

RESEARCH

Open Access



# Population genetic and biophysical evidences reveal that purifying selection shapes the genetic landscape of *Plasmodium falciparum* RH ligands in Chhattisgarh and West Bengal, India

Sharmistha Ghoshal<sup>1</sup>, Pramita Chowdhury<sup>1</sup>, Sanhita Ray<sup>1</sup>, Mitashree Mitra<sup>2</sup>, Sumana Datta Kanjilal<sup>3</sup>, Srikanta Sen<sup>4</sup>, Anjan Kr. Dasgupta<sup>1</sup> and Sanghamitra Sengupta<sup>1\*</sup> 

## Abstract

**Background:** Reticulocyte binding protein-like homologs (RHs) are currently being evaluated as anti-erythrocytic stage vaccine targets against *Plasmodium falciparum* malaria. Present study explores the possible evolutionary drivers shaping the genetic organization of *Pfrhs* in Indian parasite population. It simultaneously evaluates a putative gain-of-function variant of PFRH5, a keystone member of PFRH family.

**Methods:** Receptor binding regions of *Pfrh1*, *Pfrh2a/b*, *Pfrh4* and whole *Pfrh5* were amplified using blood samples of *P. falciparum* malaria patients from Chhattisgarh and West Bengal and sequenced. Assembled sequences were analysed using MEGA7 and DnaSPv6. Binding affinities of recombinant PFRH5 proteins with basigin (BSG) were compared using in silico (CHARMM and AUTODOCK) and in vitro (Circular dichroism, fluorescence spectroscopy and isothermal titration calorimetry) methods.

**Results:** *Pfrh1* (0.5), *Pfrh2a/b* (0.875), *Pfrh4* (0.667) and *Pfrh5* (0.778) sequence changes corresponded to low frequency (<0.05) variants which resulted in an overall negative Tajima's D. Since mismatch distribution of none of the *Pfrh* loci corroborated with the model of demographic expansion, a possible role of natural selection formulating *Pfrh* sequence diversity was investigated. Among the 5 members, *Pfrh5* displayed very high dN/dS (5.7) ratio. Nevertheless, the model of selective sweep due to presence of any advantageous substitutions could not be invoked as polymorphic nonsynonymous sites (17/18) for *Pfrh5* exceeded significantly over the divergent (62/86) ones ( $p = 0.0436$ ). The majority of extant PFRH5 sequences (52/83) differed from the reference Pf3D7 allele by a single amino acid mismatch (C203Y). This non-conservative alteration was predicted to lower the total interaction energy of that PFRH5<sub>variant</sub> with BSG, compared to PFRH5<sub>3D7</sub>. Biophysical evidences validated the proposition that PFRH5<sub>variant</sub> formed a more stable complex with BSG. Thermodynamic association constant for interaction of BSG with PFRH5<sub>variant</sub> was also found to be higher ( $K_{a_{variant}} = 3.63E6 \pm 2.02E6 M^{-1}$  and  $K_{a_{3D7}} = 1.31E6 \pm 1.21E6 M^{-1}$ ).

\*Correspondence: sanghamitrasg@yahoo.com

<sup>1</sup> Department of Biochemistry, University of Calcutta, 35, Ballygunge

Circular Road, Kolkata 700 019, West Bengal, India

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



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions:** Together, the study indicates that the genetic architecture of *Pfhrs* is principally shaped by purifying selection. The most abundant and ubiquitous PfrH5 variant harbouring 203Y, exhibits a greater affinity for BSG compared to PfrH5<sub>3D7</sub> possessing 203C allele. The study underscores the importance of selecting the functional allele that best represents circulating strains in natural parasite populations as vaccine targets.

**Keywords:** *Plasmodium falciparum*, Malaria, PfrH, Polymorphism, Selection, Vaccine

## Background

In spite of significant advancement in controlling malaria mortality, the disease particularly the one caused by *Plasmodium falciparum* remains a global health concern. Intervention of malaria is typically based on two regimens those include (a) control of anopheles mosquito vector by the use of insecticidal and (b) treatment of patients using anti-malarial drugs. Drug resistant parasites arising due to wide and indiscriminate deployment of chloroquine, sulfadoxine-pyrimethamine and mefloquine based mono-therapies, is causing resurgence of disease in many parts of the world [1–3]. The extraordinary evolutionary plasticity of *P. falciparum* has not even spared artemisinin (ART) based combination therapy, which currently serves as the first-line treatment for multidrug-resistant *P. falciparum* malaria [4]. Growing incidences of resistance to ART and its derivatives in Africa and South-East Asia are threatening the sustainability of current status of malaria control [4]. Before ART is lost to parasite's genetic adaptability, development of an efficacious vaccine and its widespread application in malarious regions are necessary to deal with rebounding malaria.

Asexual blood-stage of parasite life cycle determines the pathogenesis as well as clinical outcome of *P. falciparum* malaria. Blood stage infection begins with merozoite attachment onto RBC surface and entry inside; and culminates into intra-erythrocytic parasite replication followed by RBC rupture and induction of pro-inflammatory responses in hosts. Merozoite surface proteins that are targets of naturally acquired immunity in individuals with malaria are believed to serve as anti-erythrocytic stage vaccine candidates. Epidemiological studies have also demonstrated the feasibility of such vaccines by presenting the evidences that people living in malaria endemic areas can gradually acquire immunity against both severe malaria and clinical malaria [5, 6].

During the multi-step progression of RBC invasion, direct attachment of merozoites with erythrocytes is mediated by two protein families: *P. falciparum* reticulocyte binding protein-like homolog (PfrH) and erythrocyte binding antigen (PfeBA). Except PfrH5, both classes of ligands are transmembrane proteins which are employed redundantly by parasites for successful RBC invasion [7]. PfrH protein family is composed of five members namely PfrH1, PfrH2a, PfrH2b, PfrH4

and PfrH5. PfrH members do not share much sequence homology, although PfrH2 and PfrH5 are assumed to adopt a similar 'Kite-like' conformation with receptor binding sites as deduced by homology modelling [8, 9]. Although the identity of the other PfrH receptors remains to be established, complement receptor 1 (CR1) and basigin (BSG), have been shown to function as the erythrocyte receptors for PfrH4 and PfrH5, respectively [7, 8, 10]. In an effort to develop a blood-stage vaccine, *P. falciparum* 3D7 growth was shown to be inhibited by antibodies from rabbits immunized with a combination of EBA-175, PfrH2a/b, and PfrH4 [11]. Furthermore, expression and polymorphism analyses of parasite field isolates reveal PfrH2a, PfrH5 and EBA-181 to be key players of invasion pathway [10].

Existence of geographical variations in the invasion ligands suggests that the idea of one-size-fits-all type of vaccine needs to be revisited and emphasizes on the necessity of genetic and functional assessments of vaccine candidates. On this note and given the unequivocal importance of PfrH ligands in erythrocyte invasion, the present study takes an account of the genetic diversity of all 5 PfrH ligands in the *P. falciparum* field isolates from two malaria-prone regions of India, one of the large reservoirs of *P. falciparum* strains. It has been recently shown that PfrH5-BSG binding is essential for parasite's RBC invasion as this results in a bridge between parasite and erythrocyte through which invasion elements flow [12]. Therefore, in addition to dissecting the factors responsible for extant pattern of sequence divergence of *Pfhrs*, the functional capacity of one ubiquitously observed PfrH5 variant (C203Y) has been analysed with regard to its binding with RBC receptor BSG. Altogether the knowledge accrued from this study will be useful to comprehend parasite's evolutionary dynamics under present day drug regimens and unfold the mechanism of genetic adaptability of this important human pathogen.

