PfEMP1 DBLα Sequence Tags in Genomic DNA of P. falciparum Field Isolates from Two Malaria Endemic Sites in Kenya

Francis W. Makokha1,2,4*, Sabah A. Omar4, Francis T. Kimani4, Gabriel Magoma4, Rahma Udu4,6, Edwin Too4, Nathan Shaviya5, Charity Hungu4

1. Directorate of Research and Development, Mount Kenya, P.O Box 342-01000, Thika, Kenya
2. Institute of Tropical Medicine and Infectious Diseases, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi Kenya
3. Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi Kenya
4. Centre for Biotechnology and Research Development, Kenya Medical Research Institute, P.O Box 54840-00200, Nairobi, Kenya
5. Masinde Muliro University of Science and Technology, Department of Biomedical Sciences, P.O Box 190-50100, Kakamega, Kenya
6. Department of Biochemistry, Pwani University, P.O Box 185-80108, Kilifi, Kenya

Email: makokhafw@gmail.com

Abstract

Background
Malaria caused by Plasmodium falciparum remains a major cause of childhood morbidity and mortality in sub-Saharan Africa. PfEMP1 protein, coded for by a family of about sixty variant var genes, is a parasite protein found on infected erythrocyte membrane. PfEMP1 protein mediates cytoadherence of infected erythrocytes on endothelial cells which may lead to severe symptoms of malaria. Although PCR amplification of the whole gene is difficult due to high variability, primers targeting the DBLα domain have been designed and used to study pfemp1 genes. This objective of this study was to establish the distribution of DBLα sequence tags in isolates of Plasmodium falciparum from two malaria endemic sites in Kenya.

Methods
DNA extracted from field isolates collected from Mbita (Western Kenya) and Tiwi (Coastal region) was used to isolate and amplify DBLα domain sequence tags of pfemp1 by PCR. PCR products were sequenced by 454 next generation sequencing. After assembly, the translated protein sequences (GenBank KP085750-KP087726) were then aligned in Mega 5.2 and classified into cys/PoLV groups based on the number of cysteine residues and the motifs at PoLV1 and PoLV2 within the sequence tag. Six sequence groups were found in sequences from both endemic sites. Group 4 sequences were the most prevalent (57.35% and 57.07% in isolates from Mbita and Tiwi respectively) in the isolates from both sites. Sequence tags from Tiwi had a higher proportion of cys2 (group 1 and 2) than sequences from Mbita although individual group 2 sequence tags were slightly higher in Mbita tags. Similarly the proportion of groups 5 and 6 sequence tags was higher in sequence tags from Tiwi than those from Mbita.

Conclusion
In conclusion, the frequency of the different cyc/PoLV groups of DBLα sequence tags at both endemic sites follow almost similar pattern with group four sequence tags being the majority among the sequence tags isolated from patient isolates from both study sites. However, in the absence of expression data, the impact of this genomic distribution pattern on malaria pathology remains unknown.

Key Words: Malaria, PfEMP1, cys/PoLV, DBLα, var, Sequence tags

Background
Plasmodium falciparum expresses parasite proteins that are exported to the infected erythrocyte surface. These proteins known as P. falciparum erythrocyte membrane protein 1 (PfEMP1) (1–3) are encoded by a family of genes known as var. Each haploid genome contains approximately 60 var genes, with only one gene being expressed at any given time (4). The family of the 60 var genes is classified into three subgroups based on semi-conserved upstream sequences and the direction of transcription. Group A and B var genes are associated with telomeric regions of chromosomes and are transcribed towards the telomere. The third group, group C is associated with internal var clusters (5–9). Studies have demonstrated that the var genes have both functional and clinical significance. Expressing of group A var genes is associated with severe malaria symptoms. Group A var genes also do not bind host receptors CD36 and intracellular adhesion molecule1 (ICAM1) (7). Further these genes tend to be abundantly expressed in patients whose immunity to malaria is naïve, especially young infants. Group A var genes have been predominantly associated to cerebral malaria and have been found in patient
autopsies suffering this symptom. Using microarray technology, Claessens et al. (7) demonstrated that group A-like var genes that encode PfEMP1 protein were up-regulated in selected parasite lines that bind to human brain microvasculature endothelial cells (HBEC-5i). Although the receptor for group A var gene PfEMP1 ligand in brain cells is remained unknown, a recent study by Turner et al. (10) identified the endothelial protein C receptor (EPCR), as the endothelial receptor for DC8 and DC13 PfEMP1. The same study also demonstrated that EPCR binding was mediated by the amino-terminal cysteine-rich interdomain region (CIDRα1) of DC8 and group A PfEMP1 subfamilies (10). Group B and C var genes code for PfEMP1 ligands that bind CD36 and ICAM1 and their expression is associated to mild symptoms of malaria.

