Mapping of DBLα Sequence Tags of Field Isolates from Two Malaria Endemic Sites in Kenya

Francis W. Makokha[1,2*]   Peter C. Bull[4]   Francis T. Kimani[5]   Charity Hungu[5]   Nathan Shaviya[6]
Gabriel Magoma[3]   Sabah O. Ahmed[4]

1.Directorate of Research and Development, Mount Kenya University, 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.Kenya Medical Research Institute-Wellcome Trust, P.O Box 230-80108, Kilifi
5.Centre for Biotechnology and Research Development, Kenya Medical Research Institute, P.O Box 54840-00200, Nairobi, Kenya
6.Masinde Muliro University of Science and Technology, Department of Biomedical Sciences, P.O 190-50100, Kakamega, Kenya

Abstract
Plasmodium falciparum Erythrocyte Membrane Protein 1 (PfEMP1) found on the surface of infected erythrocytes (IEs) mediate antigenic variation during *P. falciparum* infection enabling the parasite evade host immune responses and prolong infection. These molecules mediate binding of IEs to host endothelial cells and uninfected erythrocytes. Cytoadhesion of IE to host cells leads to sequestration in tissues and PfEMP1 is thought to play an important role in parasite virulence. Here we analysed 1725 sequence tags sampled from the DBLa region of PfEMP1 encoding "*var*" genes from 27 patients in two different geographical regions in Kenya, Mbita in Western Kenya and Twiga on the Kenyan coast. The objective of this study was to construct a network to assess the extent of shared position specific polymorphic blocks (PSPBs) in sequences isolated from genomic DNA of field isolates from the two malaria endemic sites in Kenya.. Sequences from Mbita study site and those from Tiwi largely clustered into separate giant networks with only a limited number of sequences from the two sites linking to each other. This observation suggests that the parasite populations from the two endemic sites could be genetically varied and that PfEMP1 sequencing could be a useful tool of understanding the genetics of parasite populations. Thus the network approach of studying relationships between DBLα sequences is a useful tool of uncovering the genetic structure of parasite populations circulating in different malaria endemic regions.
Keywords: PfEMP1, Networks, Position Specific Polymorphic Groups, DBLα, Malaria, *P. falciparum*

Introduction

Drug resistance (1) and ability of the parasite to evade host immune responses remain major challenges in the control and eradication of malaria. One strategy employed by *P. falciparum* is to vary its molecules exported to the surface of the infected erythrocytes (IEs). One of the main parasite proteins found on IEs is *Plasmodium falciparum* membrane protein 1 (PfEMP1). PfEMP1 is coded for by a group of about 60 *var* genes per haploid genome of the parasite (2). Based on their upstream (UPS) sequences, var genes have been classified into group A-E. The UPS group A and B *var* genes are associated with sub-telomeric regions of chromosomes and they are transcribed away from the telomere. Group C genes are associated with internal *var* clusters (3–7). This classification groups also have functional significance with expression of group A being associated to immune naïve patients and severe malarial symptoms (6). A study by Tempo *et al* (16) revealed that group A genes were up-regulated in most severe cerebral malaria while group B *var* genes were up-regulated in cerebral malaria that did not show vascular pathology.

Var genes are very diverse due to many recombination events making it difficult to amplify and study the whole gene. They are instead studied by designing primers that amplify short sequence tags in the DBLα domain of PfEMP1 (8). Motifs within DBLα sequence tags have been used to classify these sequences into six groups (8). This classification system is based on the fact that DBLα sequence tags of PfEMP1 contain regions that show limited variability at specific positions known as positions of limited variability (PoLVs) and also have a characteristic number of cysteine residues (9,10). The sequence tags are sometimes named based on the number of cysteine residues in the sequences. For instance a DBLα sequence tag containing two cysteine residues is a cys2 sequence. Each sequence tag contains four positions at which PoLVs are defined, 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 (10). 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.

DBLα sequence tags are highly polymorphic and not easy to study using basic phylogenetic methods, especially when the sequence number is large. Instead a simple non-phylogenetic approach can be adopted. This approach has been based on the observation that sequences that are dissimilar when analysed globally can none the less contain short regions of exact local alignment within highly polymorphic regions of the sequence. This is an ad hoc approach that does not rely on any evolutionary model. In this approach, position specific polymorphic blocks (PSPBs) are used to construct networks (10). In the network, sequences are represented by vertices that are joined by lines if the sequences share one or more PSPBs. Our study utilized DBLα sequence tags isolated field isolates from two malaria endemic sites in Kenya with the aim of construction a network to assess sharing of polymorphic blocks in these sequences.

Methodology

Study sites, sample collection and study approval

Samples were collected from patients visiting Mbita Sub-District Hospital in Homabay County, Western Kenya and Tiwi Health Centre in Kwale County, Coastal region during drug efficacy studies. All participants showed mild symptoms of malaria. Those showing severe malarial symptoms were referred for admission and were not included in the study. This study was approved by Scientific and Ethical Review Committees of Kenya Medical Institute (KEMRI). Blood spot samples on Whatman filter paper were collected from consenting patients after microscopic confirmation that they were positive for *P. falciparum* mono-infection. Filter papers were air dried and then transported to KEMRI, malaria laboratory for analysis.