## Methods

### Study area, sample collection and DNA extraction

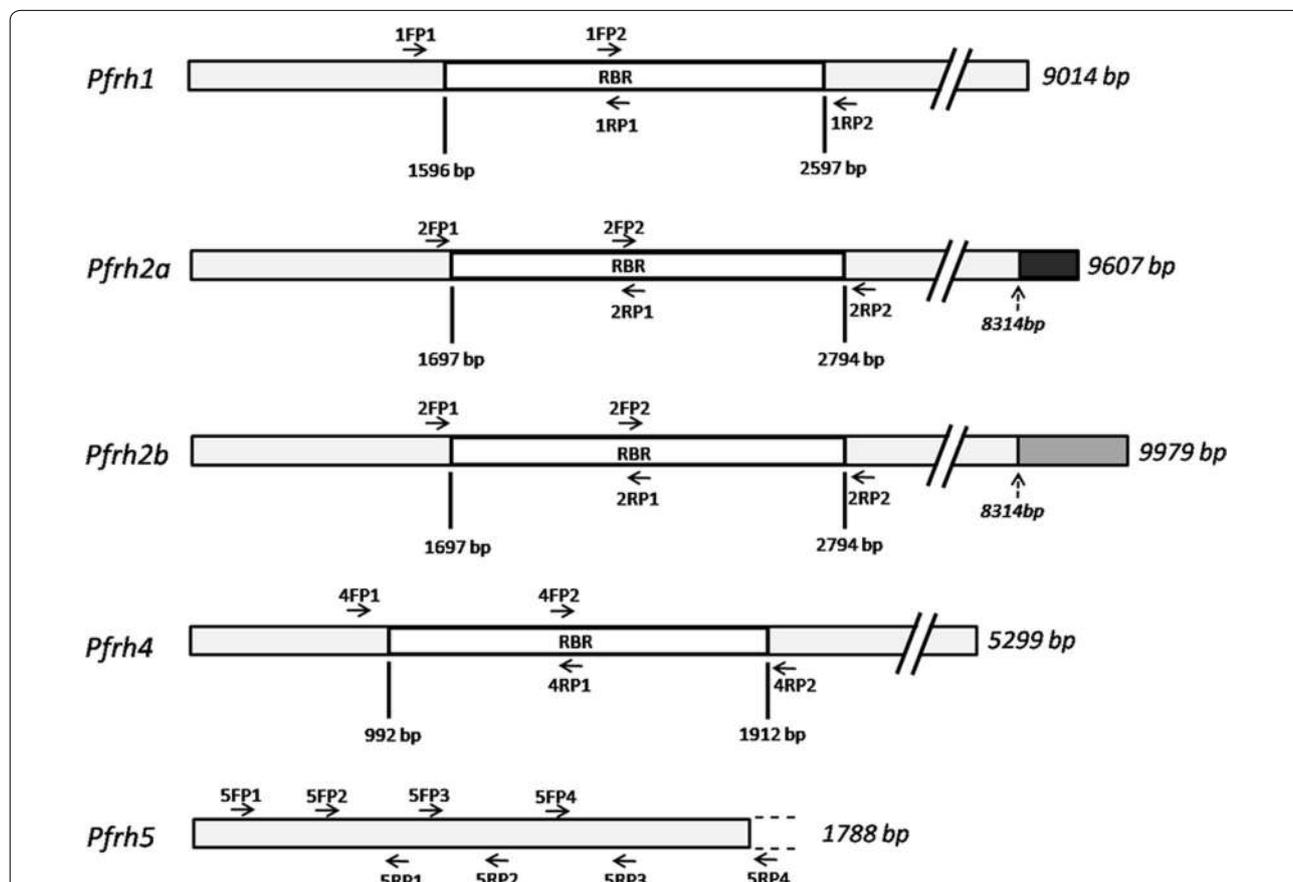
Stored genomic DNAs previously isolated from peripheral blood samples collected from *P. falciparum* malaria infected patients admitted in Calcutta National Medical College and Hospital, Kolkata, in the year 2010 were utilized in this study [13]. Kolkata is the capital of West

Bengal. Almost 10% of the total malaria cases in India are accounted in West Bengal [14]. Peripheral venous blood samples were also collected from Surguja located in Chhattisgarh, central India, during a period of 2010–2013. Surguja, incidentally a tribal dominated district, is situated in the forested area of Chhattisgarh [15]. Due to its distinct ecological and geographical conditions, Chhattisgarh contributes to ~12% of total malaria burden and carries the highest share of deaths (17%) in India [16]. Only blood samples from *P. falciparum* malaria positive patients confirmed by rapid diagnostic tests based on dual-Antigen and/or Giemsa-stained thick and thin smears were selected for this study [15]. Individuals suffering from co-infection with *Plasmodium vivax* and pregnant women were excluded. Additionally, the study excluded the patients suffering from chronic or severe disease conditions, such as cardiac, renal or hepatic diseases, G6PD deficiency, typhoid, measles, acute lower respiratory tract infection, bacteraemia, severe diarrhoea with dehydration, sickle cell anaemia, AIDS and cancer.

Genomic DNA was isolated from the blood sample of *P. falciparum* malaria infected patients using QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany) according to manufacturer’s protocol.

**Polymerase chain reaction, purification of PCR amplicons and sequencing of PfrHs**

Oligonucleotide primers for polymerase chain reaction and sequencing of genes encoding PfrHs were designed by retrieving reference sequences of *P. falciparum* 3D7 from PlasmoDB (Fig. 1) [17]. Receptor binding sites of PfrH1 and PfrH4 were amplified using two pairs of overlapping primers for each [18–20]. PfrH2a and PfrH2b possess indistinguishable ecto-domain containing the putative receptor binding site (495-860 amino acid region of PfrH2a/b in *Pf3D7*) [18, 19]. This identical receptor binding site of PfrH2a/b was considered for primer designing and amplification. An attempt was made to amplify the whole PfrH5 gene using four overlapping primer pairs. The primer sequences as well



**Fig. 1** Schematic representation of genes encoding PfrH ligands. The regions subjected to sequence analysis were shown. Regions mentioned as RBR signified receptor binding region. Solid arrows represented positions of the primers used to amplify parasite genomic DNA. Arrows with broken line indicated the non-identical regions of *Pfrh2a* and *Pfrh2b*. Sequences corresponding to RBRs were analysed for *Pfrh1*, *Pfrh2a/b* and *Pfrh4*, while the entire coding region of *Pfrh5* was subjected to sequence analysis

as cycling conditions were described in the Additional file 1: Table S1. Amplification of target regions were performed in 15  $\mu$ L reaction mixtures containing 1 U of GoTaq<sup>®</sup> Flexi DNA polymerase (Promega), 0.2 mM dNTP, 1.5 mM MgCl<sub>2</sub>, and 0.4  $\mu$ M of each primer using a GeneAmp<sup>®</sup> PCR System 9700 (Applied Biosystems). UV transillumination on gel documentation system (Biostep) was used to visualize the PCR products preceded by electrophoresis on 2% agarose gel (Promega). PCR products were purified by Qiaquick gel extraction kit (QIAGEN India Pvt. Ltd, Hilden, Germany) and sequenced using the same primers utilized in PCR. Sequencing PCR was performed using forward and reverse primers (separately) and Big Dye v3.1 dye terminator on ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) [13]. Number of samples analysed for each target gene were mentioned in Table 1.

#### Sequence alignment and data analysis of Pfrh family members

Raw sequence data files from field isolates were manually revised to exclude indefinite reads and signal noise. To compare the sequence identity, NCBI BLAST analysis was performed with all test sequences with respect to reference sequences (Gene ID of *Pfrh1*: PF3D7\_0402300; *Pfrh2a*: PF3D7\_1335400; *Pfrh2b*: PF3D7\_1335300; *Pfrh4*: PF3D7\_0424200; *Pfrh5*: PF3D7\_0424100) [21]. MEGA 7 tool was used to perform multiple sequence alignment of the nucleotide sequences and to translate

them into amino acid codes. Nucleotide sequences analysed in this study were submitted to the GenBank database under the accession numbers MN937274-MN937320, MN950796-MN950967.

Variants were identified by aligning the sequence reads with corresponding reference sequences. Varied number of single nucleotide polymorphisms (SNPs) was observed in both study zones. Genetic diversity parameters including (i) number of segregating (S) sites, (ii) average number of pairwise nucleotide differences within population (k), (iii) average number of observed nucleotide differences per site between any two sequences believing that the sample was randomly selected ( $\pi$ ), (iv) Watterson's  $\theta$  ( $\theta_w$ ) were estimated using DnaSPv6.00 [22, 23]. The estimation of the recombination phenomena represented by the minimum number of recombination events (Rm) that happened along the isolates was also performed. Tajima's D and Fu & Li's statistics were deployed using DnaSP to assess the neutral theory of evolution [24–26]. To gauge the impact of natural selection over gene sequences the ratio of non-synonymous to synonymous substitutions (dN/dS) is commonly used [27]. The method of Nei and Gojobori with the Jukes and Cantor correction was used to assess these rates of substitution within species as implemented in the MEGA7 [28, 29]. Five hundred bootstrap replications were applied to estimate p values as well as standard error. To derive an estimate of how did parasite sequence data generated in this study differ from those available in global data, particularly those from

**Table 1 Diversity estimates of Pfrhs based on sequence variants in parasite isolates**

Gene name	<i>Pfrh1</i>	<i>Pfrh2a/b</i>	<i>Pfrh4</i>	<i>Pfrh5</i>
No of samples (n)	47	43	46	83
Length sequenced (bp)	999	1092	921	1320
Polymorphic sites	4	8	6	18
Polymorphic Sites with frequency < 5%	2	7	4	14
Parsimony informative sites	4	2	2	8
No of haplotypes (H)	5	7	7	15
Haplotype diversity (Hd $\pm$ SD)	0.660 $\pm$ 0.052	0.532 $\pm$ 0.074	0.527 $\pm$ 0.071	0.594 $\pm$ 0.061
Average number of pairwise differences (k)	0.809	0.802	0.995	1.234
Raggedness index	0.182	0.108	0.482	0.055
Recombination event	0	0	0	2
Nucleotide diversity ( $\pi$ $\pm$ SD)	0.0008 $\pm$ 0.0001	0.0007 $\pm$ 0.0002	0.0011 $\pm$ 0.0002	0.0002 $\pm$ 0.0002
Watterson's $\theta$ $\pm$ SD	0.0009 $\pm$ 0.0004	0.0017 $\pm$ 0.0006	0.0015 $\pm$ 0.00061	0.0027 $\pm$ 0.0006
Tajima's D	-0.246	-1.592	-0.702	-1.922 <sup>a</sup>
Fu & Li's D <sup>a</sup>	1.011	-2.837*	-2.204	-2.915*
Fu & Li's F <sup>a</sup>	0.731	-2.866*	-1.912	-3.042*
Average number of non-synonymous sites, dN (SE)	0.0007 $\pm$ 0.00002	0.0009 $\pm$ 0.000041	0.0006 $\pm$ 0.00002	0.001 $\pm$ 0.00002
Average number of synonymous sites, dS (SE)	0.0011 $\pm$ 0.00006	0	0.003 $\pm$ 0.0001	0.0002 $\pm$ 0.00001
dN/dS	0.7	-	0.2	5.8

<sup>a</sup> indicates p < 0.05

sub-Saharan Africa, *Pfprh* sequences were retrieved from NCBI and diversity analysis was performed (Additional file 2: Table S2). Further, the evolutionary connections among nucleotide haplotypes of each *Pfprh* family members were predicted using Network program [27, 30–33].

