GENETIC DIVERSITY OF *PLASMODIUM KNOWLESI* RED CELL INVASION GENES (*PKNBPXA* AND *PKNBPXB*) AND THEIR ASSOCIATION WITH PARASITAEMIA

Mohammed Atique Ahmed

Doctor of Philosophy
(Molecular Parasitology)
2013
This thesis is dedicated to the memory of my late beloved father Late Naquib Ahmed (May Allah grant him eternal rest, Aameen), to my affectionate mother, Selima Ahmed who has always been a constant support and encouragement and my two brothers (Nazib Ahmed and Jahid Ahmed).
ACKNOWLEDGEMENTS

First and foremost I am thankful to the Almighty for giving me the strength and will power and patience to complete this work.

I would like to thank my supervisor Prof Janet Cox-Singh, St Andrews University, Scotland and Honorary researcher, Malaria Research Centre (MRC), Universiti Malaysia Sarawak (UNIMAS) who gave me an opportunity to work on a very interesting parasite. I thank her for constant guidance, support and enthusiasm throughout my PhD training. I would also like to thank my co-supervisors; Prof Balbir Singh, Director MRC, UNIMAS and Prof Sanjeev Krishna, Centre for Infection and Immunity, St Georges University of London, who had guided me during this PhD training. I would like to thank Dr Ramlah Zainuddin of the Faculty of Resource Science and Technology, UNIMAS for her valuable advice and training on phylogenetic analyses. I owe a special thanks to Tan Sri Datu Prof Dr Mohamad Taha Arif, former Dean of Faculty of Medicine and Health Sciences, UNIMAS who had been always very kind and friendly and made me feel at home during my stay in Malaysia. To the following people, whom I owe special gratitude; Prof David J. Conway of the London School of Tropical Medicine and Hygiene, Associate Prof Manoj Duraisingh of the Havard School of Public Health and Dr Julian Rayner of the Wellcome Trust Sanger Institute for their valuable advice. I owe a special thank you and grateful to one of my closest 'bm' who has been a constant company throughout my PhD training.

My deepest appreciation goes to the Hospital Staff of Sibu and Sarakei who had always helped me during my sample collection days and specially Mr Lu Chan Woon and Mr Wong of Sarakei Hospital who had been a great help and company.

I would like to thank my mom and my brothers and all my relatives who had always been with me through their love, support and encouragement. I would also like to extend my appreciation to my dearest lab mates: Angela Siner, Harttini Hatta, Suzie Voon, Kueh Bing Ling, Paul Simon Divis, Sang Nee, Dayang Shuaisyah Binti Awang Mohamad and Khamisa Abdul Kadir. Thank you all for helping me and sharing your knowledge and also being a great company during my stay in Malaysia.

I would like to thank all the staff members of the Faculty of Medicine and Health Sciences, UNIMAS who directly or indirectly contributed to the success of this work. Thank you for being so friendly.

This study was financially supported by Medical Research Council Grant, United Kingdom (MRC,UK).
ABSTRACT

Human infections with *Plasmodium knowlesi*, a parasite of long-tailed (*Macaca fascicularis*) and pig-tailed macaques (*Macaca nemestrina*), continue to be reported in most countries within Southeast Asia and majority of the humans cases occur in Malaysian Borneo. The parasite has a 24 hour erythrocytic cycle and parasitaemia has been shown to be associated with the development of severe malaria in humans. The parasite invades host red blood cells via specific receptors on these cells using multiple parasite invasion ligands and the reticulocyte binding-like (RBL) proteins are one such family of ligands. There are two RBL proteins in *P. knowlesi*; normocyte binding protein xa and xb (*Pknbpxa* and *Pknbpxb*) which are expressed at merozoite apex during invasion process. The polymorphisms at these RBL genes may be critical to the increase in peripheral parasitaemia in humans, thereby giving an invasion advantage to the parasite. Therefore, this study was designed to test the hypothesis that parasitaemia is associated with particular alleles or haplotypes of *Pknbpxa* and *Pknbpxb* genes from clinically well characterised patient isolates recruited from two study sites; Sibu and Sarakei divisions of Sarawak. Genetic diversity and signatures of natural selection acting on these loci from patient isolates were also determined. In the first instance, coding regions with highest diversity within *Pknbpxa* (8.5 kb) and *Pknbpxb* (3.1 kb) fragments from five *P. knowlesi* reference isolates recruited from geographically distinct regions of Sarawak were identified through high stringency amplification, cloning and sequencing methods. Sequence analyses of *Pknbpxa* and *Pknbpxb* from five *P. knowlesi* reference isolates revealed that the overall nucleotide diversity of *Pknbpxa* gene was high (\( \pi = 0.01429 \)). Lower diversity was observed for *Pknbpxb* gene (\( \pi = 0.00381 \)). Regions of highest diversity were localised towards the 5' erythrocyte binding region of *Pknbpxa* gene.
(nucleotide position 389 to 1388, \( \pi = 0.0241 \)). Highest diversity within the \( Pknbp\) reference sequences were observed from nucleotide position 2157 to 3156 \((\pi = 0.00582)\).

Based on these highly polymorphic regions, haplotypes were identified by sequencing \((Pknbp\)a 885 bp and \( Pknbp\)b 897 bp) fragment from each gene which identified 75 \( Pknbp\)a haplotypes with high haplotype diversity \((Hd = 0.9729)\) and 51 \( Pknbp\)b haplotypes with haplotype diversity \((Hd = 0.922)\) from 138 and 134 \( P.\) knowlesi patient isolates respectively. Phylogenetic analyses of \( Pknbp\)a fragments from the five \( P.\) knowlesi reference isolates as well as the haplotyping fragments from 138 patient isolates revealed that the gene was dimorphic (KH195 group; \( n = 77 \) and KH273 group; \( n = 61 \)).