DNA Extraction and Isolation of DBLα Sequence tags

DNA from twenty seven (27) field isolates (23 from Mbita in Western Kenya and 4 from Tiwi on the East along the Coastal region of Indian Ocean) 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 4^0C. 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 100^0C 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 -20^0C 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' GCACG (A/C) AGTTT(C/T) GC (forward primer) and DBLα BR, GCCCATTC (G/C) TCGAACCA (reverse primer) (8). 35 cycles of PCR was carried out at a denaturation temperature of 94^0C, annealing temperature of 42^0C and 65^0C extension and a final extension of 65^0C. Each reaction tube had a total volume of 25µl consisting of 6.56µl ddH$_2$O, 0.25µM of each outer primer, 1× standard PCR buffer (1.5mM MgCl$_2$, 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 454 sequencing, Roche^{TM} at the International Livestock Research Institute (ILRI), Nairobi campus. The 454-sequence reeds 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 using Newbler software.

Sequences used in this study

Sequence   data   GenBank   KP085750-KP087726 (http://www.ncbi.nlm.nih.gov/nuccore/?term=KP085750:KP087726[accn]) was used in this analysis. Briefly, 1784 *var* sequences isolated from genomic DNA of 27 field isolates were sequenced. The sequencing was done

by 454-Sequencing, Roche, at International Livestock Research Institute (ILRI), Nairobi campus. Of the 27 isolates, 23 were collected from Mbita while 4 were from Tiwi. A total of 1005 and 778 DBLα sequence tags were generated from isolates collected from Mbita and Tiwi respectfully. Sequences lacking a 5' DIGDI and 3' PQFLR or PQYLR were excluded from analysis and hence were not included in construction of the networks.

Construction of networks

Network construction was performed using the method of Bull et al 2008. A perl script was developed from this method (we will refer to a github script). It was used to extract four blocks of amino acids from specific windows of DBLa sequences defined by three anchor points (10).. Determination of sequences with shared PSPBs and formatting of the information for import into network analysis package was done using Excel spreadsheet functions.

Visualization of networks

Pajek software (V. Batagelj, A. Mrvar: Pajek – Program for Large Network Analysis. http://vlado.fmf.uni-lj.si/pub/networks/pajek/) was used to draw and visualize networks. Kamada Kawai algorithm (11) and the Fruchterman Reingold algorithm (12) within Pajek were used to draw 2D and 3D networks respectively. 3D networks were exported as *.wrl files and visualized using Cortona virtual reality modeling language client 4.2 software (http://www.parallelgraphics.com or http://software.filefactory.com). Within the network each *var* sequence was represented by vertex with an edge being formed between two vertices that shared one or more PSPBs region. During this analysis no weighting was given to edges with respect to the number of PSPBs shared. Visualization of the divisions of the sequences into cys/PoLV groups and block sharing groups was achieved through formatting the data as Pajek partition files. The data was formatted as Pajek vector files so as to be visualized.

Cys/PoLV sequence grouping

Sequences were initially classified using positions of limited variability (PoLV) based on the Bull *et al* system (8,13). Features used to group the sequences into one of six 'cys/PoLV groups' included PoLV1 motif, the PoLV2 motif and the number of cysteine residues within the tag sequence. Group 1 had MFK motif at PoLV1 and 2 cysteine residues, group 2 had a REY motif plus two cysteine residues, group3 had two 2 cysteine residues but lacked MEK or REY motifs at PoLV1 and 2 respectively, group 4 consisted of sequence tags with four cysteine residues but lacked REY motif at PoLV2, group 5 were sequence tags with four cysteine residues and REY motif at PoLV2 while group 6 consisted of sequences with 0, 3, 5 or 6 cysteine residues.

Searching for PSPBs within the sequences collected worldwide

The 14 aa PSPBs from block-sharing group 1 and 2 genes were used to search Fasta files of sequences for hits to any of the PSPBs associated with that block-sharing group. To test for overlap in genes containing 14 aa PSPBs from block-sharing groups 1 and 2, the number of cys/PoLV group 2 genes were counted from the *var* network that matched PSPBs from blocks haring group 1 only, the number that matched PSPBs from block-sharing group 2 only, the number that matched PSPBs from both block-sharing group 1 and 2 and the number that did not match any. These numbers were expressed as a 2 x 2 table, and Fisher's two-sided exact test was used to determine whether there were less sequences that matched both block-sharing group 1 and 2 PSPBs than would be expected by chance.

Global sequence alignment and tree construction

Sequences were aligned using MUSCLE (14) using default parameters. Neighbour-joining trees were constructed using MEGA 3.1 (15). Alignments were visualized using Genedoc (http://www.nrbsc.org/gfx/genedoc/index.html).

