Limited genetic variations of the Rh5-CyRPA-Ripr invasion complex in Plasmodium falciparum parasite population in selected malaria-endemic regions, Kenya

Waweru, Harrison; Kanoi, Bernard N.; Kuja, Josiah O.; Maranga, Mary; Kongere, James; Maina, Michael; Kinyua, Johnson; Gitaka, Jesse

Published in:
Frontiers in Tropical Diseases

DOI:
10.3389/fitd.2023.1102265

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Limited genetic variations of the Rh5-CyRPA-Ripr invasion complex in *Plasmodium falciparum* parasite population in selected malaria-endemic regions, Kenya

Harrison Waweru1,2, Bernard N. Kanoi1,3, Josiah O. Kuja1,4, Mary Maranga5, James Kongere3, Michael Maina1,2, Johnson Kinyua3 and Jesse Gitaka1,3*

1Centre for Research in Infectious Diseases, Directorate of Research and Innovation, Mount Kenya University, Thika, Kenya, 2Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, 3Centre for Research in Tropical Medicine and Community Development, Nairobi, Kenya, 4Department of Biology, University of Copenhagen, Copenhagen, Denmark, 5Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland

The invasion of human erythrocytes by *Plasmodium falciparum* merozoites requires interaction between parasite ligands and host receptors. Interaction of PfRh5-CyRPA-Ripr protein complex with basigin, an erythrocyte surface receptor, via PfRh5 is essential for erythrocyte invasion. Antibodies raised against each antigen component of the complex have demonstrated erythrocyte invasion inhibition, making these proteins potential blood-stage vaccine candidates. Genetic polymorphisms present a significant challenge in developing efficacious vaccines, leading to variant-specific immune responses. This study investigated the genetic variations of the PfRh5 complex proteins in *P. falciparum* isolates from Lake Victoria islands, Western Kenya. Here, twenty-nine microscopically confirmed *P. falciparum* field samples collected from islands in Lake Victoria between July 2014 and July 2016 were genotyped by whole genome sequencing, and results compared to sequences mined from the GenBank database, from a study conducted in Kilifi, as well as other sequences from the MalariaGEN repository. We analyzed the frequency of polymorphisms in the PfRh5 protein complex proteins, PfRh5, PCyRPA, PfRipr, and PfP113, and their location mapped on the 3D protein complex structure. We identified a total of 58 variants in the PfRh5 protein complex. PfRh5 protein was the most polymorphic with 30 SNPs, while PCyRPA was relatively conserved with 3 SNPs. The minor allele frequency of the SNPs ranged between 1.9% and 21.2%. Ten high-frequency alleles (>5%) were observed in PfRh5 at codons 147, 148, 277, 410, and 429 and in PfRipr at codons 190, 255, 259, and 1003. A SNP was located in protein-protein interaction region C203Y and F292V of PfRh5 and PfCyRPA, respectively. Put together, this study revealed low polymorphisms in the PfRh5
Background

The World Health Organization (WHO) estimates the latest global malaria health burden statistic at 627,000 deaths resulting from 241 million malaria infection cases (1). The scourge’s heaviest health and economic burden is borne by the developing countries in Sub-Saharan Africa, where an estimated 90% of all malaria deaths occur, with children under five accounting for 78% of all deaths (2). The emergence of multi-drug resistance Plasmodium falciparum (P. falciparum) resistant strains and insecticide resistance mosquitoes remains a significant challenge in treating and eliminating malaria (3, 4). The lack of an effective vaccine remains one of the most critical gaps in the strategies developed to eliminate P. falciparum malaria (5).

The development of an effective P. falciparum vaccine focuses on targeting pre-erythrocytic or erythrocytic stages for parasite development and malaria pathogenesis in humans or the parasite development within the mosquito vector (6). Symptoms elicited by parasite infection originate from the erythrocytic stage of malaria infection. At this stage, the merozoite invades the erythrocytes, where at initial recognition of the human erythrocytes, the merozoite orients itself such that the apical region comes to direct contact with the host’s erythrocyte membrane, followed by irreversible attachment of merozoites to erythrocytes (7). Ring-like moving junction mediates complete parasite internalization to formation of an intracellular parasitophorous vacuole (8). The whole erythrocyte invasion process is mediated by multiple merozoite proteins mainly expressed on the surface, or in the apical organelles such as rhoptry and microneme (9). Since these proteins are essential for invasion and are exposed to host immune system, they are considered ideal targets for blood stage vaccines (BSV) (10–12). However, the exposure of candidate BSV antigens to their impacts on parasite invasion process to support the consideration of PRh5 components as potential malaria vaccine candidates.