Deduced amino acid sequences obtained from the five reference isolates for both genes showed that all cysteine residues were conserved thus indicating intact binding function to host erythrocyte. Similar conservation of cysteine residues were observed for all the 138 sequences obtained from the \( Pknbp\)a haplotyping region. Analyses of natural selection using various tests revealed that the \( Pknbp\)a gene is under strong positive selection while the \( Pknbp\)b gene might be under negative selection within the parasite population. Natural selection with isolates within each dimorphic groups of \( Pknbp\)a gene revealed that there might be an ongoing selective sweep towards KH273 group due to strong positive selection. Genetic association analyses of non-synonymous SNPs of \( Pknbp\)a and \( Pknbp\)b with parasitaemia showed that there were significant associations between two \( Pknbp\)a SNPs C913T (minor allele 'C' representing 18.8\% of isolates) \((r = 0.2100, p = 0.0149)\), G1102C (minor allele 'G' representing in 30.4\% of isolates) \((r = 0.171, p = 0.047)\) and one \( Pknbp\)b SNP A3115G (major allele 'G' representing in 59\% of isolates) \((r = 0.1809, p = 0.0394)\). These results suggest that different variants of \( Pknbp\)a and \( Pknbp\)b are involved in increasing erythrocyte invasion efficiency resulting in increased parasitaemia in
patients. Further studies are necessary to determine whether more SNPs of these genes associate with patient parasitaemia.
ABSTRAK

Laporan jangkitan Plasmodium knowlesi ke atas manusia, parasit kera berekor panjang (Macaca fascicularis) dan berekor seakan ekor babi (Macaca nemestrina), terus dilaporkan di kebanyakan negara di Asia Tenggara dan majoriti kes-kes malaria manusia disebabkan P. knowlesi dilaporkan dari Borneo, Malaysia. Kitaran eritrosit 24 jam parasit ini dan parasitemia telah ditunjukkan sebagai berhubungkait dengan perkembangan penyakit malaria manusia yang tenat. Parasit menyerang sel-sel darah merah hos menggunakan ligan pencerobohan dan pelbagai reseptor khusus pada eritrosit hos dan protin seperti-mengikat retikulosit (RBL) adalah satu daripada ahli keluarga ligan seperti ini. Terdapat dua protein RBL didalam P. knowlesi; protin mengikat normocyte xa dan xb (Pknbp xa dan Pknbp xb) yang diekspresi di bahagian atas merozoite semasa proses pencerobohan. Polimorfisme yang berlaku di gen RBL mungkin kritikal kepada peningkatan parasitemia periferal manusia dan dengan itu memberi kelebihan untuk parasit melakukan pencerobohan. Oleh itu kajian ini telah direka untuk menguji hipotesis bahawa parasitemia berhubungkait dengan alel tertentu atau haplojenis gen Pknbp xa dan Pknbp xb daripada isolat pesakit yang diberikan pencirian klinikal tepat yang direkrut dari dua kawasan kajian; Sibu dan Sarakei. Kepelbagaian genetik dan tanda-tanda pilihan secara semula jadi yang bertindak pada lokus ini didalam isolat pesakit juga telah ditentukan. Dalam contoh pertama, kepelbagaian tertinggi bahagian yang dikodkan dari pecahan dalam Pknbp xa (8.5 kb) dan Pknbp xb (3.1 kb) dari lima isolat pesakit rujukan P. knowlesi yang direkrut dari daerah daripada kawasan geografi yang berlainan di Sarawak telah dikenal pasti melalui pengklonan berteliti tinggi dan kaedah penjujukan (“sequencing”). Analisa jujukan Pknbp xa dan Pknbp xb dari lima isolat rujukan P. knowlesi mendedahkan bahawa kepelbagaian keseluruhan nukleotida gen Pknbp xa adalah tinggi (π = 0.01429). Kepelbagaian nukleotida yang lebih rendah diperhatikan pada gen Pknbp xb (π =
Bahagian-bahagian berkepelbagaian tertinggi berada ke arah 5' kawasan mengikat eritrosit gen \( Pknbpxa \) (kedudukan nukleotida 389-1388, \( \pi = 0.0241 \)). Kepelbagaian tertinggi dalam jujukan rujukan \( Pknbpxb \) diperhatikan berlaku dari kedudukan nukleotida 2157 sehingga ke 3156 (\( \pi = 0.00582 \)). Berdasarkan bahagian-bahagian yang berpolimorfik tinggi, haplojenis-haplojenis dikenal pasti melalui jujukan pecahan 0.8 kb daripada setiap gen mengenal pasti 75 haplotaiip \( Pknbpxa \) dengan kepelbagaian haplojenis tinggi (Hd = 0.9729) dan 51 haplojenis \( Pknbpxb \) dengan kepelbagaian haplojenis tinggi (Hd = 0.922) masing-masing dari 138 dan 134 isolat pesakit \( P. knowlesi \). Analisis filogenetik serpihan \( Pknbpxa \) daripada lima isolat rujukan \( P. Knowlesi \) serta serpihan penghaplojenis daripada 138 isolat pesakit mendedahkan bahawa gen adalah dwi-morfik (kumpulan KH195; n = 77 dan kumpulan KH273; n = 61). Kesimpulan jujukan asid amino yang diperolehi daripada lima isolat rujukan bagi kedua-dua gen menunjukkan pemuliharaan semua sisteina menunjukkan fungsi mengikat hos eritrosit kekal sempurna. Pemuliharaan sisteina juga diperhatikan untuk kesemua 138 jujukan yang diperolehi dari bahagian penghaplojenis \( Pknbpxa \). Analisa pilihan semula jadi dalam populasi parasit menggunakan pelbagai ujian mendedahkan bahawa gen \( Pknbpxa \) adalah di bawah pemilihan positif yang kuat manakala gen \( Pknbpxb \) mungkin di bawah pemilihan negatif. Pilihan semula jadi isolat dalam setiap kumpulan dwi-morfik gen \( Pknbpxa \) mendedahkan kemungkinan terdapat pergerakan terpilih yang berterusan ke arah kumpulan KH273 kesan pemilihan positif yang kuat. Analisa penyatuan genetik bukan sinonim SNPs \( Pknbpxa \) dan \( Pknbpxb \) dengan parasitemia menunjukkan terdapat hubungan yang penting diantara dua SNPs \( Pknbpxa \) C913T (alel kecil 'C' mewakili 18.8% daripada isolat) \( (r = 0.2100, p = 0.0149) \), G1102C (alel kecil 'G' mewakili dalam 30.4% daripada isolat) \( (r = 0.171, p = 0.047) \) dan satu SNP \( Pknbpxb \) A3115G (alel utama 'G' mewakili dalam 59% daripada isolat) \( (r = 0.1809, p = 0.0394) \). Keputusan ini menunjukkan bahawa varian \( Pknbpxa \)
dan *Pknhx* yang berlainan terlibat dalam meningkatkan kecekapan eritrosit menyebabkan peningkatan parasitemia pesakit. Kajian lanjut adalah perlu untuk menentukan sama ada lebih banyak SNP daripada gen-gen ini berhubungkait dengan parasitemia pesakit.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xvi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xviii</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>xxii</td>
</tr>
</tbody>
</table>

## Chapter One: General introduction and review of the literature

1.1 Malaria: Global health burden                                        | 1    |
1.2 The life cycle of the malaria parasites                              | 2    |
1.3 Clinical symptoms of malaria                                         | 5    |
1.4 *Plasmodium knowlesi*: Simian malaria parasite                       | 6    |
1.5 Merozoite structure                                                  | 10   |
1.6 Merozoite invasion of host erythrocytes                              | 12   |
1.7 Invasion genes of *Plasmodium* in the apical cortex of merozoite     | 15   |
   1.7.1 The Duffy Binding-Like (DBP) protein family                      | 16   |
   1.7.2 Reticulocyte Binding-Like (RBP) protein family                  | 18   |
   1.7.3 RBLs in *Plasmodium knowlesi*                                    | 21   |
1.8 Dimorphic invasion genes                                             | 26   |
1.9 Polymorphisms in DBPs and RBPs and natural selection                | 28   |
1.10 Genetic association studies between polymorphic invasion genes of *Plasmodium* and clinical outcome | 31   |
1.11 Objective of present study                                          | 33   |

## Chapter two: General materials and methods

2.1 Laboratory methods                                                   | 35   |
   2.1.1 Patient recruitment process                                      | 35   |
      2.1.1.1 Blood sample collection from the field                      | 35   |
      2.1.1.2 Storage of blood samples in field sites and transportation to UNIMAS | 36   |
      2.1.1.3 Collection of clinical history of patients                 | 36   |
   2.1.2 Thick and thin blood smears                                     | 36   |
   2.1.3 Parasite counting                                               | 36   |
   2.1.4 Extraction of DNA from blood spots using InstaGene™ Matrix     | 37   |
   2.1.5 DNA extraction from blood using QIAamp Blood Mini kit          | 38   |
   2.1.6 Screening of single *Plasmodium knowlesi* infection by nested PCR | 39   |
      2.1.6.1 First nest amplification reaction (N-1)                     | 39   |
2.1.6.2 Second nest amplification reaction (N-2)  
2.1.7 Gel electrophoresis and analysis of nested PCR products  
2.1.8 Selection of five reference isolates for full-length amplification of Pknbpxa and Pknbpxb genes  
2.1.9 Primer designing for PCR amplification and sequencing of Pknbpxa and Pknbpxb genes  
2.1.10 Gel excision and gel purification of PCR amplified products  
2.1.10.1 Gel excision  
2.1.10.2 DNA recovery from gel slice  
2.1.11 Cloning and transformation of PCR products  
2.1.12 Preparation of glycerol stocks  
2.1.13 Plasmid DNA extraction  
2.1.14 Confirmation of inserts by restriction digestion  
2.1.15 Walk-in sequencing  
2.1.16 Generating high stringency reference sequences  
2.1.17 PCR amplicon purification for direct PCR sequencing  
2.1.18 PCR amplification of Pknbpxa gene with primers PknbpxaF5 and PknbpxaR2  
2.1.19 Confirmatory PCR amplification of Pknbpxa gene using primers PknbpxaF11 and PknbpxaR4  
2.1.20 Colony PCR amplification for identification of positive clones of Pknbpxa inserts  
2.1.21 PCR amplification of Pknbpxb gene using primer XB273F and XB343OR  
2.1.22 PCR amplification of the haplotyping region of Pknbpxa gene with Phusion® High Fidelity DNA Polymerase  
2.1.23 PCR amplification of the haplotyping region of Pknbpxa gene with Elongase® Enzyme mix  
2.1.24 PCR amplification of the haplotyping region of Pknbpxb gene with Phusion® High Fidelity DNA Polymerase  
2.1.25 PCR amplification of the haplotyping region of Pknbpxb gene with Elongase® Enzyme mix  
2.1.26 Ethanol-Sodium acetate precipitation  
2.2 Molecular Analysys  

Chapter three: Molecular characterisation of Plasmodium knowlesi normocyte binding protein xa (Pknbpxa) gene  
3.1 Introduction  
3.2 Materials and methods  
3.2.1 Primer design for PCR amplification of P. knowlesi normocyte binding protein xa gene from human isolates  
3.2.2 Standardising PCR amplification of the Pknbpxa gene using QIAGEN® Long Range PCR kit  
3.2.3 Cloning and high stringency sequencing of Pknbpxa gene  
3.2.3.1 Gel electrophoresis of the PCR amplified products and cloning  
3.2.3.2 Gel purification of amplified PCR products  
3.2.3.3 Cloning and transformation of PCR products  
3.2.3.4 Colony PCR standardisation  
3.2.3.5 pCR® XL TOPO plasmid extraction from positive colonies
3.2.3.6 Confirmation of inserts by EcoRI  
3.2.3.7 Walk-in sequencing and generating high stringency reference sequences  
3.2.4 DNA sequence analyses of five reference isolates  
3.3 Results  
3.3.1 Standardisation of PCR amplification of the Pknbpxa fragment (8501 bp)  
3.3.2 Cloning Pknbpxa and confirmation  
3.3.2.2 Walk-in sequencing and generation of high stringency reference sequences  
3.3.3 Sequence analyses and characterisation of 8501 bp Pknbpxa fragment from five reference isolates  
3.3.4 Tests for natural selection  
3.4 Discussion  

Chapter four: Molecular characterisation of Plasmodium knowlesi normocyte binding protein xb (Pknbp xb) gene  
4.1 Introduction  
4.2 Materials and methods  
4.2.1 Primer designing for PCR amplification of Pknbp xb gene from P. knowlesi infected human isolates.  
4.2.2 Standardising PCR amplification of the Pknbp xb gene using Elongase® enzyme mix  
4.2.3 Cloning and high stringency sequencing of Pknbp xb gene  
4.2.3.1 Confirmation of inserts by Hind III digestion  
4.2.3.2 Walk-in sequencing and generating high stringency reference sequences  
4.2.4 DNA sequence analyses of five reference isolates  
4.3 Results  
4.3.1 Standardisation of PCR amplification of the Pknbp xb fragment (3505 bp)  
4.3.2 PCR amplification, cloning and sequencing of Pknbp xb fragment (3506 bp) from reference isolates  
4.3.2.1 Cloning Pknbp xb and confirmation  
4.3.2.2 Walk-in sequencing and generation of high stringency reference sequences  
4.3.3 Sequence analyses and characterisation of 3100 bp Pknbp xb fragment from five reference isolates  
4.3.4 Tests for natural selection  
4.4 Discussion  

Chapter five: Haplotyping polymorphic fragments of Pknbp xa and Pknbp xb in Plasmodium knowlesi isolates from patients in Sarawak.  
5.1 Introduction  
5.2 Materials and methods  
5.2.1 DNA extraction from study samples  
5.2.1 Primer designing and standardisation of PCR amplification of haplotyping  

xii
5.2.1 Sequence alignment of \textit{Pknbp\textsubscript{x}a} and \textit{Pknbp\textsubscript{x}b} sequences and phylogenetic analysis

5.2.2 Sequence diversity within \textit{Pknbp\textsubscript{x}a} and \textit{Pknbp\textsubscript{x}b} aligned sequences

5.2.3 Haplotypes and haplotype diversity within \textit{Pknbp\textsubscript{x}a} and \textit{Pknbp\textsubscript{x}b} fragments sequenced

5.2.4 Polymorphisms within \textit{Pknbp\textsubscript{x}a} and \textit{Pknbp\textsubscript{x}b} sequences

5.2.5 Linkage Disequilibrium

5.2.6 Determination of recombination events and recombination break-point analysis within \textit{Pknbp\textsubscript{x}a} and \textit{Pknbp\textsubscript{x}b} sequence alignments

5.2.7 Minimum spanning tree (MST) based haplotype network

5.2.8 Signatures of natural selection within \textit{Pknbp\textsubscript{x}a} and \textit{Pknbp\textsubscript{x}b} sequences

5.2.8.1 \(dN/dS\) ratio

5.2.8.2 Test of neutrality using Tajima's D, Li and Fu's D* and F* values

5.2.8.3 McDonald and Kreitman (MK) test for neutrality

5.3 Results

5.3.1 DNA extraction and parasitaemia

5.3.2 Haplotyping \textit{Pknbp\textsubscript{x}a} from \textit{P. knowlesi} field isolates

5.3.3 Alignment and phylogenetic analysis of \textit{Pknbp\textsubscript{x}a} sequences

5.3.4 Sequence diversity within the \textit{Pknbp\textsubscript{x}a} fragment

5.3.5 \textit{Pknbp\textsubscript{x}a} Haplotypes and haplotype diversity

5.3.6 \textit{Pknbp\textsubscript{x}a} SNPs

5.3.7 Determination of recombination events and recombination break-point

5.3.8 Linkage disequilibrium within 138 \textit{Pknbp\textsubscript{x}a} sequences

5.3.9 Minimum spanning tree based haplotype network generated from 885 bp of \textit{Pknbp\textsubscript{x}a} gene from 138 field isolates

5.3.10 Signatures of natural selection within \textit{Pknbp\textsubscript{x}a} fragment

5.3.10.1 \(dN/dS\) ratio for \textit{Pknbp\textsubscript{x}a}

5.3.10.2 Codon based z-test for testing natural selection in each \textit{Pknbp\textsubscript{x}a} dimorphic groups.

5.3.10.3 Test of neutrality using Tajima's D, Li and Fu's D* and F* values

5.3.10.4 McDonald and Kreitman (MK) test for \textit{Pknbp\textsubscript{x}a} sequences

5.4.1 Haplotyping \textit{Pknbp\textsubscript{x}b} from \textit{P. knowlesi} field isolates

5.4.2 Alignment and phylogenetic analysis of \textit{Pknbp\textsubscript{x}b} sequences

5.4.3 Sequence diversity within 134 \textit{Pknbp\textsubscript{x}b} sequences

5.4.4 \textit{Pknbp\textsubscript{x}b} Haplotypes and haplotype diversity

5.4.5 \textit{Pknbp\textsubscript{x}b} SNPs

5.4.6 Recombination within \textit{Pknbp\textsubscript{x}b}

5.4.7 Linkage disequilibrium

5.4.8 Minimum spanning tree based haplotype network with \textit{Pknbp\textsubscript{x}b} haplotypes

5.4.9 Signatures of natural selection within \textit{Pknbp\textsubscript{x}b} fragment

5.4.9.1 \(dN/dS\) ratio

5.4.9.2 Codon based z-test for selection within \textit{Pknbp\textsubscript{x}b} sequences

5.4.9.3 Test of neutrality using Tajima's D, Li and Fu's D* and F* values

5.4.9.4 McDonald and Kreitman (MK) test
Chapter six: Genetic association studies of Pknbp xa and Pknbp xb alleles with parasitaemia, summary of findings and future work

6.1 Introduction
6.2 Materials and methods
6.3 Results and Discussion
6.4 Summary of findings
6.5 Future work

Chapter seven: General discussion and Conclusion

REFERENCES

APPENDICES
Appendix A Flowchart for malaria specimen collection
Appendix B Patient information sheet
Appendix C Page 1, Nested 1 PCR amplification for screening
Appendix C Page 2, Nested 2 PCR amplification
Appendix D Preparation of sequencing reaction for Plasmid DNA using Big Dye®
Terminator v3.1 Cycle sequencing kit
Appendix E Preparation of sequencing reaction for PCR products using Big Dye®
Terminator v3.1 cycle sequencing kit
Appendix F Flowchart for PCR amplification, cloning and sequencing of Pknbp xa and Pknbp xb fragments from five reference isolates
Appendix G EcoR I restriction sites of Pknbp xa insert and pCR XL TOPO vector
Appendix H Internal primers used for sequencing of Pknbp xa gene, 8501 bp
Appendix I Amino acid sequence Alignment of Pknbp xa gene starting at Exon II
Appendix I Nucleotide sequence alignment of Pknbp xa gene starting at Exon II
Appendix J Amino acid map of Exon II of Pknbp xa gene obtained from five reference isolates and the polymorphisms within them
Appendix K Hind III restriction sites of Pknbp xb insert and pCR XL TOPO vector
Appendix L Internal primers used for sequencing Pknbp xb gene
Appendix M Figure showing 3 nucleotide deletions within the five P. knowlesi reference isolates and conserved amino acid motif (ENL) with reference to published P. knowlesi H-strain Pknbp xb sequence EU867792 and ACJ54536
Appendix N Nucleotide sequence sequence alignment of Pknbp xb fragment (3100 bp) starting at Exon II
Appendix N Amino acid sequence alignment of Pknbp xb fragment from five reference isolates starting at Exon II
Appendix O Amino acid map of Exon II of Pknbp xb gene from five reference isolates and the polymorphisms within them
Appendix P Parasitemia of P. knowlesi samples per microliter
Appendix Q Summary of haplotyping Pknbp xa and Pknbp xb fragments
Appendix R Seventy five Pknbp xa haplotypes, frequency and the corresponding
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix S</td>
<td><em>Pknbpxa</em> haplotypes and nucleotide positions</td>
<td>249</td>
</tr>
<tr>
<td>Appendix S</td>
<td><em>Pknbpxb</em> haplotypes and nucleotide positions</td>
<td>251</td>
</tr>
<tr>
<td>Appendix T</td>
<td>Fifty one <em>Pknbpxb</em> haplotypes, frequency and the corresponding isolates</td>
<td>252</td>
</tr>
<tr>
<td>Appendix U</td>
<td>Minor and major alleles of <em>Pknbpxa</em> and <em>Pknbpxb</em></td>
<td>254</td>
</tr>
<tr>
<td>Appendix V</td>
<td>Nucleotide composition of <em>Pknbpxa</em> haplotyping region (885 bp)</td>
<td>255</td>
</tr>
<tr>
<td>Appendix W</td>
<td>Nucleotide composition of <em>Pknbpxb</em> haplotyping region (879 bp)</td>
<td>257</td>
</tr>
<tr>
<td>Appendix X</td>
<td>Evolutionary distances within <em>Pknbpxa</em> 138 haplotyping fragments</td>
<td>259</td>
</tr>
<tr>
<td></td>
<td>computed using the Kimura 2-method</td>
<td></td>
</tr>
<tr>
<td>Appendix Y</td>
<td>Evolutionary distances within <em>Pknbpxb</em> 134 haplotyping fragments</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>computed using the Kimura 2-method</td>
<td></td>
</tr>
<tr>
<td>Appendix Z</td>
<td>Parasite purification from archived frozen blood samples</td>
<td>281</td>
</tr>
<tr>
<td>Appendix A1</td>
<td>Ethical clearance</td>
<td>282</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.1  Human malaria parasite ligand and erythrocyte receptors interactions identified during merozoite invasion.

Table 2.1  PCR Primer Sequences used for screening.

Table 2.2  Reference isolates used for generating full-length (8.5 kb) Pknbpxa and (3.5 kb) Pknbpxb gene sequences.

Table 3.1  Forward and reverse primers designed to amplify the Pknbpxa gene from P. knowlesi field isolates.

Table 3.2  EcoRI restriction digestion sites of 8501 bp Pknbpxa (EU867791) fragment.

Table 3.3  Sequence diversity (α) of Pknbpxa gene (8501 bp) within five reference P. knowlesi isolates.

Table 3.4  Nucleotide composition of five reference isolates for the Pknbpxa gene.

Table 3.5  McDonald-Kreitman Test for neutrality; Pknbpxa (8501 bp) reference isolates (n = 5) and orthologous P. cynomolgi RBP3 (n = 3).

Table 4.1  Sequence read-lengths obtained for Pknbpxb gene from five reference isolates.

Table 4.2  Nucleotide composition of five reference isolates for the Pknbpxb gene.

Table 4.3  Sequence diversity of Pknbpxb (3100 bp) gene fragment within the five reference isolates starting at Exon II.

Table 4.4  McDonald-Kreitman Test for neutrality; Pknbpxb (3100 bp) reference isolates (n = 5) and orthologous P. cynomolgi RBP2e (n = 1).

Table 5.1  Pknbpxa nucleotide diversity, synonymous(S) and non-synonymous (NS) sites.

Table 5.2  Polymorphisms within 138 Pknbpxa sequences.

Table 5.3  Test for recombination of 138 Pknbpxa sequences using DnaSP v 5.10.

Table 5.4  Recombination detection within 75 Pknbpxa sequences and determination of recombination break point using HYK85 model in HyPhy package.
Table 5.5  Positions of 29 core nucleotides of the \textit{Pknbpxa} gene within the 135 dimorphic groups KH273 and KH195 of 138 \textit{P. knowlesi} sequences.

Table 5.6  Natural selection within \textit{Pknbpxa} sequences.
Table 5.7  McDonald-Kreitman test for neutrality; \textit{Pknbpxa} (885 bp) sequences (n = 138) and \textit{P. cynomolgi RBP3} (n = 3).
Table 5.8  Nucleotide diversity, synonymous (S) and non-synonymous (NS) mutations within 134 \textit{Pknbpxb} fragments in the haplotyping region.
Table 5.9  Polymorphisms within 134 \textit{Pknbpxb} sequences.
Table 5.10 Test for recombination of 134 \textit{Pknbpxb} sequences using DnaSP v 5.10.
Table 5.11 Recombination detection within \textit{Pknbpxb} sequences and determination of recombination break point using HYK85 model in HyPhy package.
Table 5.12 Neutrality tests and values of dN-dS and dN/dS within 134 \textit{Pknbpxb} sequences.
Table 5.13 Selection Analyses Using McDonald-Kreitman test for 134 \textit{Pknbpxb} sequences and \textit{P. cynomolgi RBP2E}.
Table 6.1  Association between \textit{Pknbpxa} allele and parasitaemia
LIST OF FIGURES

Figure 1.1 The life cycle of malaria parasite. 3
Figure 1.2 A map of Southeast Asia outlining the range of the natural hosts and mosquito vectors of Plasmodium knowlesi. 7
Figure 1.3 Merozoite structure with major organelles being highlighted. 11
Figure 1.4 Schematic depicting the early stages of red blood cell (RBC) invasion by the malaria merozoite, and the putative roles of the various protein types. 14
Figure 1.5 The DBL and RBL protein families of P. falciparum. 22
Figure 1.6 Schematic representation of Plasmodium knowlesi normocyte binding protein xa and xb. 24
Figure 1.7 Immuno-electron microscopic localization of the PkRBL proteins Pknbp xa and Pknbp xb within the micronemes. 25
Figure 3.1 Schematic representation of position of primer pairs on the Pknbp xa gene and amplified regions using combinations Pknbp xaF5 with Pknbp xaR2 and for confirmatory PCR with primers Pknbp xaF11 with Pknbp xaR4. 68
Figure 3.2A PCR amplification of Pknbp xa (8501 bp) fragment with primers Pknbp xaF5 and Pknbp xaR2 69
Figure 3.2B Confirmatory PCR amplification of Pknbp xa (8101 bp) with primers Pknbp xaF11 and Pknbp xaR4 70
Figure 3.3 Colony PCR amplification of isolate KH229. 71
Figure 3.4A EcoRI digestion profile of plasmid DNA with cloned Pknbp xa genes (8501bp) of Plasmodium knowlesi from isolate KH229. 74
Figure 3.4B Virtual gel obtained from EcoRI digest of Pknbp xa (8501 bp, EU867791) fragment using NEBcutter v2.0. 74
Figure 3.5A A snap shot of walk-in sequence assembly of Pknbp xa fragment from clone KH195AII with 13 internal forward primers and M13 primer. 75
Figure 3.5B A snap shot of walk-in sequence assembly of Pknbp xa fragment from clone KH195AII with 11 internal reverse primers and M13 primer. 75
Figure 3.6A  *Pknbp xa* gene map with synonymous and non-synonymous mutations within the five *P. knowlesi* reference isolates.

Figure 3.6B Graphical representation of nucleotide diversity (\(\pi\)) of *Pknbp xa* gene (8501 bp fragment) within the five *P. knowlesi* reference isolates.

Figure 3.7A Phylogenetic tree based on the 8501bp fragment of *Pknbp xa* genes from five *Plasmodium knowlesi* reference isolates and RBL genes of other species produced by the Neighbor-Joining method.

Figure 3.7B Phylogenetic tree based on the deduced amino acid sequences of *Pknbp xa* genes obtained from the five *P. knowlesi* reference isolates (marked in red) and RBL amino acid sequences of other species produced by the Neighbor-Joining method.

Figure 3.7C Phylogenetic tree based on the deduced amino acid sequences of *Pknbp xa* genes from the five *P. knowlesi* reference isolates and RBL amino acid sequences of other species produced by the Maximum Likelihood method based on the Poisson correction model.

Figure 4.1 PCR amplification of *Pknbp xa* fragment (3506bp) based on forward primer Xb273F and reverse primer Xb3430R.

Figure 4.2A *Hind III* digestion profile of plasmid DNA with cloned *Pknbp xb* genes (3506 bp) of *P. knowlesi* from isolate KH195. Lane numbers M represents molecular size markers, (1 – 4) represent plasmid DNA samples(A1, A2, B1 and B2) that were derived from different *E. coli* colonies. 5 represents undigested purified PCR amplicon of KH195PCRA and 6 represents blank.

Figure 4.2B Virtual gel obtained from *Hind III* digest of *Pknbp xb* cloned insert and vector sequence using NEBe cutter v2.0.

Figure 4.3A A snap shot of walk-in sequence assembly of *Pknbp xb* fragment from clone KH195A2 with five internal forward primers and M13 primer.

Figure 4.3B A snap shot of walk-in sequence assembly of *Pknbp xb* fragment from clone KH195A2 with five internal reverse primers and M13 primer.

Figure 4.4A Schematic representation of Pknbp xb (EU867792) 9571 bp. (A) Exon 1 (shaded gray) and the intron (solid line) are followed by Exon II beginning at nucleotide 346. A fragment from nucleotide1 to 3448 was amplified.

Figure 4.4B Location of synonymous (S) and non-synonymous (NS) mutations comparing five reference isolates are marked.

Figure 4.5 Graphical representation of nucleotide diversity (\(\pi\)) of *Pknbp xb* gene (3100 bp) within the five *P. knowlesi* reference isolates.
Figure 4.6A  Phylogenetic tree based on the 3100 bp fragment of Pknbpxb genes from five P. knowlesi reference isolates (marked in red) and RBL genes of other species produced by the Neighbor-Joining method.

Figure 4.6B  Phylogenetic tree based on the deduced amino acid sequences of Pknbpxb genes obtained from the five P. knowlesi reference isolates (marked in red) and RBL amino acid sequences of other species produced by the Neighbor-Joining method.

Figure 4.6C  Phylogenetic tree based on the deduced amino acid sequences of Pknbpxb genes obtained from the five P. knowlesi reference isolates (marked in red) and RBL amino acid sequences of other species produced by the Maximum Likelihood method based on the Poisson correction model.

Figure 5.1A  PCR amplification of the Pknbpxa haplotyping region with primers PknbpxaF5 and 7524R1.

Figure 5.1B  PCR amplification of the Pknbpxa haplotyping region with primers Ex1F and 7524R1

Figure 5.2 A  Unrooted phylogenetic tree based on the 885 bp polymorphic fragment of Pknbpxa genes of 138 P. knowlesi isolates produced by the Neighbour Joining method.

Figure 5.2B  Phylogenetic tree based on the 885 bp polymorphic fragment of Pknbpxa genes of 138 P. knowlesi isolates produced by the Neighbour Joining method.

Figure 5.3  Graphical representation of nucleotide diversity (π) of the 885 bp Pknbpxa haplotyping fragment from 138 P. knowlesi isolates and regions with varying diversity marked.

Figure 5.4  Polymorphic nucleotides within the 885 bp of Pknbpxa gene isolates and dimorphisms within them. The 29 core nucleotides distinguishing the dimorphism are highlighted in red within a box. Dots represent identical nucleotides and the position of nucleotides are represented vertically.

Figure 5.5  Deduced amino acid polymorphisms and dimorphism in the Pknbpxa haplotyped fragment. Amino acid dimorphism within the sequences are highlighted within boxes. Dots indicate identical amino acids and the position of amino acids are represented vertically.

Figure 5.6  Amino acid map highlighting the dimorphic residues and the conserved cysteine residues with the 138 Pknbpxa sequences. The amino acid positions are represented vertically.
Figure 5.7 Linkage disequilibrium within the 885 bp *Pknbpxa* haplotyping fragment obtained from 138 *P. knowlesi* sequences.

Figure 5.8 Minimum spanning tree based haplotype network of 75 haplotypes from the 885 bp fragment of the *Pknbpxa* gene from 138 *P. knowlesi* field isolates. The color of the circles represents *Pknbpxa* haplotype group KH273 and KH195 and number within the circles represents frequency of each haplotype which is proportional to its size. Each black bar represents one mutational step.

Figure 5.9A PCR amplification of the *Pknbpxb* haplotyping region with primers Xb273F and Xb3430R using Phusion DNA Polymerase.

Figure 5.9B PCR amplification of the *Pknbpxb* haplotyping region with primers Xb273F and Xb3430R using Elongase enzyme mix.

Figure 5.10 Phylogenetic tree obtained from the 879 bp fragment of *Pknbpxb* gene from 134 *P. knowlesi* isolates using the Neighbor-Joining method.

Figure 5.11 Graphical representation of nucleotide diversity (π) of the 879 bp *Pknbpxb* haplotyping fragment from 134 *P. knowlesi* isolates and regions with highest diversity marked.

Figure 5.12 Nucleotide polymorphisms and dimorphism within *Pknbpxb* isolates. Polymorphic nucleotides within 879 bp of 134 *Pknbpxb* isolates and dimorphisms within them. The 5 core nucleotides distinguishing the dimorphism is highlighted in red.

Figure 5.13 Deduced amino acid polymorphisms in the 879 bp *Pknbpxb* haplotyped fragment. Dots indicate identical amino acids and the position of amino acids are represented vertically.

Figure 5.14 Linkage disequilibrium within the 879 bp haplotyping fragment obtained from 134 *Pknbpxb* isolates.

Figure 5.15 Minimum spanning tree based haplotype network of 51 haplotypes from the 879 bp fragment of the *Pknbpxb* gene from 134 *P. knowlesi* field isolates. The color of the circles represents *Pknbpxb* haplotype group KH195 and KH229 and number within the circles represents frequency of each haplotype which is proportional to its size. Each black bar represents one mutational step.

Figure 6.1 Association of *Pknbpxa* (A) SNP 913 and (B) SNP 1102 with parasitaemia.

Figure 6.2 Association of *Pknbpxb* SNP 3115 with parasitaemia.
ABBREVIATIONS

Pknbpxa  Plasmodium knowlesi normocyte binding protein xa
Pknbpxb  Plasmodium knowlesi normocyte binding protein xb
bp      base pair
DNA     deoxyribonucleic acid
dNTPS   deoxynucleotide triphosphate
EDTA    Ethylenediamine tetraacetic acid
MgCl₂⁺  magnesium chloride
nt      nucleotide
hr      hour
ml      mililitre
PCR     polymerase chain reaction
Pf      Plasmodium falciparum
Pk      Plasmodium knowlesi
Pm      Plasmodium malariae
Po      Plasmodium ovale
Pv      Plasmodium vivax
RBC     red blood cell
rpm     revolution per minute
SSU rRNA small subunit ribosomal ribonucleic acid
TBE     Tris-borate EDTA
TE      Tris-EDTA
WBC     white blood cell
WHO     World Health Organisation
RBL     Reticulocyte binding-like ligand
DBL     Duffy binding-like ligand
CHAPTER ONE

General Introduction and Review of Literature

1.1 Malaria: Global Health Burden

Malaria is one of the most important infectious diseases in the world and is a serious health problem, with approximately half of the world's population at risk. According to the World Malaria Report 2011, there are about 216 million cases of malaria with an estimated 655,000 deaths annually (WHO 2011). At present, about 106 countries are malaria endemic with nearly 50% of them in Sub-Saharan Africa, and children under five years of age and pregnant women are most severely affected (WHO 2011). Moreover, the devastating consequences of malaria are a major obstacle to social and economic development in affected regions (Mendis et al. 2001; Breman et al. 2004; Snow et al. 2005). Considerable efforts have been made to combat malaria since the 1950s, including the global malaria eradication campaign launched by the World Health Organisation (WHO), which aimed to wipe out the disease at a global scale but which later had to be discontinued due to non-feasibility, social, and logistic problems (Litsios 1997). More recently, focus has been given to controlling the spread of malaria, and the World Health Organisation has initiated programmes such as the Roll Back Malaria Partnership Global Strategic Plan 2005 - 2015 with the aim of intensifying and scaling up malaria control interventions (RBM, 2012).

Malaria is a blood infection caused by a protozoan parasite of the phylum Apicomplexa and genus *Plasmodium*. Currently, there about 200 species of *Plasmodium* identified and they are known to infect a wide range of vertebrate hosts including mammals, birds and
reptiles (Garnham 1966; Perkins and Austin 2009), while some 25 species infect primates
(Garnham 1966; Coatney 1971a; Mohapatra et al. 2008). Malaria in humans is caused by
six species of *Plasmodium*; namely, *P. falciparum, P. vivax, P. malariae, P. ovale curtisi, P. ovale wallikeri* and *P. knowlesi* (Mohapatra et al. 2008; Sutherland et al. 2010; Cox-Singh 2012). *Plasmodium* found in monkeys has recently been implicated in a large human outbreak in Kapit, Malaysia (Singh et al. 2004). Also, *P. ovale* has been recently shown to have two sub species, *P. ovale wallikeri* and *P. ovale curtisi* (Sutherland et al. 2010).

1.2 The life cycle of malaria parasites

Briefly, the life cycle of the malaria parasite is characterized by two distinct phases: an asexual phase in the vertebrate host, and a sexual phase that occurs in the mosquito’s midgut which includes a zygote, the only diploid stage in the life cycle of the parasite (Figure 1.1). During a blood meal by an infected *Anopheles* mosquito, sporozoites enter the blood stream of the vertebrate host and within 30 to 45 minutes after inoculation, they reach the liver parenchyma cells where they develop into exo-erythrocytic schizonts. The sporozoites divide by multiple fission to form thousands of invasive merozoites that upon host cell rupture, attach to and enter circulating erythrocytes. In *P. vivax* and *P. ovale* some of the hepatic forms can remain dormant and can delay their schizogony up to several years, being the cause of relapses. The merozoite stage is short-lived and must invade host red blood cells in a rapid manner, within approximately within 45-60 seconds (Johnson et al. 1980). Within the erythrocyte, the merozoite develops through an asexual cycle, from early