geographical origin. Possible recombination events in the global PfCSP C-terminal region were also predicted. Higher values of recombination events were found in African PfCSP than in PfCSP from other geographical areas, suggesting that African PfCSP might allow for more opportunity for inter- or intra-allelic recombination than other geographical PfCSP. This might be due to the high multi-clonal infection rate of the parasite as well as subsequent cross fertilization and active recombination in mosquitoes in Africa. Interestingly, non-neglectable recombination parameters with a high haplotype diversity were predicted in Vietnamese PfCSP. Compared to the values for recombination parameters and haplotype diversity of other Asian PfCSP populations, these were extremely high in Vietnamese PfCSP.

Considering that Vietnam is a hypoendemic country with a low malaria transmission rate, the reason why Vietnamese PfCSP showed high recombination event and haplotype diversity is unclear and it should be elucidated further. Collectively, the results of genetic diversity analysis in the C-terminal region of global PfCSP suggested that global PfCSP showed limited genetic diversity in the region. However, the genetic diversity pattern of the PfCSP C-terminal region

differed slightly according to different geographical origins. Complicated natural selection acts on the global PfCSP C-terminal region, which produces genetic diversity of the region in global PfCSP. Recombination may also contribute to the genetic diversity of global PfCSP, although the recombination parameters differed by geographical origins. These genetic polymorphisms in the C-terminal region of global PfCSP suggest that more concern is required for design formulation of PfCSP-based vaccine.

Haplotype network analysis of 817 global PfCSP sequences indicated that Asian and Oceanian PfCSP formed limited numbers of clusters. Meanwhile, African PfCSP showed highly-branched and complicated patterns of haplotype diversity. No haplotype was identified that fully covers PfCSP from all of the geographic regions analysed in this study. Most singletons were mainly occupied by African PfCSP, supporting the notion that African PfCSP had higher genetic diversity than PfCSP from other geographical regions. The current RTS,S recombinant vaccine was constructed with PfCSP of *P. falciparum* NF54/3D7 strain [42]. Haplotype 7 with a frequency of 1.71%, which was shared by African PfCSP, was identical to 3D7 PfCSP. Many studies on evaluating the effectiveness and safety of RTS,S have been performed in Africa [5, 43–46], and it has been suggested that RTS,S is likely to be effective, at least in Africa. However, its efficacy worldwide may be challenging. As presented in this study, genetic heterogeneity of the PfCSP regions included in RTS,S, as well as the complicated haplotype diversity of PfCSP between and among global PfCSP, suggest that more attention is necessary toward developing a PfCSP-based vaccine, and a new approach for RTS,S that is effective in a variety of areas should be considered. If it is difficult to develop effective vaccine that works against global malaria populations, the development of an individual vaccine that works in particular malaria transmission areas by including genotypes prevalent in the geographical regions can also be considered. For example, considering that H45 and H48 are the most prevalent haplotypes of PfCSP in the Asian and Oceanian PfCSP populations, these haplotypes could be considered in designing a PfCSP-based vaccine for Asian and Oceanian countries.

The limitation of this study is that Myanmar PfCSP sequences analysed in this study were from *P. falciparum* isolates that collected in restricted areas of Myanmar. Therefore, nation-wide analysis of PfCSP in *P. falciparum* isolates collected from different regions of Myanmar is needed to clearly understand the overall genetic diversity and population structure of Myanmar PfCSP. Further examination of PfCSP nucleotide and amino acid variations in diverse PfCSP populations with a larger number

of global PfCSP sequences would be also necessary to better understand the polymorphic nature of PfCSP.