Results

The sequence tags appeared as shown below before analysis:

DIGDIIRGKDLFIGYDERDRKEKQKLQQSLKEIFQKIQENNPDLKTLKDEEIREYWWALNRKEVWKSLTCSEQLKGNKYFRGTCSDSGSPSMARDKCTCNNGDVPTYFDYVPQFLR

DIGDIVRGKDLFYGNTHESKQRIILENNLKTIFKKIYDKLDGKNGKTNGIEERYKDTDNYYELREDWWYANRAKVWYAMTCGAPDNAEYFRKTPCGGGKSSTPNKCRCKKEDGTHDSDQVPTYFDYVPQYLR

DIGDIVRGRDLYLGDNRKDREQREKLEKNLKEIFGIIYGKLKDKEAQTRYQDDGPNYYQLREDWWEENREKVWKAITCDVVSGKYFRATCGSGKTQTRDNCQCINFSVPTYFDYVPQYLR

DIGDIIRGKDLFLGHQQRKRKLEGNLRNIFKNIHDHLTDAKANSYYKNDNDRNYYKLREDWWTVNRDQVWKAITCDAAVEDTYSKYLGDRTTTFSNEKCGHDEGIVPTNLDYVPQYLR

DIGDIVRGKDLYSGNNKEKNRREKLENNLKTIFGNIYEELTTTNGVKSRYNDDTPDFFKLREDWWTANRETVWKAMTCSEDLKDASYFRPTCSDSHRKGSCSQANKYCRCNGDKPGEDKPNIDPPTYFDYVPQFLR

DIGDIIRGKDLYLGNSREKEKLQTNLKSIFQNIYKNLKNPAQNHYRDPHGNYYKLREDWWTANRDQVWKALTCFADGSEEYFIQSENNTQLFSNPKCGHEQGNVPTNLDYVPQFLR

DIGDIVRGRDLYRGGNNKRRQQLDENLQKIFTQIYNDVTSGKNVDKAKERYKDTKNYFQLREDWWDANRSTVWEAITCNAQGNTYFRATCSDRNGSFSQAKDKCRCEKKGGGNVNIVPTYFDYVPQYLR

DIGDIIRGKDLFLGHQQRKIQLEERLKTMFENIRNENNEKLKSLTDDQIREYWWALNRKDVWKAMTCDARDNDKYFRNTCAGGKLTEGYCRCDGDKPKADKANVDPPTYFDYVPQYLR

IFGDIYNELTSGRNGMKSAEVKERYKDDTPDFFKLREDWWTANRETVWEALTCDAPEDASYFRPTCSDRQGEAQANHKCQCPNGKDQVPTYFDYVPQYLR

DIGDIVRGKDLFYGNPQEKEKRKQLDKNLKEIFKNIKKENKSKLKSLKDEQIREYWWYANRSTVWKAITCSEDLKNSSYFRQRACAGKTATDDKCRCNGDQVPTYFDYVPQFLR

DIGDIVRGKDLYLGNPQEKKQRDQLENNLKTIFKNIYEKLLEENQKNEKKGEIETRYNDTDKNYYKLREDWWNANRATVWKALTCDAPKDSKYFRGTCGGDENTATQASHKCRCGDGKSKPGKAGEVNIVPTYFDYVPQYLR

DIGDIIRGKDLYLGNKRQNETEREKEILQKNLKNIFAKIYGDLSESAKKHYGDDENYYQLREDWWDANRLEVWKAITCHVVSGNNYFRRTCSKGQGWTQGNCRCVTDVPTYFDYVPQYLR

DIGDIVRGKDLFLGAPNKEKIKLEENLKKIFDNIKNENAELSKLSLEKVREYWWAIHRKELWEALTCNAPKGANYFVYKLDGPKFSSDRCGHNYNGDPLTNLDYVPQYLR

DIGDIVRGKRSHFMVIHKKSAQREQLDKKLKEIFAKIHEDVMKTSRNNKEVLKTRYKNDTENYYKLREDWWTANRAKVWEAMTCSDELRGASYFRATCSDSADGKSQSQARNQCRCQKKNGQHDTDQVPTYFDYVPQFLR

The analysis approach was based on a rationale described by (10). The rational presupposes that blocks of sequences in highly polymorphic regions can be shared by two sequences that are otherwise very dissimilar when compared through global sequence alignment (10) . Since DBLα sequence tags are highly variable, their alignment requires introduction of gaps. This was overcome by restricting the analysis to ungapped polymorphic sequence blocks locked within *var* sequences fixed relative to one of the three conserved anchoring points at each end and in the middle to provide four independent windows. Single sequences were then used within these windows to find out if two sequences were identical within any of the sequence blocks. Thus sequences were analysed as multiple independent blocks, with each block, (that is the 'position specific polymorphic block' [PSPBs]) acting as a genetic marker of the sequence to which it was anchored. The PSPBs were used to construct networks. In the network, the nodes (vertices) representing sequences were joined by lines (edges) if they were identical at one or more of their constituent PSPBs.

