Genetic diversity of Duffy binding protein 2 region II of *Plasmodium cynomolgi* from wild macaques in Peninsular Malaysia

Latif, E.N.M.1, Shahari, S.1, Amir, A.1, Cheong, F.W.1, Lau, Y.L.1, Abdullah, M.L.2, Fong, M.Y.1*

1Department of Parasitology, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
2National Wildlife Forensic Laboratory, Ex-situ Conservation Division Department of Wildlife and National Parks Peninsular Malaysia, 56100 Kuala Lumpur, Malaysia
*Corresponding author: fongmy@um.edu.my

**ARTICLE HISTORY**
Received: 30 January 2022  
Revised; 28 February 2022  
Accepted: 28 February 2022  
Published: 31 March 2022

**ABSTRACT**

Recent reports of natural human infection by *Plasmodium cynomolgi* indicate the increased risk of zoonotic transmission by this simian parasite. The *P. cynomolgi* Duffy binding protein 2 (PcDBP2) has a potential role in the invasion pathway of host erythrocytes, and it is a possible vaccine candidate against cynomolgi malaria. This study investigates the genetic diversity, haplotypes, and natural selection of PcDBP2 region II from isolates collected from wild macaques in Peninsular Malaysia. Blood samples from 50 *P. cynomolgi*-infected wild macaques were used in the study. Genomic DNA extracted from the blood samples was used as template for PCR amplification of the PcDBP2 region II. The amplicons were cloned into a plasmid vector and sequenced. MEGA X and DnaSP ver.6.12.03 programmes were used to analyse the DNA sequences. A genealogical relationship of PcDBP2 region II were determined using haplotype network tree on NETWORK ver.10.2. Result showed high genetic diversity (\(\theta = 0.017 \pm 0.002; \text{Hd} = 1.000 \pm 0.001\)) of the PcDBP2 region II. The Z-test indicates a purifying selection, with population expansion as shown in Tajima’s D analysis. A total of 146 haplotypes of PcDBP2 region II were observed. Phylogenetic tree analysis showed that these haplotypes were grouped into three allelic types (136 for Strain B type, 9 for Berok type, and 1 recombinant type). In the haplotype network, PcDBP2 region II revealed no geographical groupings but was divided into two distinct clusters.

**Keywords:** *Plasmodium cynomolgi*; Duffy binding protein; genetic diversity; haplotypes; natural selection.

---

**INTRODUCTION**

Thirteen species of *Plasmodium* are known to infect non-human primates in Southeast Asia (Collins, 1988), and the first description of simian malaria in Malayan macaques was made in 1908 (Daniels, 1908). Presently, majority of the human malaria cases in Malaysia are caused by the simian malaria parasite *Plasmodium knowlesi*. Recent studies from Peninsular Malaysia (Amir et al., 2020), and Singapore (Li et al., 2021) found that another simian malaria parasite, *Plasmodium cynomolgi*, was the most prevalent species in wild macaques and it has been recognized as an emerging threat of zoonotic malaria infection in the country (Dian et al., 2022).

*Plasmodium cynomolgi* was first discovered in a *Macaca fascicularis* (previously known as *Macaca cynomolgus*) monkey from Java in 1907 (Coatney et al., 2003). This parasite species shares many biological features with its sister taxon, *Plasmodium vivax*. Both species have a tertian periodicity (48h) (Mulligan, 1935), prefer to infect reticulocytes (Warren et al., 2003), and form dormant liver stage (hypnozoites) (Krotoski et al., 1982). A study by Sutton et al. (2016) described 9 strains of *P. cynomolgi*, 5 of which were isolated from Malaya, namely M/B, Smithsonian, Pig-tailed I, Berok, and Gombak. The first natural human infection of *P. cynomolgi* was reported in Malaysia (Ta et al., 2014). In recent years, several reports of naturally acquired human cynomolgi malaria have been reported in Southeast Asia (Singh et al., 2018; Grignard et al., 2019; Hartmeyer et al., 2019; Imwong et al., 2019; Nor et al., 2019; Raja et al., 2020; Putaporntip et al., 2021). The increasing number of *P. cynomolgi* cases may suggest the increased zoonotic potential of this parasite. This highlights the possible need for effective strategies in preventing and controlling this zoonotic malaria infection in the Southeast Asian region