Although var genes are highly diverse, the Duffy binding like alpha (DBLα) domains are relatively conserved. This has enabled designing of primers that target these regions and amplify short sequence tags that are used to study var genes. There are various ways of classifying DBLα sequence tags. One way of classifying DBLα sequence tags is by categorizing them based on the number of cysteine residues present in each sequence tag. This way, DBLα sequence tags can be classified into those containing two cysteine residues (cys2), four cysteine residues (cys4), none, one, three, five or six cysteine residues (cysX). The majority of DBLα sequences fall into cys4 category followed by cys2. Fewer sequence tags fall in the other cysteine classes. Some studies have shown that most cys2 sequences belong to the DBLα domain, DBLα1, found in group A var genes (11) while cys4 sequence tags were largely found in non-group A var genes. Further it has been established that cys2 DBLα sequence tags can have MFK or REY motifs at PoLV1 and PoLV2 respectively. These two motifs are mutually exclusive, implying that a single DBLα tag does not bear both motifs. Cys2 sequences are mostly expressed but not exclusively in ups A var genes. However, the MFK motif has been associated with expression of ups A var genes, hence its presence has been used to define presence of ups A var gene.

A second classification system of DBLα sequence tags is based on the fact that the tags contain regions that show limited variability at specific positions known as positions of limited variability (PoLVs) (11). Each sequence tag typically contains four PoLVs, PoLV1, PoLV2, PoLV3 and PoLV4. The positions are identified based on the anchoring motifs of amino acids, DIGDI and PQFLR at the 5’ and 3’ positions of the sequence tags respectively. This system classifies DBLα sequence tags into six groups known as cys/PoLV groups, based on the number of cysteine residues and the motifs at PoLV1 and PoLV2 (11). Group1 sequence tags consist of cys2 sequences with MFK motif at PoLV1; group 2 sequences are cys2 sequences with REY motif at PoLV2; group3 sequences are cys2 DBLα sequences without MFK and REY motifs at PoLV1 and PoLV2 respectfully. Sequence tags with four cysteine residues (cys4 sequences) that do not have REY motif at PoLV2 are group4 and Cys4 sequences with REY motif at PoLV2 belong to group5. Sequences with 0, 1, 3, 5 of 6 cysteine residues are classified as group 6.

In this study all sequence tags were analysed to determine the number of cysteine residues present in each sequence and the motif at PoLV1 and PoLV2. Sequence tags lacking 5’DIGDI or 3’ PQFLR motif were also included as long as they possessed cys/PoLV features of the particular group they were classified into but lacked a 3’ PQFLR.

METHODOLOGY

Study Sites and Sample Collections

Samples were collected from patients visiting Mbita Sub-district hospital and Tiwi health centre. Both are malaria endemic regions although Western Kenya records more malaria cases than coastal region. Mbita is located in Homa Bay County, Western Kenya on the shores of Lake Victoria. The region experiences ample rainfall and is relatively warm at a low altitude. Transmission of malaria in this area remains relatively high throughout the year, with the highest transmission rates occurring between April and August during the long rains season.

Tiwi is located in Kwale County, along the Indian Ocean coast. Tiwi is a malaria hyper-endemic region and is a moderately populated urban area where about 27% of the population consists of children below age five. It experiences two rainy seasons, long rains season runs from March to June and Short rain season from October to December. Malaria transmission occurs all year round but peaks between May and July, following the onset of long rains. Malaria is a major cause of morbidity, accounting for 30% of pediatric and adult outpatient attendances.