The cys/PoLV groups have been used to classify DBLα sequence tags into six groups. Figure 1A shows the network of sequence groups 1-3 in which each group is represented by a different colour (Brown represents group1; red, group 2 and gold group 3)
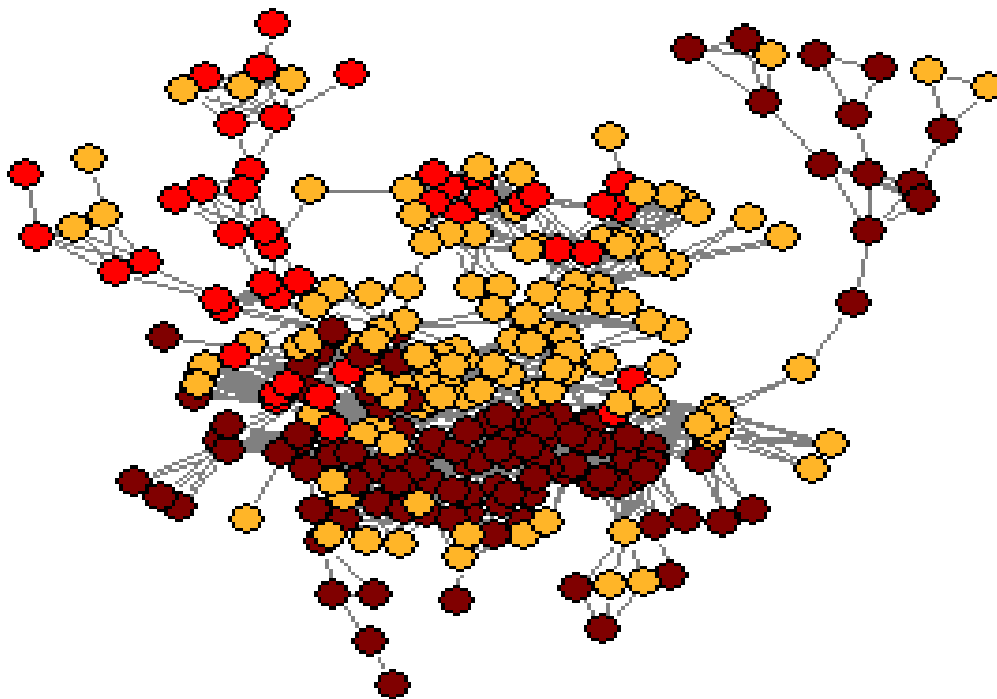
**Figure 1A: A network of sequence groups 1-3**

Some vertices representing DBLα sequence tags belonging to different cys/PoLV groups were joined. This indicated shared PSPBs on the sequence tags. Vertices representing some group3 sequence tags had edges

linking them to groups 1 and 2 separately without any of these group 3 sequences being linked to groups 1 and 2 simultaneously. However, when groups 1-3 sequences from Mbita and Tiwi isolates were analysed together, they clustered separately into two distinct groups as shown in the figure 1B, in which brown colour represents group1-3 sequence tags from mbita isolates while red represents group 1-3 sequence tags from Tiwi. This results suggested that parasites circulating in the two study sites tend to have have distinct DBLα sequence tags that do not share PSPBs.
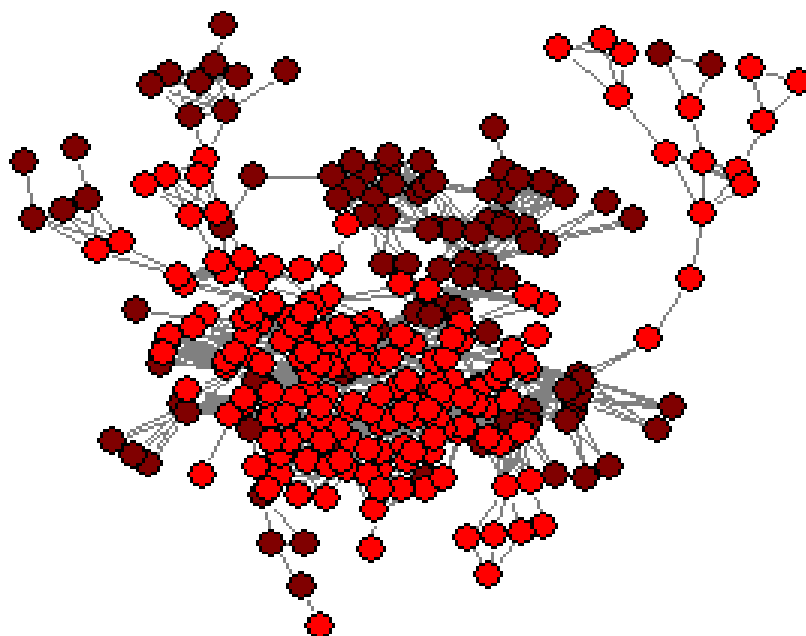

Figure 1B: Groups 1-3 sequences from Mbita and Tiwi isolates (Brown=Mbita sequence tags, Red=Tiwi Sequence tags)