*Plasmodium* Duffy binding proteins (DBPs) are parasite ligands that are secreted by micronemes and are highly expressed during the mature schizont stage (Iyer et al., 2007). The *P. cynomolgi* DBPs (PcDBPs) can be classified into 2 groups, DBP1 and DBP2. The genes for these DBP are located on chromosome 3 and 6, respectively (Tachibana et al., 2012).
PcDBPs are known to be ligands that are responsible for red cell tropism of \textit{P. cynomolgi} (Hang et al., 2021). An \textit{in vitro} study on \textit{P. cynomolgi} revealed that its invasion of human erythrocytes is restricted to reticulocytes which express both \textit{transferrin} receptor 1 (CD71) and Duffy antigen receptor for chemokines (DARC, CD234) (Kosaisavee et al., 2017).

Similar to other \textit{Plasmodium} DBPs, PcDBP2 consists of seven regions (I-VII), and it binds to the DARC on the erythrocyte surface (Adams et al., 1992; Batchelor et al., 2011). The N-terminal cysteine rich region II of PcDBP2 contains the binding motif that is necessary for binding to DARC (Chitnis & Miller, 1994; Ranjan & Chitnis, 1999). PcDBP2 region II, therefore, may be a potential vaccine candidate against human cynomolgi malaria (Okenu et al., 1997). However, to date, very little diversity studies have been done on PcDBP2 from Malaysia (Sutton et al., 2016). Here, we present our findings on the genetic polymorphism, natural selection, phylogenetic clustering, and haplotypes of PcDBP2 region II from isolates collected from wild macaques in Peninsular Malaysia.

**MATERIALS AND METHODS**

\textbf{Macaque blood sampling and screening for \textit{Plasmodium} infection}

This study was approved by the University of Malaya Institutional Animal Care and Use Committee (Reference Number: M/06122019/25022019-01/R). Blood samples were collected from wild macaques in six Peninsular Malaysian states (Pahang, Johor, Selangor, Perak, Negeri Sembilan and Kelantan) in 2019–2021, by the Department of Wildlife and National Parks Peninsular Malaysia as part of the on-going Wildlife Disease Surveillance Programme (WDSP). The sampling was conducted based on the protocol of University of California, Davis, IACUC (Protocol No. 16048). Each blood sample was screened for \textit{Plasmodium} infection. Briefly, genomic DNA was extracted from the blood using DNeasy Blood & Tissue Kit (Qiagen, Germany). Nested PCR targeting the \textit{Plasmodium} 18S rRNA was performed on the extracted DNA using the amplification conditions described by Lee et al. (2011). Samples which were positive for \textit{P. cynomolgi} were selected for the PcDBP2 region II study.

\textbf{PCR amplification of PcDBP2 region II}

Fifty \textit{P. cynomolgi} DNA-positive samples were selected for this study. Two primers pairs were designed and used in the Nested PCR: Nested 1 (PcDBP2-F1: 5’-ACGGGAGAGGTGCAAATGCCTTAACTAATG-3’ and PcDBP2-R1: 5’-GTGGTAGAATTATGTGCGCC-3’), and Nested 2 (PcDBP2-F2: 5’-GGCAATCCTTCAACTCTCATTATGCA-3’ and PcDBP2-R2: 5’-CCTCTCCTCTTCTAATGCTATG-3’). The primers were designed based on the alignment of reference sequences of \textit{P. cynomolgi} strains obtained from GenBank (Table 1). PCR cycling conditions for nested PCR were as follows: Nested 1, the amplification was initiated with denaturing at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 53°C for 1 min, and elongation at 72°C for 80 sec. This was followed by 5 min of final extension at 72°C. The reaction for nested 2 was similar, but the annealing temperature used was 52°C and elongation time of 70 sec. The final volume for both amplifications was 25 μL which consisted of 4 μL of genomic DNA, 5X Green GoTaq® Flexi Reaction Buffers (Promega, USA), 4 mM of MgCl₂, 0.2 μM of forward and reverse primers, 0.2 mM of dNTPs and 1 unit of Taq DNA polymerase. Amplicons were resolved by electrophoresis in a 1% agarose gel, stained with SYBR Safe DNA gel stain (Invitrogen, USA) and viewed under UV illumination.

**Table 1. List of \textit{P. cynomolgi} PcDBP2 sequences downloaded from GenBank**

<table>
<thead>
<tr>
<th>\textit{P. cynomolgi} strain</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain B-DBP2</td>
<td>XM_004220981</td>
</tr>
<tr>
<td>Berok-DBP2</td>
<td>JQ422036</td>
</tr>
<tr>
<td>Cambodian-DBP2</td>
<td>AB617789</td>
</tr>
<tr>
<td>Strain B-DBP1</td>
<td>XM_004221494</td>
</tr>
<tr>
<td>Berok-DBP1</td>
<td>JQ422035</td>
</tr>
<tr>
<td>Cambodian-DBP1</td>
<td>AB617788</td>
</tr>
</tbody>
</table>

\textbf{Cloning and sequencing of amplicon}

Amplicon was ligated into the cloning vector pGEM®-T (Promega, USA) and transformed into competent \textit{Escherichia coli} TOP10F' cells. Recombinant clones were screened for the presence of PcDBP2 region II by single step PCR with the following conditions: amplification initiated at 95°C for 4 min, followed by 35 cycles of denaturation at 95°C for 30 sec, annealing at 50°C for 45 sec, and extension at 72°C for 90 sec. The cycles were completed with final extension for 10 mins at 72°C. Amplicons were examined in a 1% agarose gel electrophoresis. In order to detect possible multiple PcDBP2 region II haplotypes, 3 positive clones from each sample were selected and their recombinant plasmid was extracted using QIAprep® Spin Miniprep Kit (Qiagen, Germany). The recombinant plasmids were sent to a commercial laboratory (1st Base Asia, Malaysia) for DNA sequencing.

\textbf{Sequence alignment and phylogenetic analyses of PcDBP2 region II}

Each raw sequence obtained from the commercial lab was trimmed using Gene Runner version 6.5.52 Beta to remove non-PcDBP2 region II sequences. Multiple alignments of PcDBP2 region II nucleotide and deduced amino acid sequences were performed using ClustalW tool in BioEdit v7.2. The sequences obtained were aligned together with PcDBP2 region II sequences of \textit{P. cynomolgi} strain B (GenBank Accession No. XM_004220981) as the reference sequence, and Berok strain (GenBank Accession No. JQ422036). The deduced amino acid sequences were used to construct phylogenetic trees using Neighbor Joining and Maximum Likelihood methods with 1000 bootstrap replicates in MEGA X (Kumar et al., 2018). The \textit{P. vivax} erythrocyte binding protein (PvEBP) region II sequence (\textit{GenBank} Accession No. MN853275) was used as the outgroup in the phylogenetic trees.

\textbf{Genetic diversity, natural selection, and haplotype analyses}

DnaSP ver.6.12.03 (Rozas et al., 2017) was used to determine the nucleotide diversity (\(\pi\)) and haplotype diversity (\(H_D\)) of PcDBP2 region II. To investigate the departure from neutrality, the Tajima’s D test was conducted (Tajima, 1989). Under neutrality, Tajima D is expected to be 0. A positive Tajima’s D value indicates recent population bottleneck, whereas a negative value means population expansion. For natural selection analysis, the non-synonymous (\(d_N\)) and synonymous mutation (\(d_S\)) rates were estimated and compared using the Z-test (\(P < 0.05\)), according to the Nei and Gojobori’s method in MEGA X. When a gene is under positive selection, \(d_N\) would exceed \(d_S\) (i.e., \(d_N/d_S > 1\)), while for negative (purifying) selection, \(d_N\) would be less than \(d_S\) (i.e., \(d_N/d_S < 1\)). For haplotype analyses, median joining approach in NETWORK ver.10.2 was used to generate the relationship of PcDBP2 region II haplotypes (Bandelt et al., 1999).
RESULTS

Genetic diversity and natural selection of PcDBP2 region II

A total of 150 PcDBP2 region II sequences (1002 bp in size) were obtained (GenBank Accession numbers: OM373319 – OM373468) from 50 *P. cynomolgi* DNA-positive samples. The sequences were aligned and analysed for their diversity and natural selection. Table 2 presents the result of genetic diversity and neutrality test of the sequences. Nucleotide diversity ($\pi$) and haplotype diversity ($H_d$) were 0.017 ± 0.002 and 1.000 ± 0.001, respectively. The average rates of non-synonymous ($d_N$) and of synonymous ($d_S$) mutations were 0.01 and 0.03, with $d_N/d_S$ value of 0.33 ($P < 1$), indicating negative (purifying) selection. This is also supported by the Z-test showing significant $d_N < d_S$ ($P < 0.05$). For test of neutrality, Tajima’s D shows a significant negative value (-2.700) for the PcDBP2 region II.

Amino acid polymorphism and phylogenetic analysis

Multiple alignment of 150 PcDBP2 region II amino acid sequences with those of *P. cynomolgi* strain B (GenBank Accession No. XM_004220981), and Berok strain (GenBank Accession No. JQ422036) were generated. All sequences were 334 amino acids in length, with no indels. From the 334 amino acid positions, 238 were polymorphic (Supplementary Data 1). Among these positions, 132 were monomorphic (1 amino acid change), 74 were dimorphic (2 amino acid changes), 31 were trimorphic (3 amino acid changes), and 1 was tetramorphic (4 amino acid changes: position 190 [V → E, I, A, G]).

A total of 146 PcDBP2 region II haplotypes were deduced (Supplementary Data 2), with one haplotype (H27) showing the highest frequency, i.e., 7 out of 152 sequences. Haplotypes H1 and H146 were the reference Strain B and Berok, respectively. Interestingly, one haplotype (H137) was found as a recombinant type which suggested recombination between Strain B and Berok types. Twelve predicted recombination sites were identified, in which 8 (within amino acid positions 75 – 106) were similar to the Berok type, and 4 (within positions 119 – 222) similar to the Strain B type (Supplementary Data 1).

The phylogenetic trees constructed using Neighbour Joining and Maximum Likelihood methods displayed the same profiles (Figure 1A and 1B). Two distinct allele groups, designated as group I and II were observed. Group I was the Strain B type which consisted of 142 sequences, whereas group II was the Berok type with 9 sequences. Located between these two groups was KTLT012-C1, and it was categorized as a recombinant type (corresponds to the haplotype H137 described earlier).

Haplotype network analysis of PcDBP2 region II

The network analysis identified 146 haplotypes of PcDBP2 region II from the 152 sequences. As expected, the haplotype network showed two major clusters, Strain B and Berok types (Figure 2). The Strain B type constituted the larger cluster,

<table>
<thead>
<tr>
<th>Nucleotide diversity ($\pi$ ± SD)</th>
<th>Haplotype diversity ($H_d$ ± SD)</th>
<th>Tajima’s D test</th>
<th>$d_N$ ± SE</th>
<th>$d_S$ ± SE</th>
<th>$d_N / d_S$</th>
<th>Z-test $d_N &lt; d_S$ ($P&lt; 0.05$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.017 ± 0.002</td>
<td>1.000 ± 0.001</td>
<td>-2.700</td>
<td>0.01 ± 0.00</td>
<td>0.03 ± 0.00</td>
<td>0.33</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

π: nucleotide diversity; $H_d$: haplotype diversity; SD: standard deviation; SE: standard error; $d_N$: non-synonymous mutation rate; $d_S$: synonymous mutation rate.

Figure 1. Phylogenetic analysis of Peninsular Malaysia PcDBP2 region II. A) Phylogenetic tree constructed based on Neighbour Joining method. B) Phylogenetic tree constructed based on Maximum Likelihood method. The amino acid sequences were used to construct the phylogenetic tree in MEGA X. The green font indicates outgroup, PvEBP(RII), and the red font indicates the recombinant type.
Figure 2. Haplotype network of Peninsular Malaysia PcDBP2 region II. The haplotypes were constructed with 144 sequences and 2 reference sequences (Strain B and Berok), by using median-joining algorithm in NETWORK ver.10.2. The Strain B, recombinant and Berok types are marked as yellow, red, and blue, respectively. The size of each circle represent the total frequency of the haplotypes.
which consisted of 136 haplotypes with haplotype H1 as the reference for Strain B. For Berok, the cluster consisted of 9 haplotypes (H138 – H146) with haplotype 146 as the reference for Berok strain. The recombinant haplotype (H137) was located in the middle of the two major clusters.

**DISCUSSION**

The Duffy binding protein plays a crucial role in host erythrocyte recognition and invasion of some *Plasmodium* species (Iyer et al., 2007). This protein's region II has been regarded as potential vaccine candidate antigen for *P. knowlesi* and *P. vivax*. The diversity of *P. knowlesi* and *P. vivax* DBP region II has been much studied (Fong et al., 2014; Saat et al., 2017). However, diversity of *P. cynomolgi* DBP has not been fully looked into. Therefore, this study was undertaken to investigate the genetic diversity, natural selection, and haplotypes of *PcDBP2* region II from wild macaque isolates in Peninsular Malaysia.

Our results indicated high level of genetic diversity of *PcDBP2* region II (π = 0.017 ± 0.002; Hₑ = 1.000 ± 0.001). A similar finding is also seen in its paralogue protein, *PcDBP1*, from Thailand (Putaporntip et al., 2016). The *PvEBP* region II, which is the ortholog of *PcDBP2* (M strain), also showed high level of genetic polymorphism in *P. vivax* from 8 different countries (Han et al., 2020). The Z-test (*dN < dS*) with P value less than 0.01 indicates a purifying selection on *PcDBP2* region II, this means it has functional constraints that limit polymorphism within the parasite population. Furthermore, significant negative value of Tajima's D (P < 0.05) may suggest population expansion of *P. cynomolgi*, as evidenced by the current state of zoonotic *cynomolgi* cases in humans. This is possibly due to the destruction or loss of habitats for the macaque and vector populations which leads them closer to the human populations (Vythilingam, 2010). The close interaction between hosts and vectors can be a threat to the human population, hence better strategies are required to prevent the expansion of zoonotic *P. cynomolgi* infections.

The *P. cynomolgi* used in this study were isolated from wild macaques from 6 different states in Peninsular Malaysia. Phylogenetic analysis on *PcDBP2* region II, however, showed no specific clustering or grouping based on geographical origin of the isolates. Nonetheless, *PcDBP2* region II was observed to be distributed into two different groups – group I, designated as Strain B type, and the group II designated as Berok type. Similar bifurcation has been reported in *P. cynomolgi* genes such as merozoite surface protein 1 (MSP1) (Sawai et al., 2010), MSP3 (Rice et al., 2014), and circumsporozoite protein (Pacheco et al., 2012), as well as Peninsular Malaysia *P. knowlesi* PKDBP4 region II (Fong et al., 2014). The haplotype analysis highlights remarkable variation of *PcDBP2* region II. From the 152 *PcDBP2* region II sequences, 146 haplotypes were identified with one predominant haplotype (H27) with frequency of 7.

Interestingly, we noticed one macaque blood sample (KTLT-012) had mixed *P. cynomolgi* haplotype group infection. Two of the *PcDBP2* region II sequences of this sample, KTLT012-C2 and C3, were clustered into the group II, but one sequence (KTLT012-C1) was positioned between the group I and II (Figure 1). This unique KTLT012-C1 is a recombinant type that displayed recombination between *PcDBP2* region II of Strain B and Berok allelic types. Such intragenic recombination between strains has been observed in *P. vivax* MSP1 from Myanmar (Naw et al., 2021). The possible cause of such recombination is meiotic genetic exchange or intrahelical strand slippage during DNA replication (Naw et al., 2021).

Since wild macaques mostly harbour polyclonal infections, it is not surprising that they can also be infected with multiple strains of the *Plasmodium* species (Zhang et al., 2016). There is a possibility that genetic recombination between Strain B and Berok types may increase in the future as what has been seen in *PvMSP1* from Myanmar. Thus, continuous monitoring of the genetic diversity is necessary to elucidate the polymorphic nature and gene flow in the *P. cynomolgi* populations in Malaysia.

**CONCLUSION**

This study is the first to investigate genetic diversity, natural selection, and haplotypes of *PcDBP2* region II from Peninsular Malaysia. High level of genetic diversity was observed in *PcDBP2* region II and is under purifying (negative) selection. The separation of *PcDBP2* region II into two main groups may suggest evidence of two distinct *P. cynomolgi* types or lineages. Future study should look into the diversity, natural selection, and groupings of *P. cynomolgi* erythrocyte binding proteins such as *PcDBP1* and reticulocyte binding proteins (PcRBPs).

**ACKNOWLEDGEMENTS**

This study was funded by the Fundamental Research Grant Scheme (FRGS) of the Ministry of Higher Education Malaysia (Grant Number: FRGS/1/2020/ SKKO/UM/01/3), which is awarded to MYF. SS and AA were supported by the Ministry of Education Malaysia Long Term Research Grant Scheme (LRGS/1/2018/ UM/O1/1/2). The authors thank the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN) for approval of the permit to collect wildlife samples for this study (Permit number: W-00256-15-19).

**Conflict of interest**

The authors declare no conflict of interest.

**REFERENCES**


SUPPLEMENTARY DATA