Keywords

genetic variations, Rh5-CyRPA-Ripr invasion complex, malaria, vaccines, erythrocyte (RBC)

Methods

Sampling, DNA preparation, and whole genome sequencing

Parasite DNA was extracted from archived whole blood samples from patients recruited for a drug resistance surveillance study in the Lake Victoria parasite population. However, the two mutations identified on the protein interaction regions prompts for investigation on their impacts on parasite invasion process to support the consideration of PRh5 components as potential malaria vaccine candidates.
local hospitals on four selected islands (Mfangano, Takawiri, Kibuogi, and Ngodhe) in Lake Victoria, a coastal mainland (Ungoye) between July 2014 and July 2016. The study’s approval was obtained from the Kenyatta National Hospital - University of Nairobi (KNH-UoN) ethical review committee (P609/10/2014) and the Mount Kenya University Ethics Review Committee (038/2014). Written consent was obtained from all the participants or guardians, and malaria cases were treated per the national malaria guidelines. The samples re-analyzed here were a subset of these studies which has been extensively described elsewhere (26, 27). Briefly, to increase the parasitemia, the field *P. falciparum* parasites were adapted for *in vitro* culture as previously described (28), and DNA was extracted from short-term cultures (1 month) at the schizont stage using QIAamp DNA mini kit (Qiagen, Valencia, CA). Paired-end sequencing libraries were prepared using Nextera XT DNA library preparation Kit according to the manufactures protocol. (Illumina, USA). Whole genome sequencing was performed on Illumina MiSeq technology (Illumina, USA) at 30X coverage generating reads of length 150 bps. These sequences are archived at the DDBJ BioProject, Accession number PRJDB12148. Quality control checks were performed using the FASTQC (Babraham Institute, UK) toolkit version 0.11.5.

**Comparison of polymorphisms identified with other regions**

For comparative analysis, we obtained previously reported whole genome sequences *P. falciparum* isolates collected from Kilifi, a malaria endemic region in coastal Kenya (29, 30). The mined sequences were generated from two drug trial studies that were conducted between 2005 and 2008, and the sequences deposited in the GenBank repository under accession numbers PfRipr: MW597717-MW597776, PfRh5: MW597550-MW597609, PfCyRPA: MW597610-MW597716, and PfP113: MW597459-MW597549. We also accessed the catalogue of genetic variation in *P. falciparum*, of the global MalariaGEN database v6.0, for comparing and validating the SNPs identified from the Lake Victoria sample population. This dataset comprised of genomic variation records of 7,113P. falciparum samples from 28 malaria-endemic countries. The method used to retrieve the data was previously described by Amato et al., 2016. The dplyr v1.0.9 package (Wickham H, François R, Henry L, 2022) in R v4.2.1 was used to filter out the four genes of the PfRh5 complex using the PlasmoDB unique identifiers. SNPs identified were then filtered and analyzed.

**Read mapping and coverage**

Sequence reads were aligned against *Plasmodium falciparum* 3D7 reference genome (version 8.1) (https://plasmodb.org/common/downloads/release-46/Pfalciparum3D7/fasta/data/) using Burrows-Wheeler Alignment tool (BWA) (31) (http://bio-bwa.sourceforge.net) with default parameters. The resulting alignment was further processed with Samtools (32) and Picard v1.66 (33) to remove duplicates. SNPs were called using Genome Analysis Toolkit (GATK) HaplotypeCaller with the following parameters – genotyping mode DISCOVERY, – output mode EMIT_VARIANTS_ONLY, –stand_emit_conf 10, and –stand_call_conf 30. To improve the quality of variant calling, we further discarded genotyping calls with coverage of <5 reads. The resulting variant call format (VCF) files were then merged into one file using VCF tools (34).

**Variant calling and analysis**

The VCF file containing twenty-six samples that passed the quality test from read mapping analysis was used as the input file in VCF tools for variant analysis. High-quality SNPs in four target genes, PfRh5, PfCyRPA, PfRipr, and PfP113, were functionally annotated in the SNPEFF tool (35). Called variants were further analyzed in ARTEMIS software (36). MEGA 7 tool was used to perform multiple sequence alignment and translation of nucleotide sequences to amino acid sequences. Gene variants were identified by aligning the amino acid sequence reads to their corresponding 3D7 reference gene sequence. To test sensitivity of the above approach, we analyzed the sequences at different variant calling parameters to assess the impact of these settings on downstream analysis. Additionally, prior to variant calling, we performed base quality score recalibration to adjust the base quality scores of sequencing reads as well as local realignment around indels to reduce false-positive variant calls resulting from alignment artifacts.