Conclusions

PfCSP is a main component of RST,S, the most advanced malaria vaccine currently, but the impact of natural genetic variations in the global PfCSP population on the efficacy of RTS,S remains unclear. Analysis of genetic diversity and natural selection of global PfCSP population suggested that the central repeat region was highly polymorphic in global PfCSP. Meanwhile, the N-terminal and C-terminal non-repeat regions revealed limited polymorphism, but differences were also identified between or among global PfCSP according to the different geographical origins. In particular, a high level of nucleotide diversity was identified in the Th2R and Th3R regions of the PfCSP C-terminal region. Complicated natural selection and recombination event were predicted in the global PfCSP C-terminal region, which may be major driving forces for maintaining and producing genetic diversity of the region in global PfCSP. The geographic patterns of genetic differentiation as well as the points of evidence for natural selection and recombination suggest that the functional consequences of the polymorphism should be considered for a vaccine based on PfCSP. The results of this study not only contribute insight into the genetic nature of Myanmar PfCSP, but also aid the understanding of genetic diversity of global PfCSP, and may provide valuable information for the development of an effective vaccine based on PfCSP. The findings of this study also warrant continuous monitoring of genetic diversity of PfCSP in the global population to better understand the polymorphic nature and evolutionary aspect of PfCSP in global *P. falciparum* population.

Additional file

Additional file 1: Table S1. Global PfCSP sequences analysed in this study.

Abbreviations

dS: rates of synonymous substitutions; dN: rate of non-synonymous substitutions; H: haplotypes; Hd: haplotype diversity; K: average number of pair-wise nucleotide differences within the population; PCR: polymerase chain reaction; PfCSP: circumsporozoite surface protein of *Plasmodium falciparum*; PNG: Papua New Guinea; Rm: minimum number of recombination events between adjacent polymorphic sites; S: segregating sites; π : nucleotide diversity.

Authors' contributions

HGL and JMK carried out genetic analysis of PfCSP. MM, HJ, TTL, JL, MKM and KL contributed the blood sample collection. HGL, JMK, WMS, HJS, TSK and BKN analysed and interpreted the data. BKN designed and supervised the experiments. HGL and BKN wrote the draft of the manuscript. All authors read and approved the final manuscript.

Author details

¹ Department of Parasitology and Tropical Medicine and Institute of Health Sciences, Gyeongsang National University College of Medicine, Jinju 52727, Republic of Korea. ² BK21Plus Team for Anti-aging Biotechnology and Industry, Department of Convergence Medical Science, Gyeongsang National University, Jinju 52727, Republic of Korea. ³ Department of Medical Research Pyin Oo Lwin Branch, Pyin Oo Lwin, Myanmar. ⁴ Department of Tropical Medicine and Inha Research Institute for Medical Sciences, Inha University College of Medicine, Incheon 22212, Republic of Korea. ⁵ Department of Microbiology, Ajou University College of Medicine, Suwon 16499, Republic of Korea.

Acknowledgements

We thank the staffs in Department of Medical Research Pyin Oo Lwin Branch and the health professionals in Naung Cho, Pyin Oo Lwin and Tha Beik Kyin townships for their contribution and technical support in blood collection. This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MEST) (NRF-2015K1A3A9A01034893) and the Korea government (MSIP) (NRF-2017M3A9B8069530).

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The data supporting the conclusions of this article are provided within the article and its additional file. The original datasets analysed in the current study are available from the corresponding author upon request. The newly generated sequences were deposited in the GenBank database under the Accession Number MF350670–MF350720.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Ethical approval of this study was obtained from the Ethics Review Committee, Department of Medical Research, Myanmar (1/Ethics/DMRUM/2013 and 97/Ethics 2015) or from the Ethical Review Committee of Inha University School of Medicine, Korea (INHA 15-013). Informed written consent and permission were obtained from each individual enrolled in this study.

Funding

This work was supported by the National Research Foundation of Korea (NRF) Grants funded by the Korea government (MEST) (NRF-2015K1A3A9A01034893) and the Korea government (MSIP) (NRF-2017M3A9B8069530).

Publisher's Note

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

Received: 2 August 2018 Accepted: 8 October 2018

Published online: 12 October 2018

References

1. World Health Organization. World malaria report 2017. Geneva: World Health Organization; 2017.
2. World Health Organization. Malaria in the Greater Mekong Subregion: Regional and country profiles. Geneva: World Health Organization; 2017.
3. World Health Organization. Strategy for Malaria Elimination in the Greater Mekong Subregion (2015–2030). Geneva: World Health Organization; 2015.
4. Kang JM, Cho PY, Moe M, Lee J, Jun H, Lee HW, et al. Comparison of the diagnostic performance of microscopic examination with nested polymerase chain reaction for optimum malaria diagnosis in Upper Myanmar. *Malar J*. 2017;16:119.