In the analysis of group 4 (blue) and group 5 (green) sequence tags (Figure 1C) some group four sequence tags linked to group 5 sequences suggesting share PSBPs between these sequences. This shows that group 4 sequences frequently shared PSBPs.
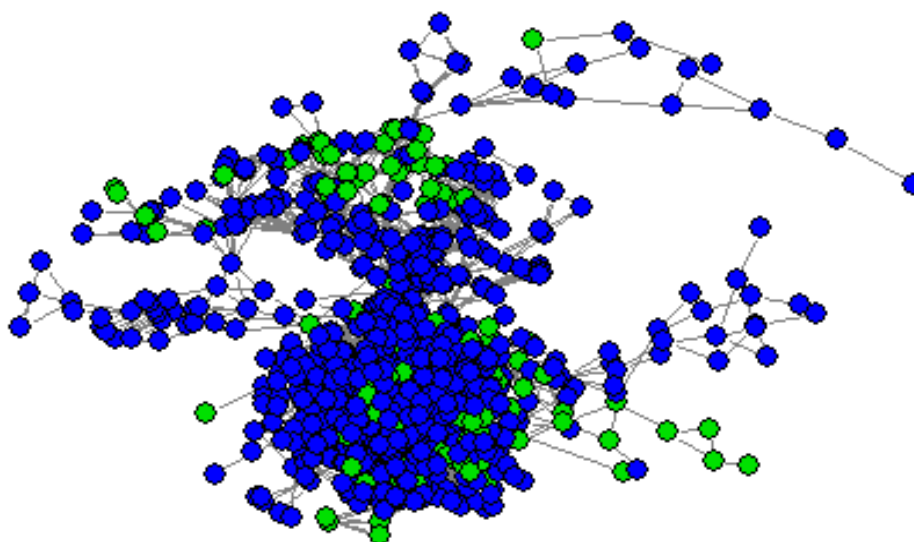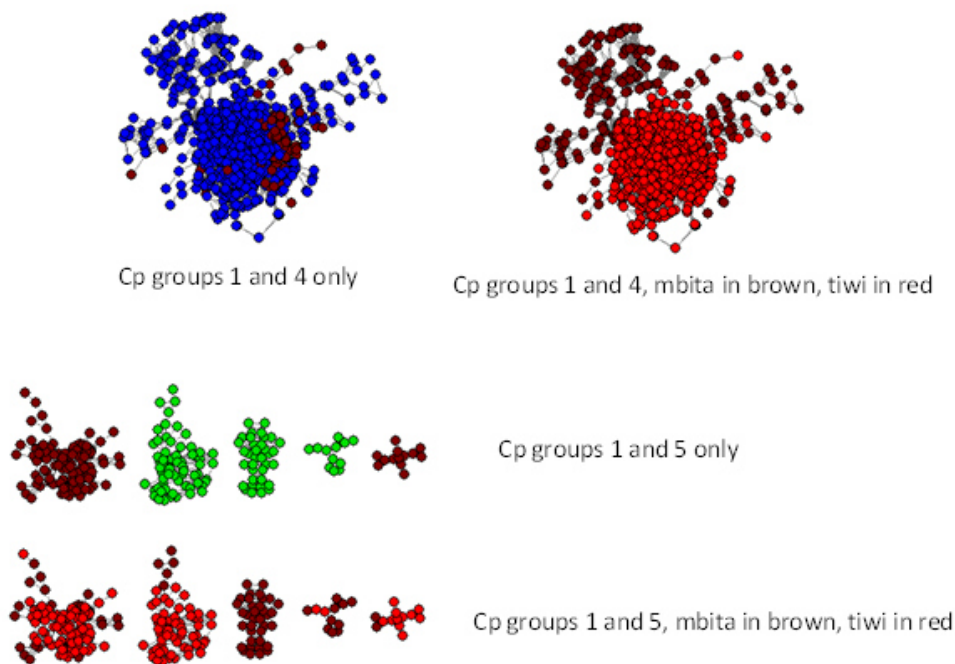

Figure 1C: Group 4 (blue) and group 5 (green) sequence tags

When group 4 and group 5 sequence tags were analysed together (figure 1D), only a few sequence tags from both sites were linked together (Mbita= brown, Tiwi=red).
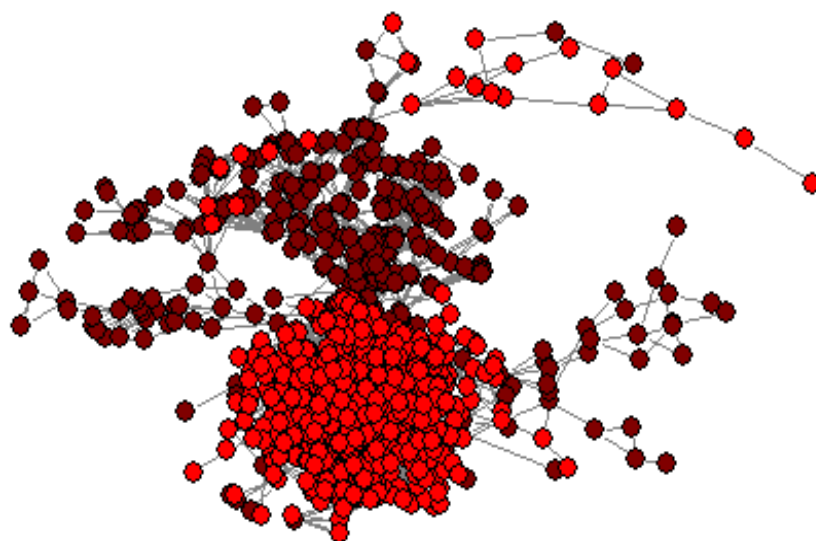
Figure 1D: Group 4 and group 5 tags from both study sites (Mbita=brown, Tiwi=red)

The analysis also revealed that occasionally, groups 1 and 4 DBLα sequence tags share PSPBs. However there were no shared PSPBs between group 1 and 5 sequence tags. This is shown in figure 2A below. The figure does not show giant components but of the smaller.