#### Ligand receptor interaction: an in silico analysis

To analyse the impact of genetic variability on ligand-receptor interaction efficacy, the most prevalent PfrH5 haplotype (rh5h1) was selected to compare with its 3D7 type counterpart (PfrH5<sub>3D7</sub>). C203 residue in PfrH5<sub>3D7</sub> was replaced by Y203 in the most prevalent PfrH5 haplotype rh5h1. 4u0q.pdb protein file representing complex of rh5h1 (hereafter, mentioned as PfrH5<sub>M</sub>) with its erythrocyte surface receptor BSG was extracted from RCSB and the interaction efficiency was compared with that of PfrH5<sub>3D7</sub> and BSG [34]. In order to introduce C203 residue, the side-chain of residue number 203 (Y) of ligand was removed from the 4u0q.pdb file and the residue name of the backbone parts was changed from Y to C. Thereafter, AUTODOCK tool was applied to construct the side chains of the newly modified residue [35]. The pdb file of the PfrH5<sub>3D7</sub> - BSG complex was thus generated. Both the interaction complexes were refined by minimizing their energies separately in equal number of steps until the difference in the energy between the last two steps is lower than 10% of the preceding step. Afterwards, the effect of the mutation was studied by comparing the residue-wise interaction patterns in the two energy minimized complexes and the H-bonding patterns of two (C/Y) residues at 203 in those structures.

To compute the residue-wise interaction pattern between ligand and receptor, partial atomic charges were assigned to all the atoms of the both complexes using CHARMM. Interaction of the atoms of individual residues in ligand with whole BSG was computed considering two components (i) electrostatic interaction and (ii) van der Waals interaction. Residue-wise interaction in the two complexes and the H-bonding patterns of the residue 203 in the two complexes were also studied.

#### Expression of recombinant proteins

To validate the result obtained from in silico analysis, the binding interaction of PfrH5 with BSG was studied in further details. A 564-bp fragment of the *Pfprh5* gene encoding the 188 aa (Glutamine-35—Serine-222 residues) which encompasses receptor binding site, was PCR amplified from DNA of a *P. falciparum* infected blood sample possessing PfrH5<sub>M</sub> type sequence using forward primer: **CGCGGATCCCCAAGAAAATAATCTGAC** and reverse primer: **ACATATGACAAAGTGAAAAGT AAGCTTGCG** [36]. Polymerase chain reaction was carried out in 15 µL reaction mixtures having 1 U of GoTaq®

Flexi DNA polymerase (Promega), 0.2 mM dNTP, 1.5 mM MgCl<sub>2</sub> and 0.4 µM of each primer. The cycling conditions for PCR consisted of an initial denaturation at 94° C for 5 min, followed by 38 cycles of denaturation at 94° C for 45 s, annealing at 58° C for 45 s, extension at 72° C for 45 s, and a final extension at 72° C for 5 min. PCR product was visualized using UV transillumination on gel documentation system (BioStep) following electrophoresis on 2% agarose gel (Promega) and were purified by Qiaquick gel extraction kit (QIAGEN India Pvt. Ltd, Hilden, Germany). The PCR amplicon was digested with *Bam*H I and *Hind* III (New England Biolabs, Beverly, MA) and inserted downstream of the T7 promoter in the *E. coli* expression vector, pET-20b(+) (Novagen, San Diego, CA) using a T4 DNA ligase (Fermentas), to obtain the plasmid rPfrH5<sub>M</sub>- pET-20b. Site-directed mutagenesis (SDM) was performed on the plasmid rPfrH5<sub>M</sub>- pET-20b to insert C in exchange of Y at 203 amino acid position to generate rPfrH5<sub>3D7</sub>- pET-20b using the SDM-specific primer SDMF: 5'-GTCCTCTACATATGGAAA GTGTATAGC-3' and SDMR: 5'-AGCATCTACAGC TATACACTTTCCAT-3'. The transcribed sequence of the recombinant plasmids (rPfrH5<sub>3D7</sub>- pET-20b and rPfrH5<sub>M</sub>- pET-20b) contained an additional His-tag (LEHHHHHH) at the C terminus. *E. coli* BL21 (DE3) cells (Novagen, San Diego, CA) were transformed with rPfrH5<sub>3D7</sub>- pET-20b and rPfrH5<sub>M</sub>- pET-20b and used for the expression of rPfrH5<sub>3D7</sub> and rPfrH5<sub>M</sub>. Sequencing of the plasmid chimera was used to confirm that two different *Pfprh5* alleles were inserted in correct reading frame.

Gene expression was performed in a 100 ml culture, using Luria–Bertani medium containing 100 mM ampicillin at 37 °C. Pilot experiments determined that 150 ml bacterial culture incubated for 16 h at 37 °C gave optimal expression of recombinant parasite proteins in culture medium. Once the optical density at 600 nm reached 0.6, the culture was incubated overnight at shaking condition (160 rpm) at 37 °C for protein expression. Cells were harvested by centrifugation next day and the cell pellet was stored at –80 °C. The recombinant protein of the expected molecular mass (24.5 kDa) was present in a total cell lysate.

#### Purification of His-tagged recombinant protein by Ni–NTA affinity chromatography

The frozen cell pellet was resuspended in 800 ml of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole (Qiagen); pH 8.0), mixed at 4 °C for 1 h, and lysed by ultrasound technology (UP200S Ultrasonic Processor; 0.7 cycle; 80% Amplitude), followed by centrifugation at 12,000 rpm for 20 min. Supernatant containing protein was applied on 133 µl of nickel Ni–NTA resin (Qiagen)

equilibrated with lysis buffer. To remove non-specifically bound proteins, 1 ml of the wash buffer (pH 7) containing 50 mM imidazole was used. Bound proteins were eluted with elution buffer containing 500 mM imidazole. Eluted fractions were analysed by SDS-PAGE and Western blotting. Concentrations of the fractions containing recombinant PfrH5 was determined by Nanodrop (Thermo Scientific) and stored at  $-80^{\circ}\text{C}$ .

#### SDS-PAGE and Western blotting

Crude lysates as well as purified proteins were separated in the presence of SDS using 15% polyacrylamide gel and visualized with Coomassie<sup>®</sup> Brilliant blue R 250 (Merck). Approximately, 30–50  $\mu\text{g}/\text{lane}$  purified protein was electrophoresed by SDS-PAGE using Prism Ultra Protein Ladder (ab116028, abcam; 10–245 kDa) and transferred to polyvinylidene fluoride membrane (Pall Corporation, Port Washington, USA). Blots were treated with primary (mouse monoclonal antibody directed against 6X His tag; ab18184, Abcam) and secondary antibodies and incubated with electrochemiluminescence (ECL) reagent (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific) and examined on Bio-Rad-Molecular imager ChemiDoc<sup>™</sup> XRSt with ImageLab<sup>™</sup> software.

#### Circular dichroism analysis, Fourier transform infrared spectroscopy (FTIR), synchronous fluorescence and fluorescence anisotropy

The folding state of the purified proteins was evaluated by circular dichroism (CD) spectroscopy. CD spectra were recorded on a J-815 CD Spectrometer (Jasco) at  $25^{\circ}\text{C}$  with the path length of 1 mm. The protein sample solution was diluted in PBS (pH 7.4) so that a final protein concentration was 2  $\mu\text{M}$ . Secondary structural changes were recorded in the range of 190–260 nm. Each spectrum represented an average of three scans and CD data were expressed as mean residue molar ellipticity  $[\theta]$ .

FTIR was carried out using PerkinElmer spectrum 100 FT-IR Spectrometer at a resolution of  $0.5\text{ cm}^{-1}$  in the wavenumber range  $1000\text{--}4000\text{ cm}^{-1}$ . 10  $\mu\text{l}$  of solution containing 1  $\mu\text{g}/\mu\text{l}$  protein was utilized for each FTIR analysis.

Synchronous fluorescence patterns of PfrH5<sub>3D7</sub>, PfrH5<sub>M</sub>, PfrH5<sub>3D7</sub>-BSG and PfrH5<sub>M</sub>-BSG were observed in different  $\Delta\lambda$  at  $25^{\circ}\text{C}$  [37, 38]. Since, in synchronous fluorescence, maximum peak intensity was shown at 379 nm while  $\Delta\lambda$  was 70, fluorescence anisotropy of BSG, PfrH5<sub>3D7</sub>, PfrH5<sub>M</sub>, PfrH5<sub>3D7</sub>-BSG (1:1), and PfrH5<sub>M</sub>-BSG (1:1) were performed in the range between 309 and 379 nm ( $\Delta\lambda=70$ ) for 300 s with both excitation and emission bandwidth of 5 nm using a Varian Cary Eclipse (USA) spectrofluorimeter [39].

#### Isothermal titration calorimetry

Isothermal calorimetry is a quantitative tool used to evaluate thermodynamic profile of interaction between two or more molecules in solution, while finding out equilibrium binding constant ( $K_a$ ) [40]. In this study, BSG was added drop by drop in a fixed volume with 1 min time interval in the sample cell containing either PfrH5<sub>3D7</sub> or PfrH5<sub>M</sub> in a VM2: Cell - ITC 200 (micro cal) titration micro-calorimeter. Concentration of BSG was 10 times higher than PfrH5. Calorimetric reference cell was loaded with 200  $\mu\text{l}$  of Dulbecco phosphate buffer saline solution (pH 7.4). Titration curve for PBS-BSG interaction was subtracted from the heat of binding reaction of PfrH5-BSG to obtain the effective heat of binding. The resulting titration curves were fitted for one set of binding site using MicroCal Origin software.