5. Agnandji ST, Lell B, Soulanoudjingar SS, Fernandes JF, Abossolo BP, Conzelmann C, et al. First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. *N Engl J Med*. 2011;365:1863–75.
6. Agnandji ST, Lell B, Fernandes JF, Abossolo BP, Methogo BG, Kabwende AL, et al. A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants. *N Engl J Med*. 2012;367:2284–95.
7. Pinzon-Ortiz C, Friedman J, Esko J, Sinnis P. The binding of the circumsporozoite protein to cell surface heparan sulfate proteoglycans is required for *Plasmodium* sporozoite attachment to target cells. *J Biol Chem*. 2001;276:26784–91.
8. Rathore D, Sacci JB, de la Vega P, McCutchan TF. Binding and invasion of liver cells by *Plasmodium falciparum* sporozoites. Essential involvement of the amino terminus of circumsporozoite protein. *J Biol Chem*. 2002;277:7092–8.
9. Coppi A, Natarajan R, Pradel G, Bennett BL, James ER, Roggero MA, et al. The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. *J Exp Med*. 2011;208:341–56.
10. Egan JE, Hoffman SL, Haynes JD, Sadoff JC, Schneider I, Grau GE, et al. Humoral immune responses in volunteers immunized with irradiated *Plasmodium falciparum* sporozoites. *Am J Trop Med Hyg*. 1993;49:166–73.
11. Nardin EH, Nussenzweig RS, McGregor IA, Bryan JH. Antibodies to sporozoites: their frequent occurrence in individuals living in an area of hyperendemic malaria. *Science*. 1979;206:597–9.
12. Calvo-Calle JM, de Oliveira GA, Clavijo P, Maracic M, Tam JP, Lu YA, et al. Immunogenicity of multiple antigen peptides containing B and non-repeat T cell epitopes of the circumsporozoite protein of *Plasmodium falciparum*. *J Immunol*. 1993;150:1403–12.
13. Hughes AL. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): Evidence for positive selection on immunogenic regions. *Genetics*. 1991;127:345–53.
14. Waitumbi JN, Anyona SB, Hunja CW, Kifude CM, Polhemus ME, Walsh DS, et al. Impact of RTS,S/AS02(A) and RTS,S/AS01(B) on genotypes of *P. falciparum* in adults participating in a malaria vaccine clinical trial. *PLoS ONE*. 2009;4:e7849.
15. Bailey JA, Mvalo T, Aragam N, Weiser M, Congdon S, Kamwendo D, et al. Use of massively parallel pyrosequencing to evaluate the diversity of and selection on *Plasmodium falciparum* csp T-cell epitopes in Lilongwe, Malawi. *J Infect Dis*. 2012;206:580–7.
16. Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, Milner DA Jr, et al. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol*. 2008;9:R171.
17. Manske M, Miotto O, Campino S, Auburn S, Almagro-García J, Maslen G, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*. 2012;487:375–9.
18. Zeeshan M, Alam MT, Vinayak S, Bora H, Tyagi RK, Alam MS, et al. Genetic variation in the *Plasmodium falciparum* circumsporozoite protein in India and its relevance to RTS,S malaria vaccine. *PLoS One*. 2012;7:e43430.
19. Snounou G, Singh B. Nested PCR analysis of *Plasmodium* parasites. *Methods Mol Med*. 2002;72:189–93.
20. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–2.
21. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
22. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3:418–26.
23. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
24. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133:693–9.
25. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16:37–48.
26. Sidjanski SP, Vanderberg JP, Sinnis P. *Anopheles stephensi* salivary glands bear receptors for region I of the circumsporozoite protein of *Plasmodium falciparum*. *Mol Biochem Parasitol*. 1997;90:33–41.
27. Escalante AA, Grebert HM, Isea R, Goldman IF, Basco L, Magris M, 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*. 2002;125:83–90.
28. Putaporntip C, Jongwutiwes S, Hughes AL. Natural selection maintains a stable polymorphism at the circumsporozoite protein locus of *Plasmodium falciparum* in a low endemic area. *Infect Genet Evol*. 2009;9:567–73.