**Population genetics analysis**

The population genetic tests for the neutral theory of evolution (37) and Tajima’s D and Fu & Li’s statistics and nucleotide diversity (Pi) were calculated using DnaSPv6.1 (38). Tajima’s D tested departure from neutrality based on allele frequency distribution in each gene. Fu and Li’s D test statistic calculated the variation between the observed number of singletons and the total number of mutations. Pi was used to test the genetic diversity of each gene of the PfRh5 complex within the parasite population from Lake Victoria region. The *P. falciparum* adenosylsuccinate lyase gene, a house keeping gene and apical membrane antigen gene were used as control in this analysis. The sequences for these genes were obtained from Lake Victoria parasite population.

**Protein structures**

The structure of the Rh5-CyRPA-Ripr complex was retrieved from the Protein Data Bank (http://www.rcsb.org/) under the protein ID 6MPV. The datasets generated from this study were used to map the polymorphic sites of PfRh5 and PfCyRPA protein structures on the Rh5-CyRPA-Ripr complex 3D structure in PyMol (The PyMOL Molecular Graphics System, Version 2.2.0, Schrödinger, LLC) to determine the location of the SNPs in the 3D protein structure and whether the SNPs were localized in the protein-protein interaction regions of the complex.
Results

Genetic variation in the PfRh5 complex genes

Whole genome sequence analysis data for the four genes of the PfRh5 complex were obtained from 26 samples from the Lake Victoria islands. A total of 45, 35, 25, and 3 Non-synonymous SNPs were identified within the PfP113, PfRipr, PfRh5, and the PfCyRPA genes, respectively. The minor allele frequency in the four genes ranged from 0.7% to 24.06%. High-frequency alleles (>5%) were identified in codons Y147, H148, C203, S277, I410, and K429 of the PfRh5 gene and codons V190, M255, Y259, and A1003 of the PfRipr gene (Figure 1). Non-synonymous SNPs identified in PfCyRPA gene at codons S25, D236, and V292 and PfP113 gene at codons L17, E234, Q620, and Q857 occurred at low frequency (Figure 1).

Comparison of polymorphisms identified with the Kilifi population

A total of nine non-synonymous previously not observed from the Kilifi population were identified across the four genes in Lake Victoria isolates (39). PfRipr gene at codon Y226, F236, and T441, PfRh5 at codon S277, PfP113 at codon L17, Q620, and Q857, and PfCyRPA gene at codon S25 and V292 (Figure 1).

Comparison of polymorphisms identified with global MalariaGEN

We further explored the MalariaGEN data to establish whether the variants identified were also present in the global database. We also screened for variants missed due to differences in analysis methodologies. This analysis confirmed that most of the variants observed from the Lake Victoria population had been previously observed elsewhere and deposited in the global MalariaGEN database giving confidence in our analysis methodology on the probability of missing variants (Table 1).

Population genetics statistics

A sliding window approach was used to calculate the nucleotide diversity (π) and Tajima’s D statistics. All four genes had a negative neutrality summary statistic. Analysis revealed that the PfP113 gene was the most conserved relative to the other three genes of the PfRh5 complex and the positive control Pfama1 gene, with a Tajima’s D summary statistic of –1.89 and a π of 0.00010 with 4 singleton mutations distributed along the 1578 bp nucleotide sequence. Relative to the negative control Pfadsl gene, the PfRh5 gene was the least conserved, with 25 non-synonymous polymorphisms distributed along the entire 2907 bp nucleotide sequence with a Tajima’s D summary value of -0.56 and a π of 0.00109 (Figures 2, 3). The Fu & Li’s statistics were not significant for PfRh5 and PfRipr genes. PfCyRPA and PfP113 gene had significant Fu & Li’s values, p < 0.05. (Table 2)

Polymorphisms on the PfRh5 protein complex

The polymorphisms established from our dataset were mapped on the PfRh5 protein complex to show whether they occurred within known protein-protein interacting regions and the PfRh5 – basigin interaction region. The previously published Rh5-CyRPA-Ripr invasion complex (16) was superimposed with the basigin structure to show the interaction of PfRh5 with basigin. PfRh5 binds to basigin via His-102 linker, α-2, α-4, and a disulfide loop.