Cp groups 1 and 4 only

Cp groups 1 and 4, mbita in brown, tiwi in red

Cp groups 1 and 5 only

Cp groups 1 and 5, mbita in brown, tiwi in red

Figure 2A: Smaller components showing occasional sharing of PSPB between 1 and 4 but never between 1 and 5

Sequences tags belong to group 1 and were also analysed together. The figure 2A below shows the giant block of this analysis
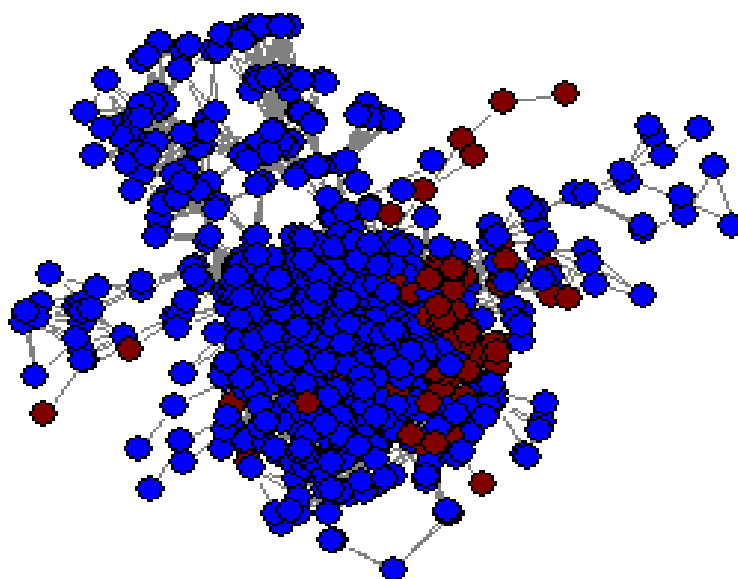
Figure 2B: A giant component showing occasional sharing of PSPB between 1 and 4 (brown=group 2, blue=group 4)

The analysis showed group 1sequences (brown) occasionally clustered within the group 4 (blue) sequences due to shared PSPBs. It should however be noted that group one sequences were much fewer than group four sequence tags. When these sequences from the two endemic sites were analysed together,the results were as shown in figure 2C below (Mbita=brown and Tiwi=Red):
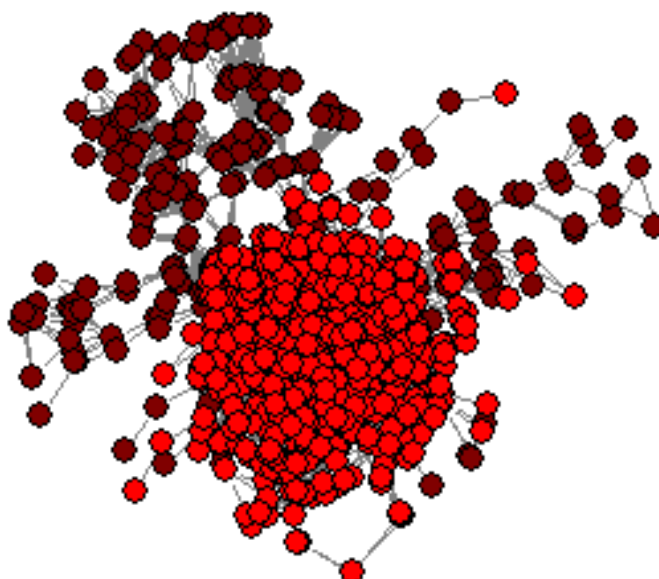

Figure 2C: A giant component of group 1 and 4 sequence tags from Mbita (brown) and Tiwi (red)