29. Zakeri S, Avazalipour M, Mehrizi AA, Djadid ND, Snounou G. Restricted T-cell epitope diversity in the circumsporozoite protein from *Plasmodium falciparum* populations prevalent in Iran. *Am J Trop Med Hyg*. 2007;76:1046–51.
30. Plassmeyer ML, Reiter K, Shimp RL Jr, Kotova S, Smith PD, Hurt DE, et al. Structure of the *Plasmodium falciparum* circumsporozoite protein, a leading malaria vaccine candidate. *J Biol Chem*. 2009;284:26951–63.
31. Rathore D, Nagarkatti R, Jani D, Chattopadhyay R, de la Vega P, Kumar S. An immunologically cryptic epitope of *Plasmodium falciparum* circumsporozoite protein facilitates liver cell recognition and induces protective antibodies that block liver cell invasion. *J Biol Chem*. 2005;280:20524–9.
32. Ancsin JB, 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*. 2004;279:21824–32.
33. Conway DJ. Natural selection on polymorphic malaria antigens and the search for a vaccine. *Parasitol Today*. 1997;13:26–9.
34. Bongfen SE, Ntsama PM, Offner S, Smith T, Felger I, Tanner M, et al. The N-terminal domain of *Plasmodium falciparum* circumsporozoite protein represents a target of protective immunity. *Vaccine*. 2009;27:328–35.
35. Tan J, Sack BK, Oyen D, Zenklusen I, Piccoli L, Barbieri S, et al. A public antibody lineage that potently inhibits malaria infection through dual binding to the circumsporozoite protein. *Nat Med*. 2018;24:401–7.
36. Ferguson DJ, Balaban AE, Patzewitz EM, Wall RJ, Hopp CS, Poulin B, et al. The repeat region of the circumsporozoite protein is critical for sporozoite formation and maturation in *Plasmodium*. *PLoS ONE*. 2014;9:e113923.
37. Garçon N, Heppener DG, Cohen J. Development of RTS,S/AS02: a purified subunit-based malaria vaccine candidate formulated with a novel adjuvant. *Expert Rev Vaccines*. 2003;2:231–8.
38. Kumkhaek C, Phra-ek K, Renia L, Singhasivanon P, Looareesuwan S, Hirunpetcharat C, et al. Are extensive T cell epitope polymorphisms in the *Plasmodium falciparum* circumsporozoite antigen, a leading sporozoite vaccine candidate, selected by immune pressure? *J Immunol*. 2005;175:3935–9.
39. Bhattacharya PR, Bhatia V, Pillai CR. Genetic diversity of T-helper cell epitopic regions of circumsporozoite protein of *Plasmodium falciparum* isolates from India. *Trans R Soc Trop Med Hyg*. 2006;100:395–400.
40. Udhayakumar V, Ongecha JM, Shi YP, Aidoo M, Orago AS, Oloo AJ, et al. Cytotoxic T cell reactivity and HLA-B35 binding of the variant *Plasmodium falciparum* circumsporozoite protein CD8+ CTL epitope in naturally exposed Kenyan adults. *Eur J Immunol*. 1997;27:1952–7.
41. Hill AV, Elvin J, Willis AC, Aidoo M, Allsopp CE, Gotch FM, et al. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature*. 1992;360:434–9.
42. Gordon DM, McGovern TW, Krzych U, Cohen JC, Schneider I, LaChance R, et al. Safety, immunogenicity, and efficacy of a recombinantly produced *Plasmodium falciparum* circumsporozoite protein-hepatitis B surface antigen subunit vaccine. *J Infect Dis*. 1995;171:1576–85.
43. The RTS,S Clinical Trial 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*. 2015;386:31–45.
44. Otieno L, Onoko M, Otieno W, Abuodha J, Owino E, Odero C, et al. Safety and immunogenicity of RTS,S/AS01 malaria vaccine in infants and children with WHO stage 1 or 2 HIV disease: a randomised, double-blind, controlled trial. *Lancet Infect Dis*. 2016;16:1134–44.
45. The RTS,S Clinical Trials Partnership. Efficacy and Safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: A phase 3 randomized, controlled trial in children and young infants at 11 African sites. *PLoS Med*. 2014;11:e1001685.
46. Kaslow DC, Biernaux S. RTS,S: toward a first landmark on the malaria vaccine technology roadmap. *Vaccine*. 2015;33:7425–32.