## Results

### Sequence diversity of genes encoding PfrH ligands

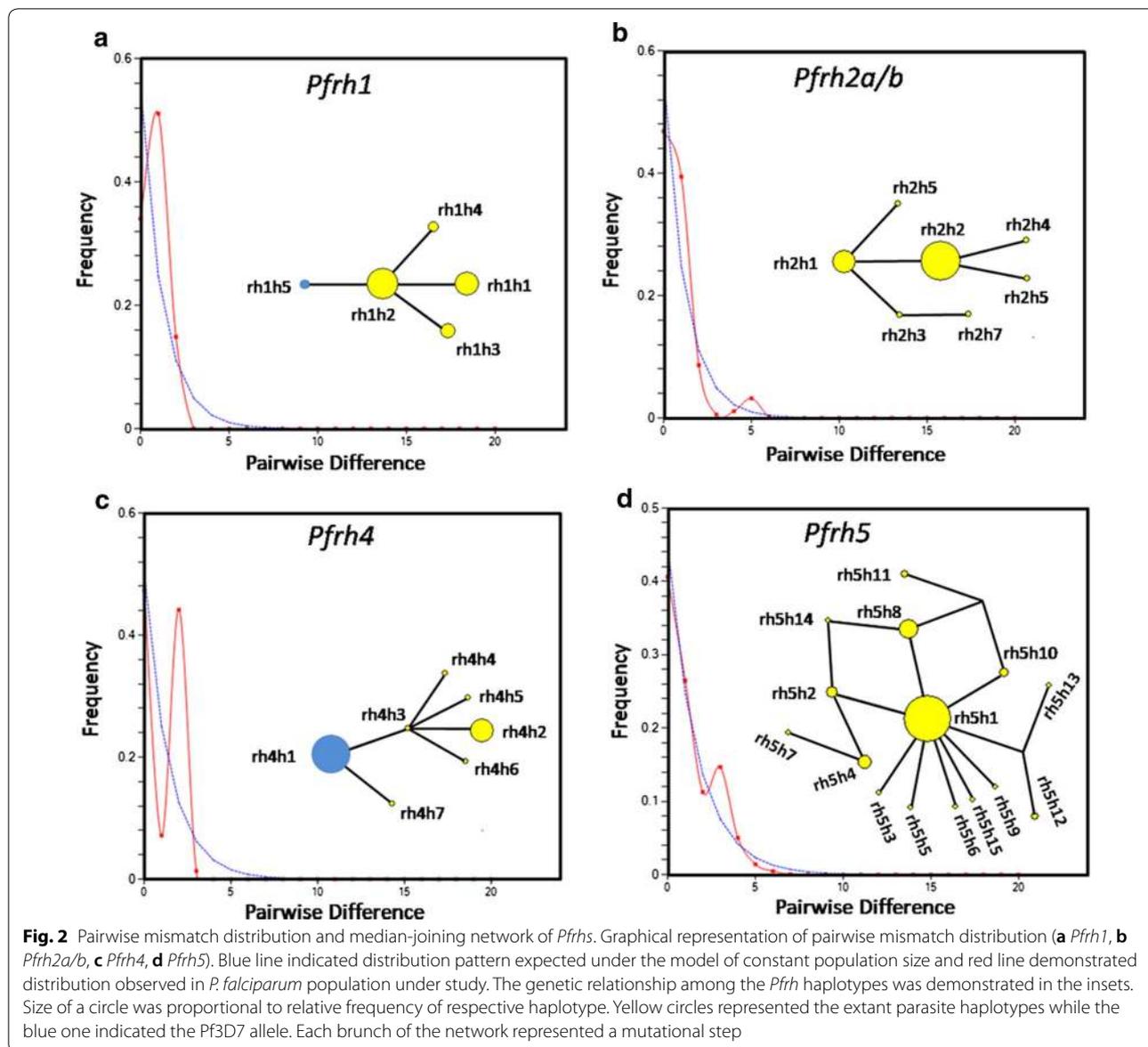
To infer the nature of the evolutionary forces shaping the genetic landscape of *Pfrh* members, fragments encoding *Pfrh1*, *Pfrh2a*, *Pfrh2b*, *Pfrh4* and *Pfrh5* were amplified from genomic DNA isolated from peripheral blood samples of malaria patients. Overlapping sequences from each sample were assembled before estimating the genetic diversity parameters.

#### *Pfrh1*

Analysis of 1 kb DNA sequences of *Pfrh1* that corresponded to the receptor binding region (500 to 832 amino acids as per PlasmoDB) from 47 malaria patients (15 from Chhattisgarh and 32 from West Bengal) identified 4 Parsimony informative sites of which 2 displayed a frequency  $>5\%$ . This resulted in an overall low genetic diversity ( $Hd=0.660\pm 0.052$ ; Watterson's  $\theta=0.0009\pm 0.0005$  and  $\pi=0.0008\pm 0.0001$ ) for the locus. Analysis of median-joining network revealed that a single parasite haplotype (rh1h2) different from the 3D7 type variant (rh1h5) dispersed in the extant population. The demographic history inferred from mismatch distribution pattern revealed no evidence for historical population size fluctuation or recent population expansion (Table 1, Fig. 2a).

#### *Pfrh2a/b*

Genes encoding PfrH2a and PfrH2b shared a common ectodomain sequence that translated into a region spanning 495 to 860 amino acids which was responsible for binding to the receptor. DNA samples from 43 malaria patients (21 from Chhattisgarh and 22 from West Bengal) were employed for the analysis. Unlike other *Pfrhs*, 4 indel polymorphisms with frequencies



ranging from 0.023 to 0.81 were detected in addition to 8 single base alterations. Only 2 of these 8 segregating sites were Parsimony informative and 7 had frequencies < 5%. Despite the predominance of low-frequency polymorphisms that yielded negative Tajima's D (−1.59174) and Fu & Li's D\* and F\* statistics (Table 1), the observed mismatch distribution did not deviate from the one expected under demographic expansion (Fig. 2b). Median-joining network analysis disclosed that all of the observed *Pfrh2a/b* haplotypes were evolved from the two most frequent haplotypes namely rh2h1 and rh2h2. The extant parasite population lacked any haplotypes corresponding to Pf3D7 sequence (Fig. 2b).

***Pfrh4***

Sequence data representing 282 to 588 amino acid region of *PfRH4* were generated using 46 malaria samples (18 from Chhattisgarh and 28 from West Bengal). Application of DnaSP6 resulted in 6 segregating sites and 7 haplotypes (Hd = 0.527 ± 0.071). The most predominant haplotype, rh4h1, resembled *Pf3D7* allele and was present in both study regions. Presence of only 2 Parsimony informative sites and an excess of low frequency (4 with frequency < 5%) variants yielded a negative Tajima's D and Fu & Li's statistics (Table 1). The bimodal mismatch distribution (raggedness index = 0.4819) persisted even when the analysis was replicated using sequences from individual

study region and was caused by coexistence of multiple high-frequency parasite haplotypes (Fig. 2c). Network analysis showed that *Pf3D7* (rh4h1) to be the most predominant allele (Frequency = 0.65) followed by rh4h2 (Frequency = 0.24) (Fig. 2c).

### ***Pfhr5***

Genomic DNAs isolated from blood samples of 83 malaria patients (21 from Chhattisgarh and 62 from West Bengal) were used to analyse the diversity of *Pfhr5* locus. Evaluation of a 1320 bp sequence covering 45 to 484 amino acid region of *PfRH5* revealed presence of 18 segregating sites giving rise to 15 haplotypes ( $Hd = 0.594 \pm 0.061$ ). Fourteen segregating sites had a frequency < 5% yielding a statistically significant negative Tajima's  $D$  (-1.92225) and Fu & Li's  $D^*$  (-2.91441) and  $F^*$  (-3.04218) statistics (Table 1). The large excess of singletons ( $n = 10$ ) produced a star-like median-joining network of *PfRH5* haplotypes emanating from rh5h1 possessing a single non-synonymous variation (G > A at 815 nucleotide position) (Fig. 2d). The observed pattern of mismatch distribution adhered strongly to the standard neutral model of panmictic population with constant size (raggedness index = 0.0698), (Fig. 2d). Seventeen out of 18 sites were non-synonymous in character yielding a very high dN/dS ratio of 5.8. However the application of McDonald-Kreitman (MK) test using *P. falciparum* and *P. gaboni* (retrieved from PlasmoDB) data showed a significantly higher ( $\chi^2 = 4.072$ ;  $p = 0.043$ ) ratio of nonsynonymous to synonymous amino acid variation within species (17:1) compared to that between species (62:24). This together with the observed reduction of the nucleotide diversity reinstated that purifying selection influenced the currently observed genetic architecture of *PfRH5*. Interestingly 13 out of 15 haplotypes possessed the variation that corresponded to C203Y non-synonymous change on *Pf3D7* background. Whether this polymorphism offered any gain-of-function from the angle of *PfRH5*-BSG interaction was next investigated.