These results indicated that group 1 and group 4 sequence tags from the two endemic sites did not tend to share similar PSBPs, hence the clear separation between the two groups of the sequence tags. Very few sequences from Tiwi shared similar PSPBs with those from Mbita at genomic level. Figure 3A shows a visualization analysis of giant components of group 1 (green) and group 5 (Brown) from Mbita isolates. In this analysis, group 1 sequence tags break into three clusters while group 5 sequence tags clustered in 2. None of these groups had their sequences clustering together
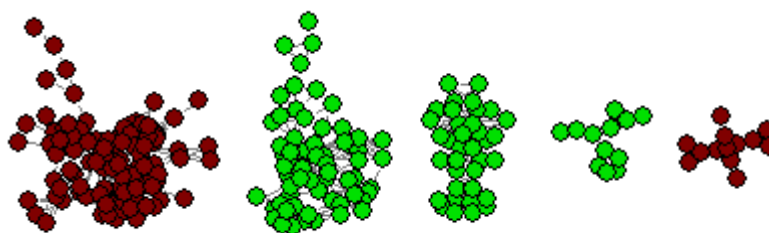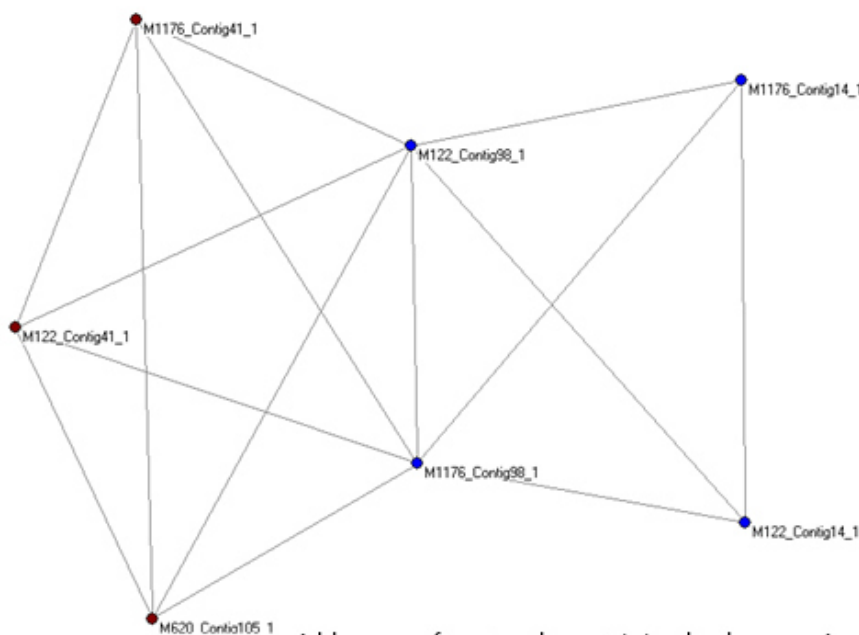
Figure 3A: A visualization analysis of giant components of group 1 (green) and group 5

Figure 3B shows group 1 group 4 sequences only from Mbita (brown) and Tiwi (red), again these sequence tags segmented differently showing lack of similarity at genomic level.
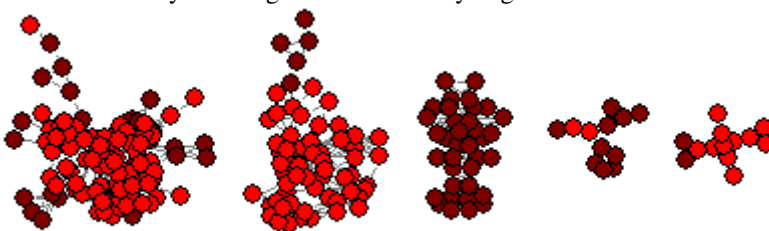
Figure 3B: Group 1 group 4 sequences from Mbita (brown) and Tiwi (red)

Another unique observation from analysis of these data is the fact that a few groups 1 and 4 sequence tags from both study sites linked to each other. This indicated that these sequence tags shared PSPBs. It could not be established however if these observation would be the same at the expression level. Figure 4 below shows these sequence tags linked to each other when the network containing both groups 1 and 4 sequences was expanded for closer observation.

M1176_Contig41_1
M1176_Contig14_1
M122_Contig98_1
M122_Contig41_1
M1176_Contig98_1
M122_Contig14_1
M620_Contig105_1

A blow up of network containing both group 1 and 4 sequences
So you can identify and look more closely at the sequences

Figure 4: An expansion of a network of both group 1 and 4 sequences to allow a closer visualization of sequences

Discussion

In this analysis, DBLα sequence tags isolated from genomic DNA of field isolates from two malaria endemic sites were used to draw networks to ascertain their relationship (http://www.ncbi.nlm.nih.gov/nuccore/?term=KP085750:KP087726[accn]). This analysis shows clearly the

tremendous level of diversity that is present in these sequences. Sequences from Mbita in western Kenya and Tiwi, coastal region clustered separately within the same giant component of the network. This suggested that there were few shared polymorphic groups within these sequences. For instance when groups 1-3 sequences from Mbita and Tiwi isolates were analysed together, they clustered separately into two distinct groups, suggesting that parasites circulating in the two study sites had distinct DBLα sequence tags that did not share PSPBs. A previous study, (10, 13) observed that block sharing groups generated independently world network showed geographical structuring. Although the sequences used in this analysis were isolated from parasites in the same country, they seemed to share very few position specific polymorphic blocks. This could be due to the geographical distribution of the parasites in that the isolates from Mbita and those from Tiwi (Mbita and Tiwi are located approximately 1100KM apart). The two sites are thus separated in such a way that there is reduced chance of genetic recombination between the parasite isolates within the mosquito vector. Since these sequences were isolated from genomic DNA, it remains to be seen if the situation is similar at the expression level. It was also not possible to determine the effect of this observed difference on the ability of host immune responses from Mbita to effectively respond to parasite from Tiwi. That is, the ability of semi-immunity acquired from exposure to *P. falciparum* parasites circulating within Mbita to protect against infection of *P. falciparum* parasites circulating in Tiwi and vice versa.