As an inclusion criterion, samples were only collected from consenting patients who tested positive for malaria after microscopic examination during an in vivo Duocotexin and Coartem drug efficacy study. Ethical approval was sought from the Scientific and Ethical Review Committees of Kenya Medical Research Institute (SSC No 1556A). The study only recruited patients with mild/non-severe malaria symptoms who had P. falciparum mono-infection as confirmed by microscopy. Patients with severe malarial symptoms were referred to admission and were not included in the study. A finger prick was used to collect blood samples as spots on Whatman filter paper. The filters were stored at room temperature at the study site and later transported to Malaria laboratory of Kenya Medical Research Institute (KEMRI), Nairobi.
DNA extraction
DNA from twenty seven (27) samples was extracted using chelex method. Briefly, a piece of filter paper (approximately 2mm x 2mm) with the blood spot was incubated in 1000µl of 0.5% saponin in 1x PBS overnight at 4°C. The resulting brown solution was then discarded and replaced with fresh 1xPBS, followed by an incubation of 15 to 30 minutes. The solution was removed and 100µl of DNAse free water was added, followed by 50µl of 20% chelex. The tubes were then incubated on a heated block at 100°C for ten minutes, being vortexed every two minutes. The solution was then centrifuged at 3000 rpm for 3minutes. The supernatant was removed, placed in fresh tubes and centrifuged again. The resultant supernatant was removed, this time ensuring that no chelex was picked, and stored at -20°C awaiting PCR. 5µl of genomic DNA extracted by chelex was then used as DNA template in PCR.

Isolation and Amplification of DBLα Sequence tags by PCR
5µl of genomic DNA was amplified using DBLα AF’ GCA CG(A/C)AGTTT(C/T)GC (forward primer) and DBLα BR, GCCCATTC(G/C)TCGAACCA (reverse primer) (Bull et al, 2007). 35 cycles of PCR was carried out at a denaturation temperature of 94°C, annealing temperature of 42°C and 65°C extension and a final extension of 65°C. Each reaction tube had a total volume of 25µl consisting of 6.56µl ddH2O, 0.25µM of each outer primer, 1× standard PCR buffer (1.5mM MgCl2, 50mM KCl, 10mM TrisHCl(pH8.3), 0.5% DMSO), 200 µm of each of the dNTPs, 1 unit of Taq polymerase (KEMTAQ®) and 8µl of DNA template.

Sequencing of PCR products by 454-Sequencing (Roche), assembly of sequence reads and Defining of DBLα tags
DNA samples from 27 patients, amplified by PCR were sequenced by the 454 sequencing, Roche™ at the International Livestock Research Institute (ILRI), Nairobi campus. The 454-sequence reads were assembled using the Newbler 2.3.5 program from Roche. The SSF files were converted into Fasta format based on quality scores. The reads from each sample were then translated into amino acid sequence tags. One hundred (100) amino acids were used as the cut-off for any single read to be translated from nucleotide to amino acid sequence. The DBLα sequence tags were then isolated from the 454 sequence reads and grouped into contigs consisting of sequence tags in each sample that were similar or had overlapping reads implying they corresponded to the same var gene and/or region and singlets consisting of sequence reads that occurred only once in a sample.

Classifying DBLα sequence tags into cys/PoLV groups
These tags were then analyzed and classified into cys/PoLV groups based on the number of cysteine residues and the motifs at PoLV1 and PoLV2. The mega such motif command was used to highlight cysteine residues in the sequence tags. The number of cysteine residues in each sequence tag was counted manually. This number of cysteines and the motifs at PoLV1 and PoLV2 were used as a basis for classifying the sequence tags in six groups.

Results
All assembled DBLα sequence tags isolate from the two study sites were deposited to Genebank, (Bankit1771917: KP085750 –KP087726). A total of 1198 PfEMP1 DBLα sequence tags were isolated from field isolated from Mbita Study site. 118 (9.84%) sequence tags were classified as group1. These were cys2 sequences with MFK motif at PoLV1. 65 (5.4%) sequence tags from isolates contained two cysteine residues and REY motif at PoLV2 and were categorized as group2 sequences. Group3 sequences are cys2 sequence tags that do not have MFK and REY motifs at PoLV1 and PoLV2 respectfully. 133 sequence tags (11.10%) belonged to group3/group3-like category. Like the other cys2 sequences, the tags were highly conserved in the 5’ region. DBLα sequence tags containing four cysteine residues (cys4) but do not have a REY motif at PoLV2 (11) were classified as group4. While most of the cys2 sequences are known to be DBLα1 domains of group A like var genes, cys4 sequences are entirely associated to non-group A like var genes. A total of 687 sequence tags, constituting 57.35% of all the DBLα sequences isolated from Mbita isolates were group4 sequences. Group5 sequence tags possess four cysteine residues (cys4) with REY motif at PoLV2, 120 DBLα sequence tags, constituting 10.02% of all sequence tags from Mbita isolates were classified as group5.