![FIGURE 1](https://frontiersin.org)  
**FIGURE 1**

Prevalence of PfRh5 complex polymorphisms. Frequency in percentage of single nucleotide polymorphisms identified across four genes of the Rh5 invasion complex (A) PfRipr, (B) PfRh5, (C) PfCyRPA, (D) PfP113 in Lake Victoria (n=26) and Kilifi (n=68) populations.
(Cys345–Cys351) (40). The F350 and W447 PfRh5 residues stabilize binding by packing into basigin hydrophobic bonds. Only one SNP corresponding to codon 203 within the PfRh5-basigin interacting region was identified. PfRipr binds to PCyRPA blade 6 at amino acid residues 281 – 311. One mutation corresponding to this interaction region at codon 292 of PCyRPA was identified. No mutation corresponding to PCyRPA and PfRh5 binding regions was identified. Other polymorphisms were localized outside the protein interaction site (Figure 4).

Discussion

*P. falciparum* infects and replicates in human host erythrocytes leading to manifestation of clinical of malaria. The invasion process by invasive merozoites involves the interaction of PfRh5 protein and the basigin receptor localized on the erythrocyte membrane (41, 42). However, PfRh5 does not function alone. Upon secretion, it forms a heteromeric complex with two micronemal proteins, PfRipr and PfCyRPA. PfRipr and PfCyRPA proteins do not interact with basigin and have been shown to lack a membrane anchor (11). The rationale for developing a blood-stage malaria vaccine targeting the components of the PfRh5 complex has been supported by *in vitro* and *in vivo* studies in non-human primates. Antibodies raised against the PfRh5 proteins have been shown to block erythrocyte invasion by inhibiting its binding to basigin receptor (5, 40, 43–45). Genes coding for proteins of the PfRh5 complex are highly conserved in *P. falciparum*, suggesting their vital role in parasite survival (30, 46). Therefore, an PfRh5-complex-based vaccine would prove effective. In the present study, we identified the genetic variations of the proteins that make up the *P. falciparum* Rh 5 complex and determined the polymorphism’s locus on the protein complex in the parasite population from the Mfangano, Takawiri, Kibuogi, and Ngodhe Islands of Lake Victoria in Western Kenya and compared with Kilifi and global databases. All genes of the PfRh5 complex were relatively conserved, and the negative population genetics statistic suggests the parasite population has limited potential to retain these mutations.

The observed negative Tajima’s D statistics from the Lake Victoria population indicated an excess of rare variants and do

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Ref</th>
<th>Alt</th>
<th>S/NS</th>
<th>Lake Victoria (n=26)</th>
<th>Kilifi (n=68)</th>
<th>MalariaGEN global dataset (n=7,113)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pfripr</td>
<td>pfripr_190</td>
<td>V</td>
<td>A</td>
<td>NS</td>
<td>7.7</td>
<td>10.3</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfripr_226*</td>
<td>Y</td>
<td>H</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfripr_236*</td>
<td>F</td>
<td>V</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfripr_255</td>
<td>M</td>
<td>I</td>
<td>NS</td>
<td>7.7</td>
<td>2.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfripr_259</td>
<td>Y</td>
<td>H</td>
<td>NS</td>
<td>23.1</td>
<td>23.6</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfripr_441*</td>
<td>T</td>
<td>N</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfripr_1003</td>
<td>A</td>
<td>S</td>
<td>NS</td>
<td>7.7</td>
<td>10.3</td>
<td>+</td>
</tr>
<tr>
<td>pfrh5</td>
<td>pfrh_147</td>
<td>Y</td>
<td>H</td>
<td>NS</td>
<td>7.7</td>
<td>22.6</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfrh_148</td>
<td>H</td>
<td>D</td>
<td>NS</td>
<td>7.7</td>
<td>22.6</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfrh_203</td>
<td>C</td>
<td>Y</td>
<td>NS</td>
<td>42.3</td>
<td>10.3</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfrh_277</td>
<td>S</td>
<td>N</td>
<td>NS</td>
<td>7.7</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfrh_371*</td>
<td>V</td>
<td>I</td>
<td>NS</td>
<td>3.8</td>
<td>4.1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfrh_407*</td>
<td>I</td>
<td>V</td>
<td>NS</td>
<td>3.8</td>
<td>0.7</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfrh_410</td>
<td>I</td>
<td>M</td>
<td>NS</td>
<td>7.7</td>
<td>3.4</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pfrh_429</td>
<td>K</td>
<td>N</td>
<td>NS</td>
<td>34.6</td>
<td>34.9</td>
<td>+</td>
</tr>
<tr>
<td>pfcyrpa</td>
<td>pfcyrpa_25*</td>
<td>S</td>
<td>N</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfcyrpa_236*</td>
<td>D</td>
<td>N</td>
<td>NS</td>
<td>3.8</td>
<td>3.4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfcyrpa_292*</td>
<td>V</td>
<td>F</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>pf113</td>
<td>pf113_17*</td>
<td>L</td>
<td>V</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pf113_234*</td>
<td>E</td>
<td>K</td>
<td>NS</td>
<td>3.8</td>
<td>1.4</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>pf113_620*</td>
<td>Q</td>
<td>H</td>
<td>S</td>
<td>3.8</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pf113_857*</td>
<td>Q</td>
<td>E</td>
<td>NS</td>
<td>3.8</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