### **Comparison of the *Pfhr* sequences**

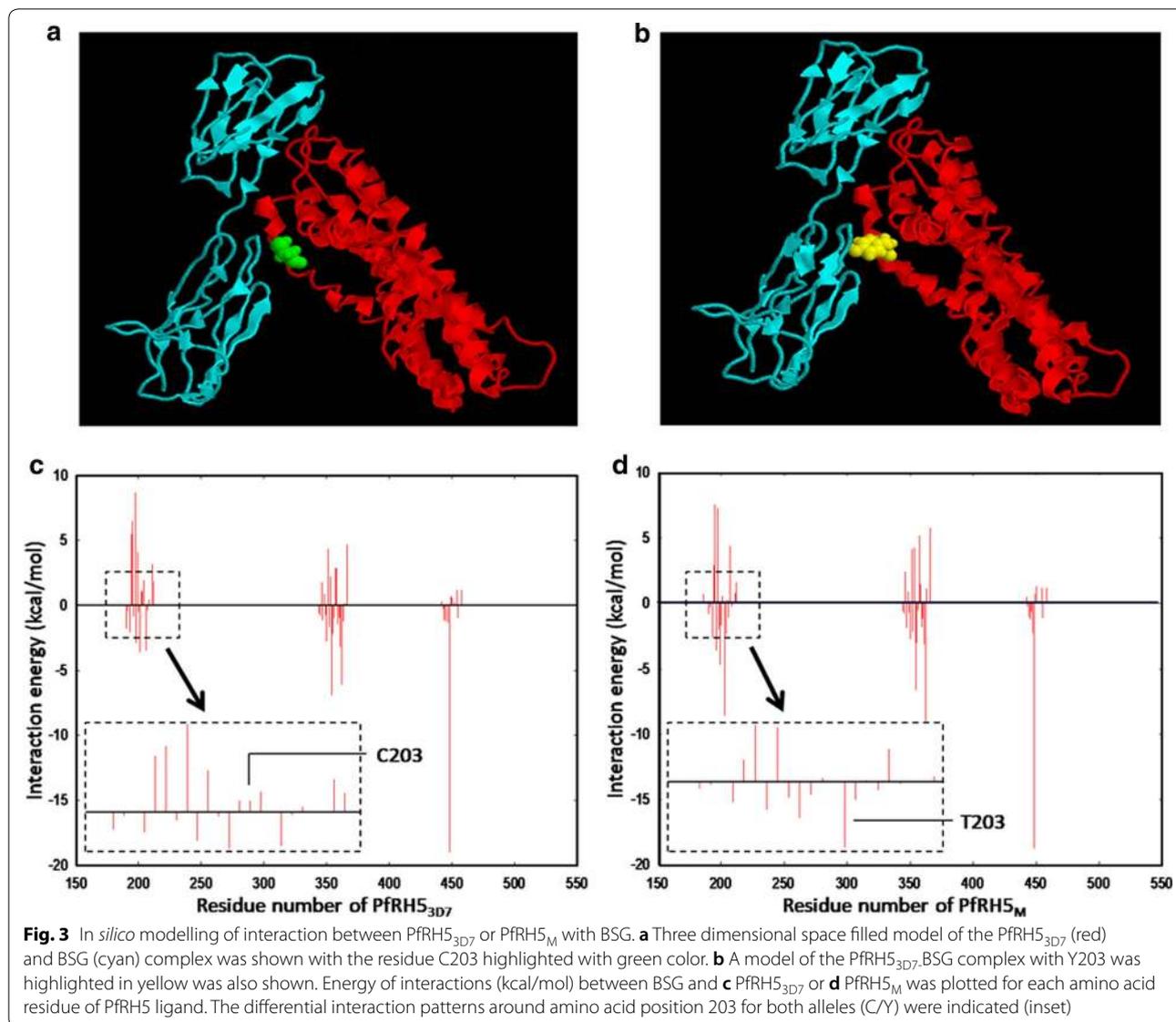
A comparison of the sequence data generated in this study was carried out by using sequences from central India, Papua New Guinea, Kenya and Mali (Additional file 2: Table S2.). Despite the excess of low frequency sites, a statistically significant genetic differentiation ( $F_{ST}$ ) was observed for majority of the loci and geographical regions analysed (Additional file 2: Table S2.). Analysis of *Pfhr5* sequences from Kenya and Mali revealed a higher abundance of C203Y (0.9 and 0.67, respectively) variation as observed in data from Chhattisgarh and West Bengal of the present study (Additional file 3: Fig. S1.) [41–44].

### **In silico evaluation of C203Y variation on *PfRH5*-BSG interaction**

To understand the exact role of C203Y, a ubiquitously observed nonsynonymous variation, the efficacy of interaction of BSG with *PfRH5* encoded by the most prevalent (frequency = 0.63) haplotype, rh5h1 (hereafter mentioned as r*PfRH5<sub>M</sub>*) was compared to that encoded by the reference *Pf3D7* allele (r*PfRH5<sub>3D7</sub>*). Protein file, 4u0q.pdb, representing a complex of r*PfRH5<sub>M</sub>* with BSG was extracted from RCSB protein data bank. A pdb file containing only r*PfRH5<sub>M</sub>* was produced using AUTODOCK followed by generation of r*PfRH5<sub>3D7</sub>* pdb from the one containing r*PfRH5<sub>M</sub>*. Ribbon models constructed using pdb files of r*PfRH5<sub>3D7</sub>*-BSG and r*PfRH5<sub>M</sub>*-BSG complexes indicated that the Y203 was closer to BSG than C203 (Fig. 3a, b). The interaction efficiencies of two *PfRH5* variants with BSG estimated using CHARMM showed that the presence of Y203 resulted in a lowering of total self-energy of r*PfRH5<sub>M</sub>*-BSG complex (-11200 kcal/mol) compared to that of r*PfRH5<sub>3D7</sub>*-BSG (-11127 kcal/mol) indicating that the former was a more stable complex. This was validated by estimation of interaction energy of individual residues of r*PfRH5<sub>3D7</sub>* and r*PfRH5<sub>M</sub>* in the energy minimized complexes of r*PfRH5<sub>M</sub>*/r*PfRH5<sub>3D7</sub>* with BSG (-10.9 kcal/mol for *PfRH5<sub>3D7</sub>*-BSG and -29.9 kcal/mol for *PfRH5<sub>M</sub>*-BSG) (Fig. 3c, d). Residue-wise interaction energies varied not only at amino acid position, 203, but in neighbouring positions as well. This signified different conformational arrangement of r*PfRH5<sub>3D7</sub>* and r*PfRH5<sub>M</sub>* at the time of interaction with BSG (Fig. 3c, d).

### **Purification of soluble, biochemically active recombinant *PfRH5* proteins**

DNA sequence representing the receptor binding region of *PfRH5* which spanned amino acids 141 to 756 of the protein and harboured Y203 (r*PfRH5<sub>M</sub>*) allele was PCR amplified from genomic DNA isolated from one *P. falciparum* infected blood sample. The PCR product was cloned downstream to T7 promoter in the *E. coli* expression vector pET-20b (+) in the correct reading frame. C203 (r*PfRH5<sub>3D7</sub>*) allele was obtained by site-directed mutagenesis using the Y203 containing clone. The chimeric plasmids were mentioned as r*PfRH5<sub>3D7</sub>*-pET-20b and r*PfRH5<sub>M</sub>*-pET-20b. Recombinant proteins (r*PfRH5<sub>3D7</sub>* and r*PfRH5<sub>M</sub>*) of expected molecular mass (24.5 kDa) detected in cell lysates were purified by one-step affinity chromatography on Ni-NTA resin (Fig. 4a). Fractions eluted with 500 mM imidazole were separated by SDS-PAGE and detected by Coomassie staining and Western blotting using anti-His tagged antibody (Fig. 4b, c). Yields of purified *PfRH5* proteins were 1 mg/ml of *E. coli* culture.



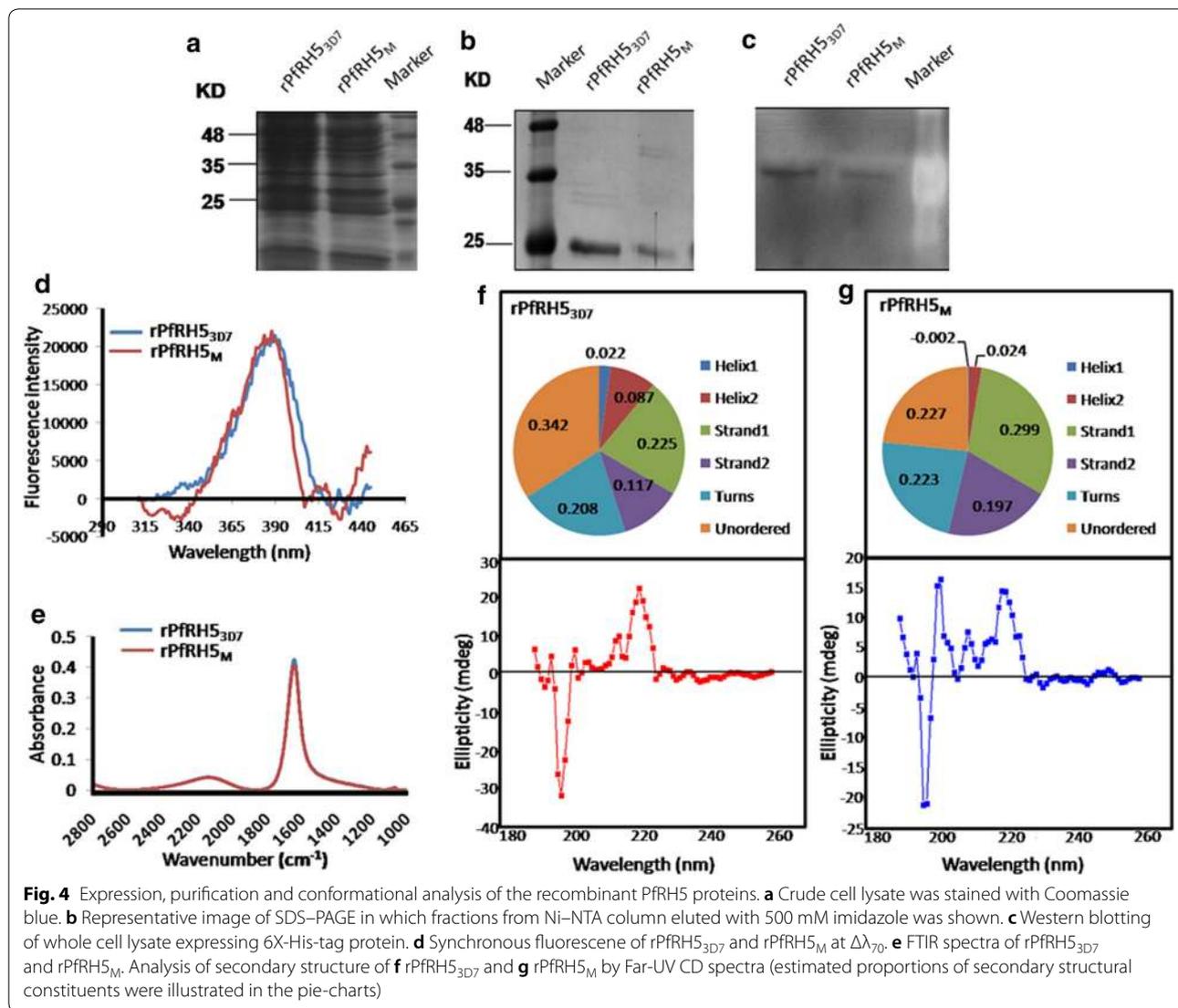
### Conformation analysis of rPfrH5 proteins