Conclusion

The observation from this study indicates that the parasite populations circulating at the two endemic sites could be genetically varied as evidenced by the fact that only a few sequences shared PSPBs. Thus PfEMP1 sequencing can be a very useful tool of understanding the genetics of parasite populations from different regions of varied malaria endemicity. The network approach of studying relationships between DBLα sequences is a useful tool of uncovering the genetic structure of parasite populations circulating in different malaria endemic regions. This approach may be useful in determining the use of PfEMP1 as vaccine candidate. Only sequences shared by majority of parasite population would be useful in inducing protective immune responses against parasites circulating in different regions. These observations indicate that DBLα sequence tags from field isolates from Mbita and Tiwi study sites in Kenya are varied and only few share PSPBs. The majority of the sequence tags, however, do not share PSPBS, suggesting that they totally different

References

1. Khalid B Beshir CJS. Residual Plasmodium falciparum parasitemia in Kenyan children after artemisinin-combination therapy is associated with increased transmission to mosquitoes and parasite recurrence. J Infect Dis. 2013;

2. Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. PLoS Comput Biol. 2010;6(9).

3. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. 2002 Oct 3;419(6906):498–511.

4. Berger SS, Turner L, Wang CW, Petersen JEV, Kraft M, Lusingu JPA, et al. Plasmodium falciparum Expressing Domain Cassette 5 Type PfEMP1 (DC5-PfEMP1) Bind PECAM1. PLoS ONE. 2013 Jul 9;8(7):e69117.

5. Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, et al. A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. Proc Natl Acad Sci U S A. 2012 Jun 26;109(26):E1772–81.

6. Warimwe GM, Keane TM, Fegan G, Musyoki JN, Newton CRJC, Pain A, et al. Plasmodium falciparum var gene expression is modified by host immunity. Proc Natl Acad Sci U S A. 2009 Dec 22;106(51):21801–6.

7. Bull PC, Kyes S, Buckee CO, Montgomery J, Kortok MM, Newbold CI, et al. An approach to classifying sequence tags sampled from Plasmodium falciparum var genes. Mol Biochem Parasitol. 2007 Jul;154(1):98–102.

8. Bull PC, Kyes S, Buckee CO, Montgomery J, Kortok MM, Newbold CI, et al. An approach to classifying sequence tags sampled from Plasmodium falciparum var genes. Mol Biochem Parasitol. 2007 Jul;154(1):98–102.

9. 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 Dec 30;4(12):e8497.

10. 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 Jun;68(6):1519–34.

11. Kamada T, Kawai S. An Algorithm for Drawing General Undirected Graphs. Inf Process Lett. 1989 Apr;31(1):7–15.

12. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exp. 1991;21(11):1129–64.

13. Bull PC, Berriman M, Kyes S, Quail MA, Hall N, Kortok MM, et al. Plasmodium falciparum Variant Surface Antigen Expression Patterns during Malaria. PLoS Pathog [Internet]. 2005 Nov [cited 2015 Mar 14];1(3). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1287908/

14. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

15. Kumar RA, Sridevi K, Kumar NV, Nanduri S, Rajagopal S. Anticancer and immunostimulatory compounds from Andrographis paniculata. J Ethnopharmacol. 2004 Jun;92(2-3):291–5.

16 Tembo DL, Nyoni B, Murikoli RV, Mukaka M, Milner DA, et al. Differential fEMP1 Expression Is Associated with Cerebral Malaria Pathology. PLoS Pathog 10(12): 2014. e1004537. doi:10.1371/journal.ppat.1004537
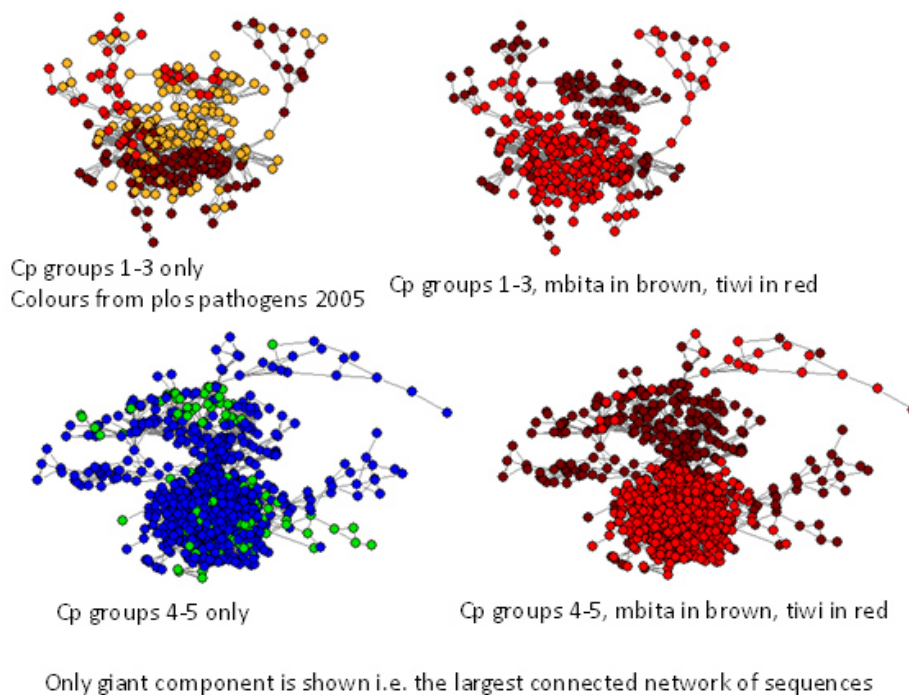
Figure 5: Giant Components within networks (Colours adopted from Bull et al, 2005)