Polymorphisms identified in both populations and the global MalariaGEN dataset are indicated with a plus +, and singleton SNPs are indicated by an asterisk*. Ref refers to the 3D7 *Plasmodium falciparum* strain reference amino acid, while Alt refers to the amino acid variation. The SNPs are classified under S (synonymous or NS (non-synonymous) and frequency of SNPs is expressed as percentages.
not suggest balancing selection (30). Genes with a significant negative Tajima’s D value indicate that the parasites population has a limited potential to retain polymorphisms, especially PfP113 and PfRipr genes (47). These findings are consistent with previous studies of P. falciparum in the African population, which showed a majority of genes having a negative Tajima’s D value, suggesting a historical parasite population expansion event (48, 49).
In contrast to other merozoite antigens that are considered potential vaccine candidates such as Apical membrane antigen 1 (AMA1), merozoite surface protein 1 (MSP1), and merozoite surface protein 10 (MSP10) (50, 51), majority of polymorphisms of the PfRh5 complex components were rare variants and did not indicate balancing selection. Recent studies from Nigeria and Kenya reported one non-synonymous mutation on PfRh5 protein at codon C203Y (30, 52). We identified the C203Y mutation in the Lake Victoria population while mutations at codon Y147H, H148D, and K429N were reported in Kilifi samples and MalariaGEN global database as rare variants, which suggests a need for P. falciparum to maintain these mutations across various populations. Mutation at codon S277N observed from Lake Victoria isolates was not reported from the Kilifi population. Three singleton mutations at codons Y226H, F236V, and T441N of the PfRipr gene were identified in Lake Victoria. Among the three polymorphisms, only the mutation at codon S277N was not reported from the Kilifi population. The components of the PfRh5 complex are located in different subcellular locations; thus, the complex only forms during erythrocyte invasion when they are secreted from the rhoptries or micronemes (42). Field studies have demonstrated that recombinant PfRh5 with the C203Y mutation binds to basigin with the same affinity as the wild type (54).

Put together, developing an effective malaria vaccine remains a priority among strategies to eliminate and eradicate the disease. One major hindrance to achieving this is the emergence of polymorphisms within the various vaccine antigen targets leading to allele-specific immune responses. Among the PfRh5 complex,
PfRh5 is the most advanced vaccine target. However, the presence of low-frequency mutations raises concerns about immune system evasion. This study recommends functional assay studies to investigate the immunological and biological relevance of the identified mutations.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ddbj.nig.ac.jp/, PRJDB12148.

Ethics statement

The study’s approval was obtained from the Kenya National Hospital - University of Nairobi (KNH-UoN) ethical review committee (P609/10/2014) and the Mount Kenya University Ethics Review Committee (038/2014). Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

Author contributions

HW, BK, and JG conceived and designed the study; JG provided the analysed data sets; HW performed the bioinformatic and statistical analysis, under supervision of JOK and BK; HW and BK wrote the first draft of the manuscript, and the final version included edits from all authors. The final manuscript was read and approved by all authors.

References


Funding

This work was supported by the Royal Society of Tropical Medicine and Hygiene (RSTMH) small grants 2019 (HW). BK is an EDCTP Fellow under EDCTP2 programme supported by the European Union grant number TMA2020CDF-3203. JG received support from the African Academy of Sciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We wish to acknowledge the Mount Kenya University research team for their insights in the improvement of this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.