Sequence tags with zero, one, three, five or six cysteine residues were classified as group6. There were 40 sequence tags that fell into this group. This constituted 3.34% of all the DBLα sequence tags isolated from patient samples collected from Mbita study site. 18 of these sequence tags had three cysteine residues (cys3) and constituted 45% of all group 6 sequences. Cys5 sequences were the next most abundant group 6 sequences, with 13 sequence tags, constituting 32.50% belonging to this group. Cys1 and cys0 were the least frequent, with 5 (1.25%) and 4 (1%) sequence tags respectively.
Grouping of DBLα sequence tags from Tiwi isolates based on the number of cysteine residues and positions of limited variations (cys/PoLVs)

A total of 783 DBLα sequence tags were isolated from field isolates from Tiwi. Sequence tags containing MFK motif at PoLV1 and lacking REY motif at PoLV2 were classified as group1 sequences. A total of 94 sequence tags, constituting of 12.08% sequence tags from Tiwi were classified as group1 sequences. 4.5% of Cys2 sequence tags containing REY motif at PoLV2 but without MFK motif at PoLV2 were classified as group 2 sequences.

Sequence tags containing two cysteine residues but lacking MFK and REY motifs at PoLV1 and PoLV2 respectively were classified as group3 sequences. 85 sequence tags from Tiwi field isolates, constituting 10.93% were classified as group3 sequences. Sequence tags containing four cysteine residues (cys4 sequences) that lacked REY motif at PoLV2 position were classified as group 4 sequences. Thus 444 (57.07%) sequence DBLα tags from Tiwi isolates were classified as group 4. This group constituted the majority of the sequence tags extracted from Tiwi field isolates. 87 (11.18%) sequence contained four cysteine residues (cys4 sequences) with REY motif at PoLV2. They were classified as group5 sequences. 38 DBLα sequence tags (constituting 4.88%) had either one, three, five, six or no cysteine residues and were classified into group 6.

Discussion

In the sequence tags isolated from Mbita field isolates, group4 sequences were the most prevalent, accounting for 57.35% of all the sequence tags. This was followed by group3, (11.10%), group5 (10.02%), group1 (9.84%) then group2 (5.40%). Group6 had the least frequency (3.34%). Groups3, 4 and 5 DBLα sequence tags are all cys4 sequences that contain four cysteine residues. Cys4 sequences have been associated with uncomplicated and mild malaria symptoms of malaria (8). One of the inclusion criteria of subjects into the study was that they had to be positive for malaria but showed uncomplicated (non-severe) malaria symptoms. Parasites causing uncomplicated malaria are more likely to express non-group A var genes which are characterized by possessing two cysteine residues (cys2 sequences). Expression of group A var genes, that is cys/PoLV group1, is characteristic of non-immune patients who end up developing severe symptoms of malaria like cerebral malaria, convulsion and severe anemia. The study was carried out in hyper-endemic areas of malaria where transmission occurs throughout the year with peak transmission corresponding to long rains between March and June. There is therefore a high likelihood that most of the patients were semi-immune, having been severally exposed to malaria and possessing anti-malarial memory T- and B-cells that limited the severity symptoms (12). However, since this study targeted individuals with non-severe malaria symptoms, the exact implication of this observation could not be determined. It cannot therefore be stated that these patients could have progressed to develop severe symptoms of malaria since existence of sequences at genomic level does not necessarily result in a similar expression pattern.