Synchronous fluorescence of rPfrH5<sub>3D7</sub> and rPfrH5<sub>M</sub> were measured at a range of wavelength differences ( $\Delta\lambda$ ) (Additional file 4: Fig. S2.). Synchronous fluorescence of both proteins showed peak with maximum intensity at  $\Delta\lambda_{70}$ . The spectral pattern at  $\Delta\lambda_{70}$  did not show any differences between the two variants (Fig. 4d). FTIR spectra of rPfrH5<sub>3D7</sub> and rPfrH5<sub>M</sub> also did not differ from each other (Fig. 4e). Variation in folding states of purified rPfrH5 proteins detected as CD data were analysed by SELCON3 programme in Dichroweb (Fig. 4f, g) [45]. CD spectra of rPfrH5<sub>M</sub> showed an increase in the proportion beta-sheet (49.6% vs 34.2%) at the expense of alpha-helix (2% vs 10.9%) and random coil (22.7% vs 34.2%) compared to that of rPfrH5<sub>3D7</sub>. Secondary structural

components of rPfrH5<sub>M</sub> differed significantly than that of rPfrH5<sub>3D7</sub> ( $p=0.02$ ) which appeared to be a result of C to Y replacement.

### Characterization of binding affinity of PfrHs with BSG

To explore the conformation of two ligand-receptor complexes, FTIR, synchronous fluorescence, fluorescence anisotropy UV CD spectroscopy were applied. No significant differences in backbone conformation could be identified when PfrH5<sub>3D7</sub>-BSG and rPfrH5<sub>M</sub>-BSG complexes were subjected to FTIR spectroscopy (Fig. 5a). However, contrary to its unbound state, rPfrH5<sub>M</sub>-BSG exhibited a higher fluorescence peak intensity at  $\Delta\lambda_{70}$ , indicating that either Y203 change or exposure of aromatic amino acids in the rPfrH5<sub>M</sub>-BSG complex was



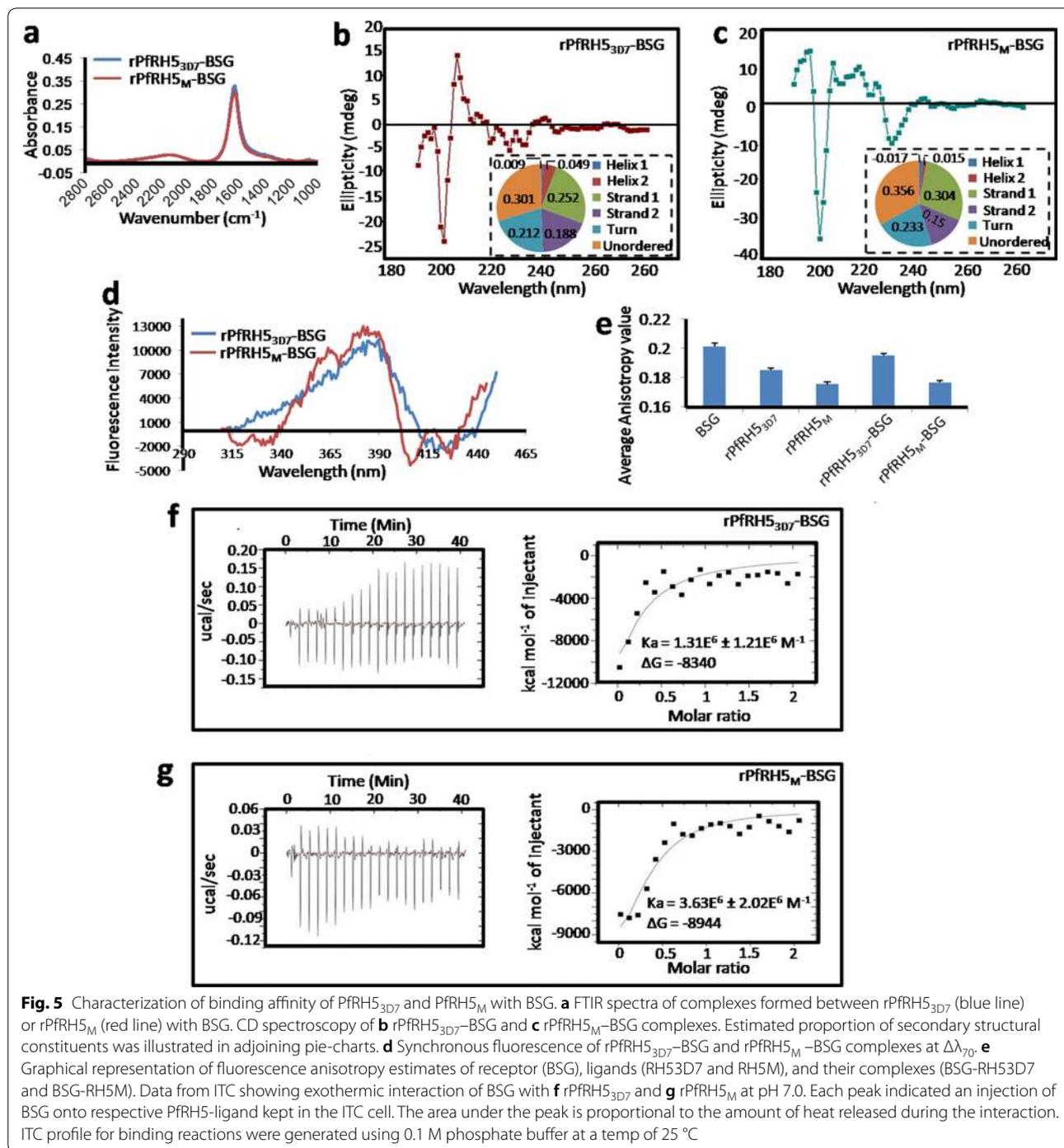
responsible for the observed difference (Fig. 5d). Estimated fluorescence anisotropy indicated a more compact conformation of rPFRH5<sub>M</sub>-BSG complex (0.176) compared to rPFRH5<sub>3D7</sub>-BSG (0.198) (Fig. 5e). CD spectra of rPFRH5<sub>3D7</sub>-BSG and rPFRH5<sub>M</sub>-BSG complexes also showed differences in the percentage of beta-sheet and random coil contents of rPFRH5<sub>3D7</sub>-BSG (50.6% beta-sheets, 25.3% random coil) in comparison to rPFRH5<sub>M</sub>-BSG (43.4% beta-sheet, 33.3% random coil) (Fig. 5b, c).

Finally, isothermal titration calorimetry was employed to compare the thermodynamic aspects of complexes formed between rPFRH5s and BSG. ITC yielded negative heat deflection, indicating the binding between rPFRH5<sub>M</sub> and BSG to be an exothermic event. The association constant ( $K_a$ ) of rPFRH5<sub>M</sub> with was found to be 2.77 fold higher ( $K_a = 3.63E^6 \pm 2.02E^6 \text{ M}^{-1}$ ) than that between

rPFRH5<sub>3D7</sub> and BSG ( $K_a = 1.31E^6 \pm 1.21E^6 \text{ M}^{-1}$ ). This validated a distinct functional impact of C203Y polymorphism in relation to ligand-receptor interaction (Fig. 5f, g). Gibbs free energy of binding ( $\Delta G$ ) was also lower for rPFRH5<sub>M</sub>-BSG (-8944) than rPFRH5<sub>3D7</sub>-BSG (-8340).

### Discussion

More than 50 years of continued research and control initiatives have caused a commendable reduction of malaria incidence and mortality globally in the last 15 years [46]. In keeping with this progress, the World Health Organization (WHO) has compiled a technical strategy for achieving 90% reduction of malaria incidence and mortality with particular emphasis on development of a malaria vaccine with greater than 75% protective



efficacy by 2030 [47]. Till date most malaria vaccines are directed against single stage-specific parasite proteins and are often met with limited success due to the mutational diversity of parasite antigens and complex developmental stages of *P. falciparum*. For example, RTS,S which is close to obtaining licensure, displays a modest vaccine efficacy in terms of blocking infection [48]. In addition,

there always remains a probability that the vaccine efficacy would deteriorate further due to parasite's genetic adaptability. These underscore the need for an efficacious vaccine decorated with epitopes from multiple parasite proteins preferably representing pre-erythrocytic and erythrocytic stages. Inclusion of the geographically predominant version of epitopes is another essential

criterion. Due to extensive polymorphic nature of merozoite proteins such as AMA-1 and MSP-1, focus has currently been shifted to PfrH antigens [27, 49, 50]. Since, India serves as a large and diverse reservoir of *P. falciparum* parasites, the level and pattern of genetic variability of *Pfrh* family of ligands, which take important part in host-parasite interaction during erythrocyte invasion by *P. falciparum*, are thoroughly investigated.

Not all members of PfrH family are essential for parasite propagation, as they can be knocked out without affecting parasite growth and few members of EBA and PfrH families even perform overlapping function [51, 52]. PfrH5-BSG interaction, nevertheless, is crucial for erythrocyte invasion by *P. falciparum* and PfrH5 antibodies have been shown to exhibit strain-transcending invasion inhibition [19, 53]. Thus a detailed functional analysis of one important PfrH5 variant is also performed.

Analysis of receptor binding regions of *Pfrh1*, *Pfrh2a/b*, *Pfrh4* and *Pfrh5* sequences reveals majority of the polymorphic sites (2/4, 7/8, 4/6 and 14/18, respectively) are rare (frequency < 0.05) among the parasite isolates of Chhattisgarh and West Bengal. For each of these loci, the estimate of nucleotide diversity,  $\pi$  is lower than Watterson's  $\theta$ , resulting in a negative Tajima's D, which typically denotes that evolutionary drivers such as recent population expansion and/or purifying selection/selective sweep may reconstruct the genetic landscape of *Pfrh* genes in *P. falciparum* parasites of these study regions. The observed pattern of mismatch distribution rules out possibility of any past demographic processes. Despite the abundance of low frequency variants, a very high  $d_N/d_S$  of 5.7 is observed for *Pfrh5*, which suggests a possibility of genetic hitchhiking effect. Even so, a model for selective sweep due to any advantageous amino acid substitutions cannot be invoked, as the locus harbours significantly high ratio of within-species non-synonymous to synonymous amino acid variations compared to between-species divergence ( $p=0.0436$ ). It is thus surmised that genetic architecture of *Pfrh* family members has been shaped predominantly by purifying selection. To compensate for the limited sample size, additional sequences for each of *Pfrh* genes retrieved from NCBI were analysed. Despite an overall abundance of low frequency polymorphisms, the structures of the parasite population from different geographical regions showed remarkable genetic differentiation possibly due to factors such as local adaptation and/or random genetic drift. Nevertheless, frequency of C203Y was found to be high in Kenya and Mali, as well [43, 44].

Since, 17 (D53Y, I60K, N143I, Y147H, H148D, S197Y, C203Y, N245K, N246T, E259A, R298I, M304R, Y307N,

V371I, I407V, I410N, K429M) out of the 18 *Pfrh5* variants results in amino acid replacements, an attempt has been undertaken to weigh their possible functional values. Since only C203Y is predominantly present among Indian *P. falciparum* strains, the present study evaluates any functional advantages associated with this replacement in relation to the PfrH5's interaction with BSG. Analyses of circular dichroism data of purified PfrH5 alone or in complex with BSG show significant differences in the composition of alpha-helices, beta-sheets, and random coils due to C203Y replacement. Pattern of intrinsic fluorescence and anisotropy measures for rPfrH5<sub>M</sub>-BSG complex indicate that alteration in secondary structure of rPfrH5<sub>M</sub> has further affected the three-dimensional folding of ligand-receptor complex. Estimation of interaction energy and thermodynamic association constant rPfrH5<sub>M</sub>-BSG further establish a favourable biochemical impact of C203Y polymorphism on the affinity between rPfrH5<sub>M</sub> and BSG. This data however contrasts a recent study conducted in Mali which fails to find any correlations between amino acid substitutions on PfrH5 with malaria risk [43]. An in vivo validation of the present study in animal or cellular model or by setting up an appropriate epidemiological study specifically addressing whether the parasites with C203Y variant are endowed with increased pathogenicity may resolve this ambiguity. It is noteworthy in this regard that none of the PfCSP polymorphisms have been associated with any clinical risks in paired consecutive infections, but RTS,S which targets CSP still shows allele-specific vaccine efficacy [43, 54, 55]. Taken together, the study unequivocally encourages 203Y version of PfrH5 to be chosen as a prospective vaccine target because of its limited variability across globe in addition to the allele specific effect of the polymorphism in PfrH5-BSG interaction to minimize the possibility of vaccine escape.

## Conclusions

The data presented here clearly demonstrate that genetic architecture of *Pfrhs* is dominated by presence of rare variants due to possible action of purifying selection. C203Y polymorphism located on PfrH5 is present in overwhelming proportion in the study population. This C to Y replacement strengthens the interaction between PfrH5 and BSG. The implication of this genetic variant in the context of host-parasite conversation during merozoite invasion needs to be explored further. Nonetheless, present data is sufficiently informative while designing a rational multi-valent malaria vaccine with PfrH5 being one of the components.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12936-020-03433-z>.

**Additional file 1: Table S1.** Details of oligonucleotide primers used for amplification and sequencing.

**Additional file 2: Table S2.** Comparison of genetic diversity of Pfrh loci in publicly available global data

**Additional file 3: Fig. S1. a** Neighbor joining network of *Pfrh5* haplotypes (excluding the singletons) observed in India, Kenya and Mali. Size of a circle was proportional to relative frequency of respective haplotype and each branch of the network represented a mutational step. **b** Frequency of the haplotypes in respective regions.

**Additional file 4: Fig. S2.** Synchronous fluorescence data of recombinant target proteins and their complexes with BSG at different  $\Delta\lambda$ s.

### Abbreviations

PfRHs: *P. falciparum* reticulocyte binding protein-like homologs; PCR: Polymerase chain reaction; SNPs: Single nucleotide polymorphisms; CD: Circular dichroism; FTIR: Fourier transform infrared spectroscopy.

### Acknowledgements

Authors are thankful to Souvik Roy and Arijit Pal of DBT-CU-IPLS CORE FACILITY and Anirban Bose of Department of Biochemistry, University of Calcutta for their assistance in instrument handling and data analysis. Authors are indebted to CAS (UGC), Department of Biochemistry, University of Calcutta for providing sequencing facility and DBT-CU-IPLS CORE FACILITY, University of Calcutta for providing Isothermal Titration Calorimeter.

### Authors' contributions

SSG conceived and designed the study. MM and SDK helped in collecting blood samples from malaria patients. SG performed the experiments. SG, PC, SS, SR, ADG and SSG carried out the statistical and bioinformatic analysis. SG and SSG wrote the manuscript. All authors read and approved the final manuscript.

### Funding

This work has been supported by the funding from Department of Biotechnology, Government of West Bengal (Project no: 215-BT(Estt)/RD-08/2015 dated 16.03.2015) and fellowship from Council of Scientific & Industrial Research (CSIR), India (Sanction No.: 09/028(0957)/2015-EMR-1).

### Availability of data and materials

All sequence data are publicly available at NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

### Ethics approval and consent to participate

Ethical clearance to conduct the study was obtained from the Calcutta National Medical College and Hospital and Institutional Ethics committee for human research, Pt. Ravishankar Shukla University, Raipur. Written informed consent, prior to sample collection, was collected from each study participant or guardian, in case where the participant was a child.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interest.

### Author details

<sup>1</sup> Department of Biochemistry, University of Calcutta, 35, Ballygunge Circular Road, Kolkata 700 019, West Bengal, India. <sup>2</sup> School of Studies in Anthropology, Pt. Ravishankar Shukla University, Raipur 492010, Chhattisgarh, India. <sup>3</sup> Department of Pediatric Medicine, Institute of Post Graduate Medical Education & Research, Kolkata, West Bengal, India. <sup>4</sup> Mitra Tower, Lake Town, Block-A, Kolkata 700 089, India.

Received: 9 April 2020 Accepted: 1 October 2020

Published online: 14 October 2020

### References

- Happi CT, Gbotosho GO, Folarin OA, Bolaji OM, Sowunmi A, Kyle DE, et al. Association between mutations in *Plasmodium falciparum* chloroquine resistance transporter and *P. falciparum* multidrug resistance 1 genes and in vivo amodiaquine resistance in *P. falciparum* malaria-infected children in Nigeria. *Am J Trop Med Hyg*. 2006;75:155–61.
- Folarin OA, Bustamante C, Gbotosho GO, Sowunmi A, Zalis MG, Oduola AM, et al. *In vitro* amodiaquine resistance and its association with mutations in *pfcr* and *pfmdr1* genes of *Plasmodium falciparum* isolates from Nigeria. *Acta Trop*. 2011;120:224–30.
- Nosten F, Vugt MV, Price R, Luxemburger C, Thway KL, Brockman A, et al. Effects of artesunate-mefloquine combination on incidence of *Plasmodium falciparum* malaria and mefloquine resistance in western Thailand: a prospective study. *Lancet*. 2000;356:297–302.
- Imwong M, Suwannasin K, Kunasol C, Sutawong K, Mayxay M, Rekol H, et al. The spread of artemisinin-resistant *Plasmodium falciparum* in the Greater Mekong subregion: a molecular epidemiology observational study. *Lancet Infect Dis*. 2017;17:491–7.
- Beeson JG, Drew DR, Boyle MJ, Feng G, Fowkes FJ, Richards JS. Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *FEMS Microbiol Rev*. 2016;40:343–72.
- Marsh K, Kinyanjui S. Immune effector mechanisms in malaria. *Parasite Immunol*. 2006;28:51–60.
- Tham WH, Healer J, Cowman AF. Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends Parasitol*. 2012;28:23–30.
- Wright KE, Hjerrild KA, Bartlett J, Douglas AD, Jin J, Brown RE, et al. Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature*. 2014;515:427–30.
- Aniweh Y, Suurbaar J, Morang'a CM, Nyarko PB, Wright KE, Kusi KA, et al. Analysis of *Plasmodium falciparum* Rh2b deletion polymorphism across different transmission areas. *Sci Rep*. 2020;10:1498.
- Lopez-Perez M, Villasis E, Machado RL, Póvoa MM, Vinetz JM, Blair S, et al. *Plasmodium falciparum* field isolates from South America use an atypical red blood cell invasion pathway associated with invasion ligand polymorphisms. *PLoS ONE*. 2012;7:e47913.
- Lopatnicki S, Maier AG, Thompson J, Wilson DW, Tham WH, Triglia T, et al. Reticulocyte and erythrocyte binding-like proteins function cooperatively in invasion of human erythrocytes by malaria parasites. *Infect Immun*. 2011;79:1107–17.
- Weiss GE, Gilson PR, Taechalerpaisarn T, Tham WH, de Jong NW, Harvey KL, et al. Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during *Plasmodium falciparum* invasion of erythrocytes. *PLoS Pathog*. 2015;11:e1004670.
- Basu M, Das T, Ghosh A, Majumder S, Maji AK, Kanjilal SD, et al. Gene-gene interaction and functional impact of polymorphisms on innate immune genes in controlling *Plasmodium falciparum* blood infection level. *PLoS ONE*. 2012;7:e46441.
- Joshi H, Valecha N, Verma A, Kaul A, Mallick PK, Shalini S, et al. Genetic structure of *Plasmodium falciparum* field isolates in eastern and north-eastern India. *Malar J*. 2007;6:60.
- Chakraborty B, Mondal P, Gajendra P, Mitra M, Das C, Sengupta S. Deciphering genetic regulation of CD14 by SP1 through characterization of peripheral blood mononuclear transcriptome of *P. falciparum* and *P. vivax* infected malaria patients. *EBioMedicine*. 2018;37:442–52.
- Jain V, Basak S, Bhandari S, Bharti PK, Thomas T, Singh MP, et al. Burden of complicated malaria in a densely forested Bastar region of Chhattisgarh State (Central India). *PLoS ONE*. 2014;9:e115266.
- PlasmoDB. The Plasmodium Genomics Resource. <https://plasmodb.org>.
- Sahar T, Reddy KS, Bharadwaj M, Pandey AK, Singh S, Chitnis CE, et al. *Plasmodium falciparum* reticulocyte binding-like homologue protein 2 (PFRH2) is a key adhesive molecule involved in erythrocyte invasion. *PLoS ONE*. 2011;6:e17102.
- Pandey AK, Reddy KS, Sahar T, Gupta S, Singh H, Reddy EJ, et al. Identification of a potent combination of key *Plasmodium falciparum* merozoite antigens that elicit strain-transcending parasite-neutralizing antibodies. *Infect Immun*. 2013;81:441–51.

20. Tham WH, Wilson DW, Reiling L, Chen L, Beeson JG, Cowman AF. Antibodies to reticulocyte binding protein-like homologue 4 inhibit invasion of *Plasmodium falciparum* into human erythrocytes. *Infect Immun*. 2009;77:2427–35.
21. BLAST: Basic Local Alignment Search Tool – NIH. <https://blast.ncbi.nlm.nih.gov/>.
22. Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*. 1999;15:174–5.
23. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003;19:2496–7.
24. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
25. Tajima F. The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics*. 1996;143:1457–65.
26. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133:693–709.
27. Basu M, Maji AK, Mitra M, Sengupta S. Natural selection and population genetic structure of domain-I of *Plasmodium falciparum* apical membrane antigen-1 in India. *Infect Genet Evol*. 2013;18:247–56.
28. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3:418–26.
29. Jukes T, Cantor C. Evolution of protein molecules. In: *Mammalian Protein Metabolism*, Vol. III, Munro HN, Ed. New York: Academic Press; 1969.
30. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16:37–48.
31. Bull PC, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, et al. *Plasmodium falciparum* antigenic variation. Mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol Microbiol*. 2008. 68:1519–34.
32. Barry AE, Schultz L, Buckee CO, Reeder JC. Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite *Plasmodium falciparum*. *PLoS ONE*. 2009;4:e8497.
33. Friedman SR, Aral S. Social networks, risk-potential networks, health, and disease. *J Urban Health*. 2001;78:411–8.
34. RCSB PDB: Homepage. <https://www.rcsb.org>.
35. AutoDock. [autodock.scripps.edu](http://autodock.scripps.edu).
36. Rodriguez M, Lustigman S, Montero E, Oksov Y, Lobo CA. PfrH5: a novel reticulocyte-binding family homolog of *Plasmodium falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE*. 2008;3:e3300.
37. Lloyd JBF. Synchronized excitation of fluorescence emission spectra. *Nature*. 1971;231:64–5.
38. Li R, Goswami U, King M, Chen J, Cesario TC, Rentzepis PM. In situ detection of live-to-dead bacteria ratio after inactivation by means of synchronous fluorescence and PCA. *Proc Natl Acad Sci USA*. 2018;115:668–73.
39. Chin J, Längst G, Becker PB, Widom J. Fluorescence anisotropy assays for analysis of ISWI-DNA and ISWI-nucleosome interactions. *Methods Enzymol*. 2004;376:3–16.
40. Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, et al. Insights into protein–ligand interactions: mechanisms, models, and methods. *Int J Mol Sci*. 2016;17:144.
41. Reiling L, Richards JS, Fowkes FJL, Barry AE, Triglia T, Chokejindachai W, et al. Evidence that the erythrocyte invasion ligand PfrH2 is a target of protective immunity against *Plasmodium falciparum* malaria. *J Immunol*. 2010;185:6157–67.
42. Reiling L, Richards JS, Fowkes FJL, Wilson DW, Chokejindachai W, Barry AE, et al. The *Plasmodium falciparum* erythrocyte invasion ligand Pfrh4 as a target of functional and protective human antibodies against malaria. *PLoS ONE*. 2012;7:e45253.
43. Ouattara A, Tran TM, Doumbo S, Adams M, Agrawal S, Niangaly A, et al. Extent and dynamics of polymorphism in the malaria vaccine candidate *Plasmodium falciparum* reticulocyte-binding protein homologue-5 in Kalifabougou Mali. *Am J Trop Med Hyg*. 2018;99:43–50.
44. Ochola-Oyier LI, Wamae K, Omedo I, Ogola C, Matharu A, Musabyimana JP, et al. Few *Plasmodium falciparum* merozoite ligand and erythrocyte receptor pairs show evidence of balancing selection. *Infect Genet Evol*. 2019;69:235–45.
45. DichroWeb - Online Circular Dichroism Analysis. [dichroweb.cryst.bbk.ac.uk](http://dichroweb.cryst.bbk.ac.uk).
46. WHO. World Malaria Report 2014. Geneva, World Health Organization. [http://www.who.int/malaria/publications/world\\_malaria\\_report\\_2014/en](http://www.who.int/malaria/publications/world_malaria_report_2014/en).
47. WHO. Global Technical Strategy for Malaria 2016–2030. Geneva, World Health Organization. <http://www.who.int/malaria/publications/atoz/9789241564991/en>.
48. Penny MA, Verity R, Bever CA, Sauboin C, Galactionova K, Flasche S, et al. Public health impact and cost-effectiveness of the RTS, S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models. *Lancet*. 2016;387:367–75.
49. Ghoshal S, Gajendra P, Datta Kanjilal S, Mitra M, Sengupta S. Diversity analysis of MSP1 identifies conserved epitope organization in block 2 amidst high sequence variability in Indian *Plasmodium falciparum* isolates. *Malar J*. 2018;17:447.
50. Chowdhury P, Sen S, Kanjilal SD, Sengupta S. Genetic structure of two erythrocyte binding antigens of *Plasmodium falciparum* reveals a contrasting pattern of selection. *Infect Genet Evol*. 2018;57:64–74.
51. Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. *Cell*. 2006;124:755–66.
52. Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, Duraisingh MT, et al. Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science*. 2005. 309:1384–7.
53. Reddy KS, Pandey AK, Singh H, Sahar T, Emmanuel A, Chitnis CE, et al. Bacterially expressed full-length recombinant *Plasmodium falciparum* RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infect Immun*. 2014;82:152–64.
54. Gandhi K, Thera MA, Coulibaly D, Traoré K, Guindo AB, Ouattara A, et al. Variation in the circumsporozoite protein of *Plasmodium falciparum*: vaccine development implications. *PLoS ONE*. 2016;11:e0148240.
55. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and protective efficacy of the RTS, S/AS01 malaria vaccine. *N Engl J Med*. 2015;373:2025–37.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