Figure No. 1 is a graph showing the comparison between the distribution of the cys/PoLV groups between DBLα sequence tags from Mbita and Tiwi field isolates. Although the frequency of group1 sequences appeared to be slightly higher among sequence tags from Tiwi than those from Mbita isolates, the difference was not statistically significant (P value = 0.359, Fishers exact test). These sequences were from four and twenty five Tiwi and Mbita field isolates respectively. The results suggest that the occurrence of group1 DBLα sequence tags per isolate could be fewer in the Mbita isolates compared to isolates from Tiwi study site. It has been established by (12) that a subset of group A-like var genes encoded ligands bind to brain endothelial cells. The same study also demonstrated that infected cells selected for high binding affinity to brain endothelial also bound antibodies from African children with severe malaria at a higher affinity than the non-selected infected cells. It is expected that DBLα sequence tags from group A var genes be present in genomic DNA of parasite isolates from patients at both study sites. However, the implication of these sequence tags on disease severity could not be established from genomic data. Neither was it possible to establish the expression pattern of the var genes based on genomic DNA. It was also not possible to establish the role played by the patient’s immunity status in limiting virulence of the parasite.

The frequency of Group2 DBLα was field isolates from the two sites was almost similar (5.4% and 4.5% in isolates from Mbita and Tiwi respectively). However, samples from both sites yielded fewer group2 DBLα sequences compared to group1 sequences. Cys2 sequences are expressed by but not limited to group A var genes (8). While group1 sequence tags are usually indicative of expression of group A var genes, the other cys2 sequences can belong to either group B or group C.

Cys/PoLV group 3 sequences tags had an almost similar frequency in parasite isolates from both study sites. The same distribution pattern was observed in group 4 sequence tags. These group of sequence tags had the highest frequency in isolates from both sites. They constituted 57% of all sequence tags from isolates from both sites. Cys 4 sequences have been largely associated with mild malarial symptoms. They are also associated with expression of non-group A var genes. When expressed, sequences from these var genes could aid in the evasion of the host immune responses by the parasite. They may also be responsible for maintaining the equilibrium
between host immune responses and persistence of *P. falciparum* infection in semi-immune hosts within malaria endemic zones. This may in turn maintain a constant population of parasites in circulation that sustain host responses, which keep the individual semi-immune to malaria infection. Such responses and immune memory maintained would limit the infection by *P. falciparum* or even resolve progressive infections and thus limit the development of severe malarial symptoms such as cerebral malaria, severe malarial anaemia and respiratory distress. The average age of the study participants was 3.40 years and having been born in malaria endemic zones, are likely to have had several previous infections of *P. falciparum*, therefore had likely started developing some level of immunity although may have not become totally semi-immune as compared to mature individuals within the population. This is because malaria immunity is developed in piece-meal, since it takes a long period for a person to be exposed to the various repertoires of PfEMP1, due to differential expression of var genes within the parasite and the high antigenic variation that accompanies expression of these parasite antigens. However, the immunological and clinical significance of the high frequency of group 4 DBLα sequence tags in genomic DNA from the field isolates could not be established.

Group6 sequence tags were least among the sequences isolated from Mbita field isolates and second last among sequence tags isolated from Tiwi isolates. Studies from Brazil have indicated that deletion of 1-2 cysteine residues could be associated with severe non-cerebral malaria (13). However, the exact role played by group 6 sequences in genomic DNA in disease progression could not be established.

In conclusion, there was a similar distribution pattern of the types DBLα sequence tags in field isolates from Mbita and Tiwi study sites, with group 4 sequence tags being the most frequent in isolates from both sites. However, in the absence of expression data, the impact of this genomic distribution pattern on malaria pathology remains unknown. Secondly, although their impact could not be established, presence of group 1 sequence tags in genomic DNA could have indicated the potential of patients from both endemic sites being at risk of developing severe symptoms of malaria. This is because the expression of sequence type is more important in influencing the binding capacity of infected erythrocytes to endothelial cells rather than the number of sequences expressed.

References


Figure 2: Distribution of Cys/PoLV Groups of PfEMP1 DBLα Sequence Tags in Field Isolates from Mbita and Tiwi, Two Malaria Endemic Sites in Kenya
The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage: http://www.iiste.org

**CALL FOR JOURNAL PAPERS**

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

**Prospective authors of journals can find the submission instruction on the following page:** [http://www.iiste.org/journals/](http://www.iiste.org/journals/)  All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

**MORE RESOURCES**


**IISTE Knowledge Sharing Partners**

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar