GENETIC DIVERSITY AND NATURAL SELECTION OF THE RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) OF RECENT *Plasmodium knowlesi* CLINICAL ISOLATES FROM MALAYSIA

MIRA SYAHFRIENA AMIR RAWA

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

GENETIC DIVERSITY AND NATURAL SELECTION OF THE RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) OF RECENT *Plasmodium knowlesi* CLINICAL ISOLATES FROM MALAYSIA

MIRA SYAHFRIENA AMIR RAWA

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF MEDICAL SCIENCE

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MIRA SYAHFRIENA AMIR RAWA**

Registration/Matric No: **MGN 140005**

Name of Degree: **MASTER OF MEDICAL SCIENCE**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

GENETIC DIVERSITY AND NATURAL SELECTION OF THE RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) OF RECENT *Plasmodium knowlesi* CLINICAL ISOLATES FROM MALAYSIA

Field of Study: **PARASITOLOGY**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

The *Plasmodium* rhoptry-associated protein 1 (RAP-1) plays a role in the formation of the parasitophorous vacuole following the parasite's invasion of red blood cells. Although there is some evidence that the protein is recognized by the host's immune system, a study of *Plasmodium falciparum* RAP-1 suggests that it is not under immune pressure. A previous study on five old *P. knowlesi* strains (1953-1962) suggested that RAP-1 has limited genetic polymorphism and might be under negative selection. In the present study, 30 recent *P. knowlesi* isolates were studied to obtain a better insight into the polymorphism and natural selection of *P. knowlesi* RAP-1. Blood samples from 30 knowlesi malaria patients were used. These samples were collected between 2010 and 2014. The *P. knowlesi* RAP-1 gene, which contains two exons, was amplified by PCR, cloned into *Escherichia coli* and sequenced. Genetic diversity and phylogenetic analyses were performed using MEGA6 and DnaSP ver. 5.10.00 programs. Thirty *P. knowlesi* RAP-1 sequences were obtained. The nucleotide diversity (π) of exons 1, 2 and the total coding region (0.00915, 0.01353 and 0.01298, respectively) were higher than those of the old strains. Further analysis revealed a lower rate of non-synonymous (d_N) than synonymous (d_S) mutations, suggesting negative (purifying) selection of *P. knowlesi* RAP-1. Tajima's D test and Fu and Li's D test values were not significant. At the amino acid level, 22 haplotypes were established with haplotype H7 having the highest frequency (7/34, 20.5%). In the phylogenetic analysis, two distinct haplotype groups were observed. The first group contained the majority of the haplotypes, whereas the second had fewer haplotypes. The present study found higher genetic polymorphism in the *P. knowlesi* RAP-1 gene than the polymorphism level reported in the previous study. This observation may stem from the difference in sample size between the present (n=30) and the previous (n=5) study. Synonymous and

nonsynonymous mutation analysis indicated purifying (negative) selection of the gene. The separation of *P. knowlesi* RAP-1 haplotypes into two groups provides further evidence on the postulation of two distinct *P. knowlesi* types or lineages.

ABSTRAK

Rhoptry-associated protein 1 (RAP-1) dalam spesies *Plasmodium* memainkan peranan penting dalam pembentukan vakuol parasitoforus sejurus selepas pencerobohan parasit di dalam sel darah merah. Walaupun terdapat beberapa bukti bahawa protin tersebut dikenal pasti oleh sistem pertahanan hos, kajian mengenai RAP-1 dalam *Plasmodium falciparum* telah menunjukkan bahawa ia tidak berada di bawah tekanan imun. Kajian sebelum ini mengenai lima pencilan lama (1953-1962) *P. knowlesi* mencadangkan RAP-1 mempunyai polimorfisme genetik yang terhad dan mungkin berada di bawah pemilihan negatif. Dalam penyelidikan terbaru ini, 30 pencilan *P. knowlesi* yang baru telah dikaji untuk mendapatkan gambaran yang lebih baik mengenai polimorfisme dan pemilihan semula jadi RAP-1. Sampel darah daripada 30 pesakit malaria knowlesi telah digunakan. Sampel-sampel ini telah diperolehi dari Semenanjung Malaysia di antara tahun 2010 dan 2014. Gen RAP-1, yang mengandungi dua ekson, telah diamplifikasi secara PCR, diklon ke dalam *Escherichia coli* dan dijujukan. Analisis dalam diversiti genetik dan filogenetik telah dilakukan dengan menggunakan program MEGA6 dan DnaSP ver. 5.10.00. Tiga puluh jujukan *P. knowlesi* RAP-1 telah diperolehi. Diversiti nukleotida (π) dalam ekson 1, 2 dan jumlah kawasan pengekodan (0.00915, 0.01353 dan 0.01298, masing-masing) adalah lebih tinggi berbanding dengan diversiti nukleotida dalam pencilan lama yang dilaporkan sebelum ini. Analisis perbandingan mendedahkan kadar mutasi bukan sinonim (d_N) lebih rendah daripada mutasi sinonim (d_S) yang mencadangkan pemilihan negatif dalam RAP-1. Nilai daripada ujian Tajima's D dan ujian Fu and Li's D tidak ketara. Di peringkat asid amino, 22 haplotaip telah dijana, yang mana haplotaip H7 memperoleh frekuensi tertinggi (7/34, 20.5%). Dalam analisis filogenetik, terdapat dua kumpulan haplotaip berbeza. Kumpulan pertama mengandungi majoriti haplotaip, manakala yang kedua mempunyai haplotaip yang

sedikit. Kajian yang baru ini telah mempamerkan polimorfisme genetik yang lebih tinggi dalam *P. knowlesi* RAP-1 gen berbanding dengan laporan sebelumnya. Pemerhatian ini mungkin berpunca daripada perbezaan dalam saiz sampel antara kajian masa kini (n = 30) dan yang sebelumnya (n = 5). Analisis mutasi sinonim dan bukan sinonim menunjukkan pemilihan gen negatif. Pemisahan haplotaip *P. knowlesi* RAP-1 kepada dua kumpulan memberikan bukti selanjutnya kepada postulasi terhadap dua jenis perbezaan atau keturunan *P. knowlesi*.

ACKNOWLEDGEMENTS

It has been an exciting and thrilling journey throughout my experience as a Masters student. I gained the knowledge, skills and experience that have enriched my life. Learning new techniques and meeting a lot of friends with different backgrounds definitely were and always will be the best parts of my time here and I will miss it. And for that, it is my great pleasure to thank my supervisors, Professor Dr. Fong Mun Yik and Associate Professor Dr. Lau Yee Ling for their unconditional guidance and support. Without them, I would not have been who I am today. Their knowledge, research expertise and assistance have helped me a lot in order to complete my research.

The endless support and love from my family will always be my reason, strength and encouragement to make my graduation ceremony possible. Therefore, I thank them for always being there for me at every single point of my life and being the best parents and family members I could have wished for.

Also, my heartiest gratitude goes to my close friends, Tengku Idzzan Nadzirah, Jonathan Liew Wee Kent, Ng Yit Han and Lai Meng Yee for always lending me a helping hand and staying with me through thick and thin. Thank you for making my research dissertation possible and showing me one of the best times of my life. And last but not least, I would like to wish all the best to all my fellow lab mates. I am sure we all have had our great time, both sad and happy, and may it be our good memory to cherish.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

<	less than
>	greater than
%	percent
&	and
°C	degree Celcius
μm	micrometer
μM	micromolarity
μl	microliter
π	nucleotide diversity
g	gram
h	hour
L	litre
m	meter
M	molarity
min	minute
ml	milliliter
s	second
V	volt
rpm	revolutions per minute

ACT	artemisinin-based combination therapy
AIDS	acquired immunodeficiency syndrome
AMA-1	apical membrane antigen 1
ARDS	acute respiratory distress syndrome
BLAST	Basic Local Alignment Search Tool
bp	base pair
$CaCl_2$	calcium chloride
CSP	circumsporozoite surface protein
ddH_2O	distilled and deionized water
d_N	nonsynonymous substitution
d_S	synonymous substitution
DGs	dense granules
DMF	dimethyl formamide
DNA	deoxyribonucleic acid
DBP	Duffy-binding protein
EBA	erythrocyte-binding antigen
EDTA	ethylenediaminetetraacetic acid
GPI	glucose phosphate isomerase
H	number of haplotypes
Hd	haplotype diversity
HCL	hydrochloric acid
HIV	human immunodeficiency virus
IgG	immunoglobulin G
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITNs	insecticide-treated mosquito nets
IRS	indoor residual spraying

kDa	kiloDalton
LB	Luria-Bertani
$lbf/inch^2$	pound-force per square inch
LLINs	long-lasting insecticidal nets
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
MHC	major histocompatibility complex
MK	McDonald-Kreitman
ML	maximum likelihood
MP	maximum parsimony
MSP-1	merozoite surface protein 1
n	number of isolates
NaCl	sodium chloride
NaOH	sodium hydroxide
NGS	next-generation sequencing
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycerol
PFS	*Plasmodium falciparum* surface protein
PVM	parasitophorous vacuole membrane
RALP	rhoptry-associated, leucine zipper-like protein
RAMA	rhoptry-associated membrane antigen
RAP-1	rhoptry-associated protein 1
RAP-2	rhoptry-associated protein 2
RAP-3	rhoptry-associated protein 3

Rh	reticulocyte-binding-like homologue protein
Rhop	rhoptry protein
RNA	ribonucleic acid
RON-3	rhoptry neck protein 3
RON-6	rhoptry neck protein 6
ROPE	repetitive organellar protein
S	segregating sites
SD	standard deviation
SE	standard error
SNP	single nucleotide polymorphism
SSP-2	sporozoite surface protein 2
TAE	tris-acetate-EDTA
UV	ultraviolet

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Plasmodium knowlesi infection is a newly described zoonosis that causes morbidity and mortality. It is one of the six malaria species that can infect humans (Cox-Singh *et al*., 2008; White, 2008). Infection by *P. knowlesi* is common in Malaysia, where majority of the cases are reported in Sabah and Sarawak. It has also been reported in other Southeast Asian countries including Myanmar (Jiang *et al*., 2010), Thailand (Sermwittayawong *et al*., 2012), Cambodia (Khim *et al*., 2011) and Singapore (Ng *et al*., 2008). In 2013, 68% of malaria patients hospitalized in Malaysian Borneo were positive with this zoonotic malaria (Piotrowski, 2014). The increasing number of knowlesi malaria cases indeed highlights the importance of detailed studies on *P. knowlesi* in developing successful intervention strategies to control the infection in Malaysia.

The pathogenesis of malaria parasites involves the orchestrated action of various proteins, a few of which are primary targets for anti-malarial vaccines. These proteins frequently exhibit high levels of heterozygosity, and rapid rates of evolution may be essential for the parasite to escape the host's immune defence (Hughes & Hughes, 1995). Highly polymorphic proteins are often favoured by positive selection, in which selective forces, such as immune responses and drugs, drive genes expressing these antigenic proteins to accumulate mutations and maintain them in the population (Escalante *et al*., 2004). This strategy enables the parasite to manifest antigenically different alleles to thwart the host's immune response. Alternatively, these alleles may be eliminated or negatively selected in the case of less fit genetic variants.

The *Plasmodium* merozoite invasion of red blood cells involves binding, apical orientation and secretion of apical organelle contents known as rhoptries, micronemes and dense granules (Baum *et al.*, 2008; Counihan *et al.*, 2013; Kats *et al*., 2008). This very crucial stage is responsible for most of the clinical manifestations of malaria. Proteins in these organelles have been implicated in key aspects of invasion. These include the formation of moving junctions between the merozoite and erythrocyte surfaces, which subsequently leads to the formation of the parasitophorous vacuole in which the parasite resides. Rhoptry-associated protein 1 (RAP-1) plays a role in the latter process (Counihan *et al.,* 2013), although its precise function is unknown. RAP-1 forms a complex with smaller proteins, RAP-2 or RAP-3, and deletion of the RAP-1 gene results in mistargeting of RAP-2 to the rhoptries (Baldi *et al*., 2000).

P. falciparum RAP-1 gene is highly conserved between isolates, with only 9 amino acid substitutions from 783 amino acids that make up the protein (Fonjungo *et al.,* 1998). Limited polymorphism in *P. falciparum* RAP-1 may suggest that it is not under immune pressure. However, there is evidence that the protein is recognized by the host's immune system and that antibodies to the protein might inhibit merozoite invasion (Pacheco *et al*., 2010). For example, monoclonal antibodies raised against *P. falciparum* RAP-1 hindered erythrocyte invasion *in vitro* (Harnyuttanakorn *et al*., 1992; Schofield *et al*., 1986), and partial protection against *P. falciparum* challenge infection was observed in *Saimiri sciureus* and *Saimiri boliviensis* monkeys immunized with RAP-1 and RAP-2 (Baldi *et al.,* 2000; Collins *et al*., 2000; Perrin *et al.,* 1985). Although there have been extensive studies of *P. falciparum* RAP-1 (Fonjungo *et al.*, 1998; Howard & Peterson, 1996; Perez-Leal *et al*., 2006), studies on the *P. knowlesi* orthologues are limited.

In a recent investigation, it was demonstrated that negative selection might be acting on the RAP-1 of non-human primate parasites, including *P. knowles*i (Pacheco *et al*., 2010). The study, however, used only five old (isolated 1953-1962) *P. knowles*i strains, which may not reflect the actual picture of polymorphism in *P. knowles*i RAP-1. Further studies on the genetic diversity and natural selection of *P. knowlesi* RAP-1 may provide valuable information and extend the knowledge on population genetics of malaria antigens. In the present study, the RAP-1 of 30 recently isolated *P. knowlesi* from Malaysia was analysed to obtain a better picture of the parasite's diversity. The RAP-1 gene of each isolate was cloned and sequenced. Sequence analyses were performed to elucidate the possible selective forces that might contribute to the diversity and grouping of the RAP-1 of *P. knowlesi*.

In summary, the objectives of this research were:

1. To amplify and sequence the RAP-1 gene of *P. knowlesi* isolates from Malaysia
2. To determine the genetic diversity and natural selection of the gene
3. To perform the phylogenetic analyses on the gene

CHAPTER 2: LITERATURE REVIEW

2.1 Malaria

2.1.1 Malaria in general

Malaria is one of the most critical parasitic diseases worldwide, with more than 3 billion people at risk of getting the disease. It is caused by protozoan parasites called *Plasmodia* of the phylum Apicomplexa (Kats *et al*., 2008). This mosquito-borne disease is widespread in tropical and subtropical regions, including Sub-Saharan Africa, Asia, and the Americas. Due to the high morbidity and mortality rates it can induce, malaria is considered to be the primary parasite disease of humans globally (Preiser *et al*., 2000). Natural transmission of the disease begins with an infected female *Anopheles* mosquito that introduces the parasite into a person's circulatory system from the saliva through a blood meal. A single blood meal is enough to spread the disease. It can lead to flu-like symptoms, which if not treated can lead to coma or death.

In 2015, an estimated 214 million cases of malaria and 438,000 deaths were reported, where the majority occurs in sub-Saharan Africa (WHO, 2015). Asia, Latin America and parts of Europe and Middle East are also at risk, in which 97 countries and territories have ongoing malaria transmission. People from malaria-free countries can also acquire infection whilst travelling to malaria endemic countries or through blood transfusion (although the cases are very rare) (WHO, 2015). The parasite can also be transmitted vertically from mother to fetus, which may be fatal. The costs spent for malaria including treatments and intervention strategies are enormous, resulting in heavy economic burdens in affected countries.

2.1.2 History of malaria

Malaria or a malaria-resembling disease was discovered more than 4,000 years ago. It is named based on the Italian word mal'aria for "bad air" and its symptoms have been described in ancient Chinese medical writings (CDC, 2015). The malaria parasite was discovered in 1880 by Laveran, a French army surgeon (Cox, 2010). The malignant tertian malaria parasite, *P. falciparum,* was initially known as *Oscillaria malariae* by Laveran and was thought to be the only species to cause malaria. Other species including *P. vivax* and *P. malariae* were later identified and named by the Italian investigators, Grassi and Filetti (CDC, 2015). In 1922, Stephens described *P. ovale* as the fourth human malaria parasite. The fifth human malaria parasite, *P. knowlesi* was named after Robert Knowles and its first human infection was documented in 1965. In 2013, *P. ovale wallikeri* (the variant type) was distinguished from *P. ovale curtisi* (the classic type) using real-time PCR molecular method (Calderaro *et al*., 2013). It was reported to be the sixth human malaria parasite.

2.2 *Plasmodium*

2.2.1 Human *Plasmodium* species

More than 100 species of *Plasmodium* have been identified. These *Plasmodium* species can infect various animal species including reptiles, birds, and mammals. Among them, five species have been recognized to infect humans. Of the five species, *P. falciparum* is responsible for the most serious form of the disease, followed by *P. vivax, P. ovale,* and *P. malariae* (Kats *et al.*, 2008). The fifth species, *P. knowlesi,* is a parasite of long-tailed macaque monkeys, has now been found to cause severe infections in humans (Vythilingam *et al*., 2008). This zoonotic species is prevalent in Southeast Asia and the first case was reported in Peninsular Malaysia in 1965 (Chin *et al.*, 1965).

P. falciparum is the most lethal form of human malaria species. It can cause severe infections and death if the treatment is not promptly administered. This parasite is found worldwide in tropical and subtropical areas, but endemic in many countries in Africa, especially south of the Sahara desert. Nearly 1 million deaths are reported annually due to *P. falciparum* infection in these regions (CDC, 2015). It proliferates rapidly in the blood, which leads to severe blood loss and subsequently blocks small blood vessels. This complication can be fatal as it can trigger cerebral malaria affecting the brain.

P. vivax and *P. ovale* are two human malaria species that can develop dormant liver stages (hypnozoites). These stages can remain latent up to 2 years (*P. vivax)* or 4 years (*P. ovale*) (CDC, 2015). Unlike *P. vivax*, *P. ovale* is capable of infecting individuals of non-Duffy blood group. This explains the greater number of cases of *P. ovale* in sub-Saharan Africa. *P. malariae* is known to be the only human malaria species that develops a quartan cycle (three-day cycle). It is found worldwide and causes a long-lasting chronic infection if left untreated. It is generally correlated with uncomplicated clinical course and low parasitemia. However, in some cases, it can contribute to serious complications including nephrotic syndrome.

2.2.2 *Plasmodium knowlesi*

P. knowlesi is the only natural pathogen of long-tailed and pig-tailed macaques known to infect humans and cause malaria disease. *P. knowlesi* is genetically related to *P. vivax* but they differ in the length of life cycle and the occurrence of hypnozoite stage (Cox-Singh *et al*., 2008). *P. knowlesi* has a 24-hour asexual life cycle, which is the shortest replication cycle among the malaria species (Knowles & Gupta, 1933; Sinton & Mulligan, 1933). This rapid progress can lead to a severe infection due to elevated

parasitemia in the infected individuals.

Previously, *P. knowlesi* was mistakenly diagnosed by microscopy as *P. malariae* due to similar morphology (Singh *et al*., 2004). In 2004, a large-scale screening of malaria patients' blood was performed using PCR methods and found that 58% of patients from Kapit Hospital, Sarawak, Malaysia were actually infected with *P. knowlesi* (Cox-Singh *et al*., 2008). Subsequently, human knowlesi malaria cases were reported in West Malaysia (Lee *et al.,* 2010; Vythilingam *et al*., 2008) and other Southeast Asian countries including Thailand (Jongwutiwes *et al.,* 2004), Myanmar (Zhu *et al.,* 2006), Singapore (Jeslyn *et al*., 2011; Ng *et al*., 2008), the Philippines (Luchavez *et al*., 2008), Vietnam (Van den Eede *et al*., 2009) and Indonesia (Figtree *et al*., 2010; Sulistyaningsih *et al.*, 2010).

The first isolation of *P. knowlesi* was performed in 1931 from a long-tailed macaque (*Macaca fascicularis*) imported from Singapore to India (Napier & Campbell, 1932). Knowles and Das Gupta, who conducted the early experiments, observed a few clinical manifestations caused by the species (Knowles & Gupta, 1933). Infected natural host *M. fascicularis* was noted to be asymptomatic at low levels of parasitemia. However, *P. knowlesi* was found to be lethal for Indian rhesus macaques (*M. rhesus*). Studies performed on humans demonstrated that *P. knowlesi* infection could be acquired through blood passage (Knowles & Gupta, 1933). In 1965, a report on an American army traveller infected with *P. knowlesi* in Pahang was described (Chin *et al.,* 1965). This was the first case of a human infection by mosquito bite reported under natural conditions. However, investigation on rhesus macaques inoculated with blood samples from villagers in the same area concluded that human acquiring knowlesi infection was rare (Coatney, 1968; Warren *et al.,* 1970).

2.2.3 Life cycle of *Plasmodium*

The life cycle of *Plasmodium* is complex, involving mosquito as the vector and human as the host (Figure 2.1). Three main phases- sporogonic, pre-erythrocytic, and erythrocytic stages, are required to complete one cycle (CDC, 2015). During the first stage, a mosquito ingests gametocytes of *Plasmodium* during a blood meal. Fertilization then results in the production of zygotes, which takes place in the mosquito's gut. Zygotes develop into motile, elongated ookinetes that can penetrate the mosquito's midgut wall. The ookinetes will mature into oocysts, which then grow, divide, and rupture to release sporozoites that move to the salivary glands.

The infectious cycle starts when the mosquito feeds on blood from a human host. The saliva that contains sporozoites is injected into the human bloodstream. Hepatocyte invasion occurs after one to two weeks, where sporozoites undergo asexual reproduction to develop into schizonts. Once a liver schizont has matured, the erythrocytic cycle begins to take place, where thousands of merozoites are released into the bloodstream. The merozoites transform into trophozoites (ring-form), which begin another round of asexual multiplication, and develop into schizonts. When schizonts rupture, merozoites will be released into the bloodstream, where it can subsequently infect other red blood cells. One schizont can produce up to 32 merozoites. The merozoite stage is responsible for most of the clinical manifestations of malaria.

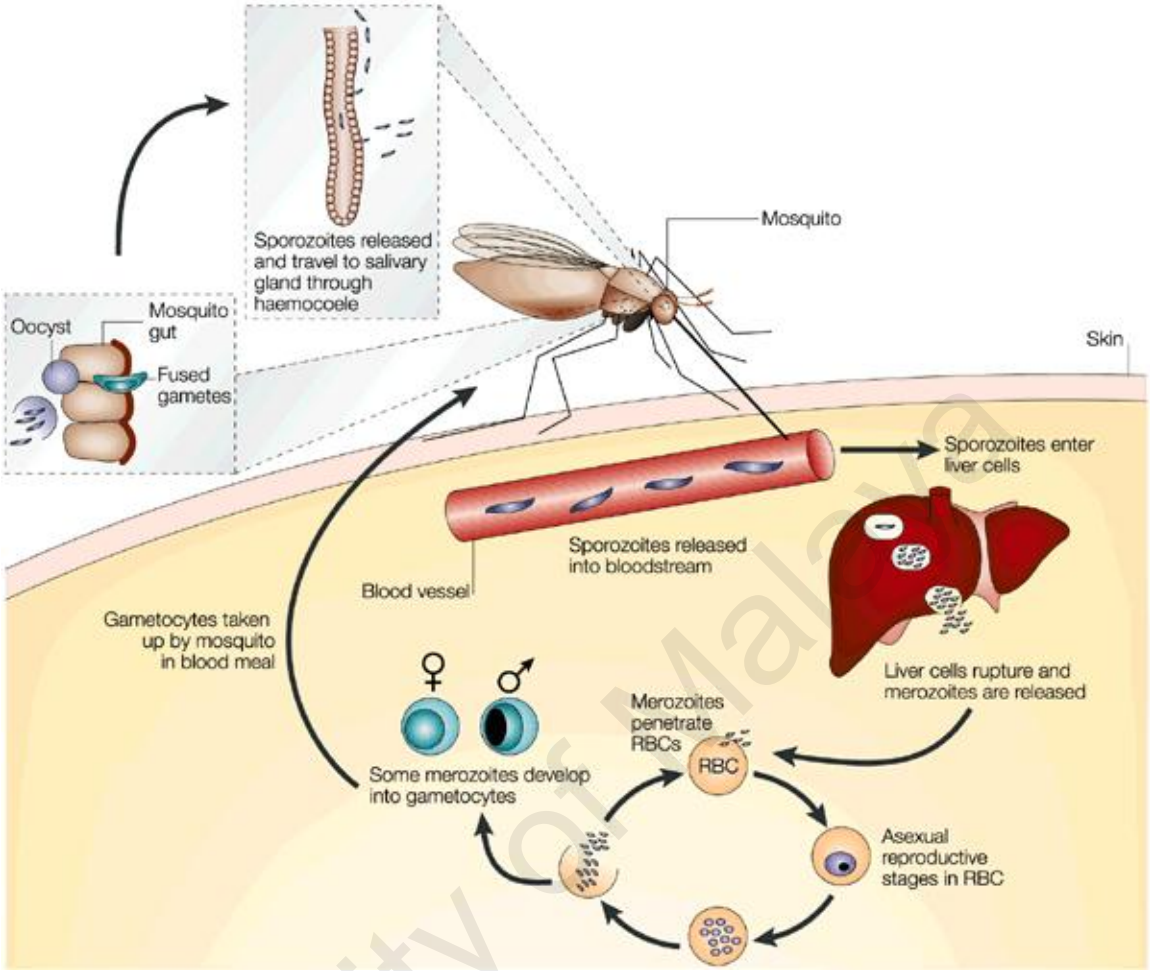

Figure 2.1: The life cycle of *Plasmodium* (adapted from Cooke & Hill, 2001)

2.3 Malaria vector

In nature, only 30-40 *Anopheles* species have been identified to be the vectors for malaria. *Anopheles* mosquito can be found worldwide except Antarctica (CDC, 2015). Similar to other mosquitoes, anophelines have four stages of life cycle, which are egg, larva, pupa, and adult. The first three stages require an aquatic environment and last about 5-14 days, depending on the temperature and species. The *Anopheles* mosquito plays its role as a malaria vector during the adult stage. It normally lives not more than 1-2 weeks in nature but is capable to live up to a month or more in captivity. *Plasmodium*'s life cycle from the gametocyte stage to the sporozoite stage in the vector depends on several factors. Ambient temperature and humidity contribute to the successful development of the malaria parasite in the mosquito (WHO, 2015). Survival of *Anopheles* mosquito with the parasite infection is crucial for parasite growth.

2.4 Malaria epidemiology and prevalence

Distribution of malaria infection can be affected by several factors. Climatic factors such as temperature, humidity, and rainfall can influence the transmission of malaria (CDC, 2015). Particularly, temperature is critical for the parasite to grow; for example, *P. falciparum* cannot survive at temperatures below 20°C as it cannot complete its growth cycle in the *Anopheles* mosquito. *P. vivax* is more prevalent in cooler regions as it can sustain at lower ambient temperatures. Malaria is widespread in tropical and subtropical areas except during colder seasons, in deserts (excluding the oases) or at very high altitudes. Transmission is generally more intense in warmer regions closer to the equator. The highest transmission occurs in sub-Sahara Africa and in Oceania regions such as Papua New Guinea (CDC, 2015). In Western Europe and the United States, strategies on eliminating malaria have been successful due to economic development and public health awareness.

2.5 Malaria symptoms

Prompt treatment and correct diagnosis can cure malaria. Malaria is generally an acute febrile illness. A wide variety of symptoms can develop from *Plasmodium* infection, ranging from very mild symptoms to very severe manifestations, which can be fatal. *P. falciparum* infection can give rise to severe illnesses and often causes death, if not treated within 24 hours (WHO, 2015). For *P. vivax* and *P. ovale* infections, clinical relapse can occur for weeks or months after the first infection. This appearance arises due to re-activation of hypnozoites in the liver. Children with severe malaria are commonly associated with severe anaemia, respiratory distress due to metabolic acidosis, and cerebral malaria. In adults, multi-organ complications can occur. Individuals acquiring malaria infection can also develop partial immunity in malaria endemic areas, whereby asymptomatic infections are manifested.

Asexual erythrocytic stage of *Plasmodium* is the main cause for all the clinical symptoms of malaria (CDC, 2015). Various known and unknown waste substances such as hemozoin pigment and other toxic components will be released and accumulated in the infected erythrocyte as the parasite develops. Once the infected cells burst, the invasive merozoites will be discharged along with the waste substances that are deposited into the bloodstream. This results in an activation of macrophages and other cells to produce cytokines that can induce fever and rigors. Severe pathophysiology related to malaria takes place subsequently on account of stimulation by the hemozin and other toxin factors including glucose phosphate isomerase (GPI).

Malaria can be classified into two categories: uncomplicated and severe (complicated). The classical symptoms of uncomplicated malaria can be a combination of fever, chills, sweats, headaches, nausea or vomiting, and body aches (CDC, 2015).

The attack usually occurs every 24 hour for *P. knowlesi* infection, every second day for *P. falciparum*, *P. vivax*, or *P. ovale* and every third day for *P. malariae* infection. Patients with mild malaria infection can be mistakenly diagnosed as influenza, a cold or other common infections as they manifest almost the same symptoms.

Patients with severe malaria may exhibit cerebral malaria that creates seizures, coma, consciousness dysfunction or other neurologic abnormalities (CDC, 2015). Other symptoms of severe malaria include acute anemia (impairment of the red blood cells), hemoglobinuria (hemoglobin in the urine), blood coagulation abnormalities, severe kidney failure, hyperparasitemia, and hypoglycemia. Severe malaria patients can also acquire acute respiratory distress syndrome (ARDS), a syndrome that inhibits oxygen exchange, due to an inflammatory reaction in the lungs. Urgent and vigorous treatment with a medical emergency is required to administer severe malaria patients.

2.6 Treatments and diagnosis of malaria

Early diagnosis and treatment of malaria is critical in reducing the severity and transmission of the disease. Artemisinin-based combination therapy (ACT) is the best current treatment available, particularly for *P. falciparum* malaria (WHO, 2015). All suspected malaria cases confirmed by parasite-based diagnostic tools are recommended before conducting the treatment. Microscopy remains the gold standard for confirming malaria species. Malaria rapid diagnostic test is also widely used in which the results can be obtained within 15 minutes or less. Due to similar morphology found in certain species, microscopy may produce false results. Polymerase chain reaction has been proven to be a reliable method for confirming and differentiating malaria species. It provides more reliable results and more specificity than microscopic examination due to its ability to identify malaria parasites at the species level when microscopy is

ambiguous. The need to differentiate malaria species is vital especially during *P. falciparum* or *P. knowlesi* infection. *P. falciparum* and *P. knowlesi* can lead to fatal; prompt identification of species is required to determine effective treatment and to avoid more serious complications that can lead to death.

2.7 Malaria intervention strategies

Vector control remains the major approach on minimizing malaria transmission at the community level. It is capable of reducing the transmission by the mosquito to humans and vice versa. Vector control comprises mainly of two approaches: insecticide-treated mosquito nets (ITNs) and indoor residual spraying (IRS) (WHO, 2015). One of the preferred forms of ITNs during public health distribution programs is the long-lasting insecticidal nets (LLINs). Indoor spraying with residual insecticides is also efficient in reducing malaria. It is recommended that at least 80% of houses in targeted areas be sprayed in order to effectively eradicate the mosquitoes. Indoor spraying can be effective up to 3-6 months. Annual rotation of IRS is endorsed in the areas where IRS is the main intervention for vector control to maintain its effectiveness.

Antimalarial medicines are also available to prevent malaria. Chemoprophylaxis, a prevention that uses a chemical agent to suppress the blood stage of malaria parasites, is handy for individuals, particularly travellers (WHO, 2015). In 2012, Seasonal Malaria Chemoprevention as an additional malaria intervention strategy is advised for areas of the Sahel sub-region of Africa (WHO, 2015). This strategy includes a monthly-course administration of amodiaquine and sulfadoxine-pyrimethamine to all children under 5 years old during peak season of transmission. For pregnant women who live in high transmission areas, sulfadoxine-pyrimethamine intermittent preventive treatment is implemented at each scheduled antenatal visit after

the first trimester. Similarly, 3 doses of intermittent treatment with sulfadoxine-pyrimethamine are given to infants.

Vaccine against *P. falciparum*, known as RTS, S/AS01, is currently under evaluation in seven countries in Africa (WHO, 2015). The recommendation of this vaccine to be an intervention strategy for malaria control is expected to be in late 2016. On the other hand, research vaccines against other *Plasmodium* species is still at premature stage.

The design of antimalarial vaccines is crucial as vaccine generates the first line of defence to protect the host from getting more serious complications. In general, vaccine candidates are from the surface proteins of parasite's asexual stage (Richie & Saul, 2002). These proteins are required for initial interaction of the sporozoites or merozoites with the hepatocytes or erythrocytes respectively. They are greatly exposed to the surrounding, whereby the bloodstream contains lymphocytes and platelets to provide protection against the infection. These proteins are usually antigenic, therefore able to provoke the host's immune system to excrete protectors such as T cells and antibodies.

Low genetic variability of the candidate proteins is also an important consideration in vaccine design (Richie & Saul, 2002). Such strategy however is hindered by the parasite's multiplex biology. Rapid exposure has driven the surface proteins to express high polymorphism in order to escape from the host's immune system. These variants are said to provide protection for the parasite to continually replicate and invade the host. Other proteins, which are not located on surface, may involve in adhesion and invasion processes such as during tight junction and

parasitophorous vacuole formations. They may not be directly exposed on the parasite membrane; they are less polymorphic as genetic variation is not crucial for them to provide protection for the parasite. These low polymorphic proteins are now the focus of vaccine development.

2.8 *Plasmodium* organelles and proteins

2.8.1 Importance of organelles in blood invasion

The malaria parasites are characterized with three main distinct organelles, collectively known as apical organelles as they cluster at one end of the parasites (Baum *et al*., 2008; Kats *et al*., 2008). These main apical organelles are known as rhoptries, micronemes and dense granules (DGs) (Figure 2.2). Exoneme is also one of the organelles, but is less characterized. These organelles discharge a myriad of the parasite proteins during invasion to aid the parasite's rapid invasion into the human hosts (Kats *et al*., 2008). The formation of the apical organelles takes place during the final 10-12 hours of parasite development inside the host's circulatory system (Margos *et al*., 2004). Following two rounds of nuclear division, the rhoptry organelles are synthesized. The micronemes and dense granules are generated at the completion of nuclear division, resulting in fully segmented merozoites (Lobo *et al*., 2003; Topolska *et al*., 2004).

Release of various proteins into the host occurs in two locations. The first involves proteins on the parasite surface. The other involves proteins within the specialised secretory organelles (Kats *et al*., 2008). The three most invasive forms of *Plasmodium*, i.e., ookinetes, sporozoites, and merozoites, contain these organelles. However, the ookinetes lack the rhoptry organelle (Vlachou *et al.,* 2006). The micronemes are found less in the merozoite (Cowman & Crabb, 2006), but are abundant in the sporozoites (Prudencio *et al.,* 2006). Larger rhoptries are observed both in the

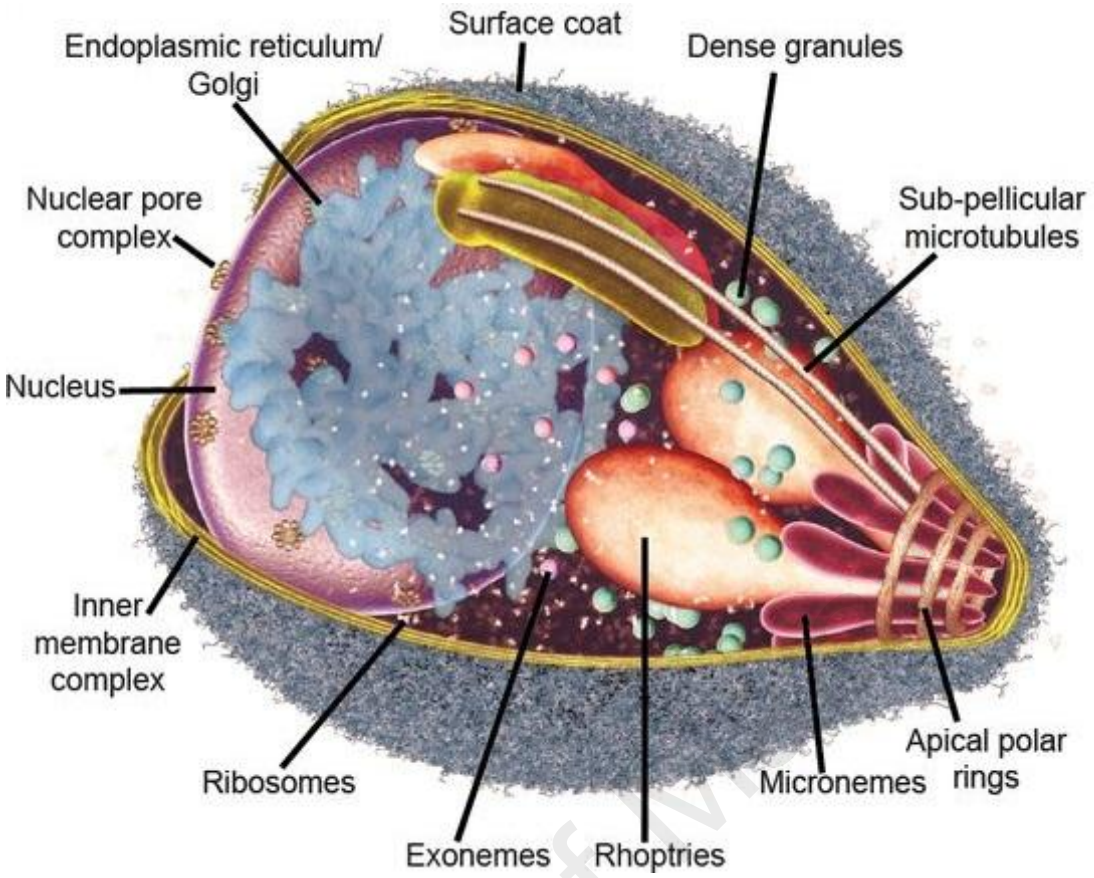

Figure 2.2: Three-dimensional diagram of a *Plasmodium* merozoite and its main secretory organelles (adapted from Cowman *et al*., 2012)

merozoite and sporozoite.

The most important stage in the human host, the merozoite involves initial merozoite binding to the erythrocytes. This reversible interaction requires proteins such as merozoite surface protein 1 (MSP-1) to interact with the host erythrocytes (Cowman *et al.*, 2012; Holder, 1994). Then, proteins such as apical membrane antigen 1 (AMA-1) are released during parasite reorientation and erythrocyte deformation (Mitchell *et al*., 2004). Subsequently, tight junctions are formed to establish a strong connection between the erythrocyte and the parasite. This is the time when microneme starts to secrete major important proteins. Proteins involved in the junction formation are Duffy-binding protein (DBP), sporozoite surface protein 2 (SSP-2), and erythrocyte-binding antigen (EBA) (Wiser, 1999). The rhoptry organelle plays its role in secreting the proteins during parasite entry. It releases numerous proteins to assist in parasitophorous vacuole membrane (PVM) formation. This vacuole fuses to surround the invaded parasite and separates it from the host-cell cytoplasm to provide a hospitable environment for parasite replication (Counihan *et al*., 2013). At a later time, dense granule's proteins are discharged for post invasion and host cell modification.

2.8.2 Rhoptry organelle

The rhoptry is a paired, pear-shaped membrane bound organelle, located at the apical pole of the parasite (Richard *et al*., 2009). It is most visible during the sporozoite and merozoite stages of the parasite. Known to be the primary contributor for the parasitophorous vacuole formation, it is rich in lipids and proteins that are secreted during the invasion. The rhoptry is the biggest apical organelle known. More than 30 rhoptry proteins have been classified (Kats *et al.,* 2006). They are sub-compartmentalised either at the neck or the bulb of the organelle, where the neck is

electron-lucent while the bulb is electron-dense (Counihan *et al*., 2013). Rhoptry-associated protein (RAP), rhoptry-associated membrane antigen (RAMA), rhoptry neck protein 3 (RON3), rhoptry protein (Rhop), rhoptry-associated, leucine zipper-like protein (RALP), and repetitive organellar protein (ROPE) are among the proteins known to be located at the rhoptry bulb, while reticulocyte-binding-like homologue protein (Rh) and rhoptry neck protein 6 (RON6) are situated at the rhoptry neck. The rhoptry bulb proteins are generally distinct in each species due to different host cells that each species infects, whether they are nucleated or nonnucleated cells (Boothroyd & Dubremetz, 2008). The rhoptry proteins are essential for post-invasion processes.

2.8.3 Rhoptry-associated protein 1 (RAP-1)

The rhoptry bulb contains two well-characterized protein complexes: RAP and RhopH complexes, which are conserved among the malaria species (Counihan *et al*., 2013). RAP-1 and RAP-2/RAP-3 form the low molecular weight (RAP) protein complex during the merozoite invasion, thus are important in maintaining the blood-stage cycle. The development of the protein complexes for accurate targeting is pivotal in order to avoid mislocalization. It has been shown that localization of RAP-2 to the rhoptry is dependent upon the RAP-1 protein (Baldi *et al*., 2000). Removal of RAP-1 sequence that covers RAP-2 or RAP-3 binding sites can result in mistargeting of RAP-2 or RAP-3 to endoplasmic reticulum (ER), which subsequently leads to degradation. It has also been shown that mislocalization is likely to occur at the rhoptry neck instead of the bulb when C-terminal of RAP-1 sequence is deleted. This will hinder localization of the proteins to PVM. The RAP protein biogenesis is suggested to occur through secretory pathway, having hydrophobic N-terminal sequences as a signal peptide (Howard & Schmidt, 1995).

The RAP complex has been implicated in key aspects of invasion including interaction with red blood cell receptors (Sterkers *et al*., 2007). This is evident by the capability of antibodies against RAP-1 and other rhoptry proteins to hinder invasion. The RAP complex is found to interact with rhoptry-associated membrane antigen (RAMA) at Golgi body via RAP-1 N-terminus. RAMA serves as an escorter to link the complex to cytoplasmic trafficking machinery (Topolska *et al*., 2004). During the last 6 hours of the intra-erythrocyte cycle, the RAP-1 encoding gene is transcribed (Bozdech et al., 2003). RAP-1 is a soluble protein, having a serine-rich N-terminal sequence, with no transmembrane regions (Kats *et al*., 2006). Like other *Plasmodium* rhoptry bulb proteins, RAP-1 shares no homology with other Apicomplexan's rhoptry proteins (Counihan *et al*., 2013). This suggests that they have evolved to adapt with their target host cell. Antibodies against RAP-1 have been proposed to be the protector against disease caused by toxin-like exoantigens (Jakobsen *et al*., 1993). RAP-1 triggers *in vitro* production of tumor necrosis factor by human mononuclear cells, whereby acting as a component of the endotoxin-like exoantigen to remove the antigens from the circulation.

In vivo maturation of RAP-1 involves proteolytic cleavage with sequential steps from the N-terminus (Howard *et al.,* 1998). The immature RAP-1 (84 kDa) is cleaved with multiple processing steps until it becomes matured (82 kDa). Further processing of RAP-1 into a 67-kDa molecule is also observed, but is precluded from involving in the blood invasion. Lymphocytes giving response to RAP-1 N-terminal (from amino acids 23-294) suggests that T-cell epitopes are present in this region (Jakobsen *et al*., 1993). Although more epitopes are found outside of the region, immunoglobulin G (IgG) antibodies are mostly reactive toward the N-terminal, rather than the C-terminal (Fonjungo *et al*., 1998). Antigenicity studies on recombinant RAP proteins showed that

rRAP-1 was strongly recognized by the antibodies, while rRAP-2 was poorly identified. Further studies on RAP-1 C-terminus, whether it is removed or modified is vital for further characterization of the protein.

Although RAP-1 is among the vaccine candidate proteins, its polymorphism is not fully studied (Pacheco *et al*., 2010). RAP-1 from *P. falciparum* isolates shows a maximal conservation of amino sequence between isolates, with only 9 amino acid substitutions identified (Howard, 1992; Howard *et al.,* 1993). This limited polymorphism in the *P. falciparum* RAP-1 suggests that it is not under immune pressure (Pacheco *et al.*, 2010). However, there is evidence that this protein is recognized by the host's immune system, thus antibodies against it may block the merozoite invasion. For example, monoclonal antibodies raised against *P. falciparum* RAP-1 hindered erythrocyte invasion *in vitro* (Harnyuttanakorn *et al*., 1992). Partial protection against *P. falciparum* challenge infection was also seen in *Saimiri sciureus* and *S. boliviensis* monkeys immunized with RAP-1 and RAP-2 (Collins *et al.*, 2000; Ridley *et al.*, 1990).

Although there have been extensive studies of *P. falciparum* RAP-1 (Fonjungo *et al*., 1998; Howard *et al*., 1998; Howard & Peterson, 1996; Stowers *et al*., 1997) studies of the *P. knowlesi* orthologues are limited. In a recent investigation, it was demonstrated that negative selection might be acting on RAP-1 in non-human primate parasites, including *P. knowlesi* (Pacheco *et al*., 2010). The study, however, only examined five old strains of *P. knowlesi*, which may not reflect the actual picture of polymorphism in *P. knowles*i RAP-1.

2.9 Evolutionary studies

2.9.1 Parasite immune evasion

Immune evasion is a strategy of the parasite to evade the host's specific immunity by accumulation of genetic mutations that express proteins with different antigenicity. This extensive diversity of malaria antigens enables the parasite to manifest antigenically different alleles to thwart the host's immune responses (Hisaeda *et al.*, 2005). It is also proposed that these strategies are crucial for malaria pathogenesis. Successful completion of each cycle in malaria pathogenesis is achieved when the parasites are able to escape the immune recognition. Several mechanisms have been outlined in malaria immune evasion. Intracellular parasitism provides the initial escape mechanism by inhabiting the interior of the cell (Hisaeda *et al.*, 2005). In the context of malaria parasites, sporozoites and merozoites invade the host without provoking the immune system by entering hepatocytes and erythrocytes respectively. Additionally, merozoites are able to evade recognition by CD8+ T cell, as the red blood cells express no MHC molecules.

Antigenic diversity or polymorphism is a powerful strategy to escape host immunity. Antibodies and T cells work in a very strict manner, whereby recognition of antigen epitopes is dependent on the memory developed from the first infection (Pancer & Cooper, 2006). T cell recognition is precisely based on the primary structure of proteins (amino acid sequence) rather than protein conformation. Single nucleotide polymorphism (SNP) that changes the amino acid sequence can contribute to failure of triggering protective T cells. The parasites rapidly express antigenically different alleles of a gene, which will be selectively expanded afterward to avoid recognition by host immunity.

2.9.2 Polymorphism and natural selection

Extensive polymorphism detected in the parasites is due to the strategy to evade the host's immune responses; this phenomenon is strongly associated with selection on the parasites. Natural selection is discussed whenever elicited immune response or drug resistance is observed, although it can occur as a result of any process that can affect the reproduction of the parasite population (Escalante *et al*., 2004). Immune responses or antimalarial drugs function as selective pressures, affecting the outcomes of differences in variants reproduction in a given population. A gene is subject to strong selection when differences in alleles or variants are observed in the gene, and the immune responses or drugs drive the gene to express polymorphism to confer drug resistance or immune escape (Escalante *et al.*, 1998).

Changes in genetic variants (nucleotide variation) that have a higher capacity of altering the results of phenotypes (amino acid variation) rather than unchanged ones, cause positive (balancing) selection to occur (Escalante *et al*., 2004). This event exhibits higher nonsynonymous substitutions (nucleotide substitutions that change amino acids) than synonymous substitutions (nucleotide substitutions that don't change amino acids) on the respective genes, in a given population. The genetic variants are said to be favoured, and are maintained in the population. It leads to the development of immunity against the host and sustainment of diversity within local populations (Tetteh *et al*., 2009).

Negative (purifying) selection is observed when the selective forces drive the genes to manifest more nucleotide substitutions, but less greatly alter the amino acid sequences (synonymous substitutions are greater than nonsysnoymous substitutions) (Escalante *et al*., 2004). This scenario happens when the mutations detected are less

advantageous or more harmful toward the parasite; the genetic variants will be decreased or eliminated throughout the population. The selective alleles that are deleterious are removed, which result in stabilization of the selection and reduced chance of the genotypes to be maintained throughout the population (Loewe, 2008).

Under neutrality, nonsynonymous substitutions often are not preferable. Due to the factor of changes in amino acids, they are vulnerable to a deleterious effect, a consequence that influences the substitutions to remain in low frequencies or to be completely eliminated (Kimura, 1977). On the other hand, many times sysnonymous substitutions are regarded as neutral or almost neutral, on account of the nucleotide mutations that do not change the protein sequences and functions (Ohta, 1992). There has been a demanding interest in polymorphism studies, particularly on molecular evolution of genes encoding malaria antigens (Escalante *et al.,* 1998). *P. falciparum* and *P. vivax* are the most widely studied *Plasmodium* parasites (Cole-Tobian & King, 2003; Escalante *et al.,* 2001; Polley & Conway, 2001), although formal analyses for revealing evidence of natural selection acting on the parasite population are relatively new. It was shown that surface proteins such as apical membrane antigen 1 (AMA-1), circumsporozoite surface protein (CSP), and merozoite surface protein (MSP) are more polymorphic than erythrocyte-binding antigen 175 (EBA-175), *P. falciparum* surface protein 25 (PFS-25) and RAP-1 (Anders & Saul, 1994; Escalante *et al.,* 1998). The proteins expressed on the surface of merozoites or sporozoites are more exposed to the host's immune system than those expressed during the asexual stage or inside the parasite (Escalante *et al.,* 1998; McCutchan *et al.,* 1988). These proteins are subject to strong selection pressure to accumulate polymorphism as a means of immune escape.

Currently, no gold standard procedure has been assigned for identifying natural selection, but several tests available utilize different approaches for estimating the selection (Escalante *et al.*, 1998). Normally, these tests are run simultaneously to confirm the possible selection. Two different ways of estimating natural selection are developed and implemented in most software such as MEGA6 (Tamura *et al.*, 2013). The first approach is based on the distribution of allele frequencies and/or segregating sites, while the second one takes the pattern of polymorphism as a basis and associates it with a phenotypic change (Escalante *et al.*, 1998).

Comparison of the synonymous (d_S) and nonsynonymous (d_N) mutation rates is the most common applied method, which explicitly uses the phenotypes (frequency of amino acid changes) to estimate the selection. Under neutral circumstances, it is assumed that the nucleotide changes to occur naturally would be around 0.0001 or 1 base pair change per 10000 nucleotides. Similar or nearly similar rates of synonymous and nonsynonymous substitutions are expected under neutrality, constituting a value of (d_N/d_S) approximately to 1 when compared. A faster accumulation of synonymous substitutions will be observed as a replacement of nonsynonymous substitutions, which often have the tendency to be negatively selected (Kimura, 1977). Departure of neutrality takes place when the observed nucleotide changes are more or less than the expected nucleotide changes. Nucleotide substitutions are produced excessively, provided that the values of (d_N/d_S) are significantly more or less than 1. McDonald-Kreitman (MK) test also compares the patterns of polymorphism within species (Mcdonald & Kreitman, 1991), adding divergence into the method to present substitutions between species. Synonymous and nonsynonymous substitutions in both estimates (within and between species) are counted and compared by using a 2×2 contingency table. MK test is a very convincing method for detecting selection, as it

shows less sensitive to departures of neutrality including population structure and population growth.

Tajima's D test utilizes the first approach based on the neutral model prediction that suggests two different ways of estimating the expected heterozygosity from a random sample of alleles obtained from a population (Tajima, 1989). The number of segregating sites and pairwise differences among alleles are estimated using parameter θ. The neutral model, which is the null hypothesis for the test, is rejected when those two estimates show θ with statistically significant disparity. Other tests that are known to have a similar approach are Fu and Li's D test and Fu and Li's F test (Fu & Li, 1993).

2.9.3 Phylogenetic analysis

The study of evolutionary relationships among the *Plasmodium* species and populations through phylogenetics has been a subject of considerable interest in genetics (Gillespie, 1991; Kimura, 1983; Ohta, 1992, 1996). Phylogenetic analysis is a means of estimating relationship through evolutionary history among organisms, depicting branching, tree-like diagrams as estimated lineage of the inherited relationships (Baxevanis & Ouellette, 2004). Members who are grouped together are thought to share unique features or a common evolutionary history, and are more related to each other than to members of another group.

A phylogenetic analysis encompasses four fundamental steps: alignment, substitution model determination, tree building, and tree evaluation (Baxevanis & Ouellette, 2004; Larkin *et al*., 2007). It usually involves multiple sequence alignments of nucleotide or amino acid sequences. They are positioned vertically according to their similar types of nucleotide base pairs or amino acids, which are individually referred to

as ‘sites’. Sites with more than one nucleotide or amino acid difference is called segregating sites. These sites determine the heterozygosity of each analyzed sequence and branches that are formed in the tree. CLUSTALW is a typical alignment procedure widely used to align multiple sequences, with manual alignment editing performed subsequently (Larkin *et al*., 2007).

A few methods of tree building are developed based on the suitability of the data sets and effectiveness in constructing the tree. Two major methods, distance and character-based methods have different approaches on establishing a tree, and are implemented in current available software (Saitou, 1996). Distance methods incorporate pairwise distances, depending on some measure, and derive trees by eliminating the actual data and using only the fixed distances. The amount of dissimilarity (the distance) between two aligned sequences will be computed and compared, but is disadvantageous due to the discarding of the actual character data. The distance matrix methods, however, are less computationally comprehensive than the character-based methods; rapid results of the tree can be obtained easily. Neighbor-joining (NJ) and Fitch-Margolish (FM) methods are frequently used methods for pairwise distances (Baxevanis & Ouellette, 2004).

The character-based methods construct trees based on the actual data patterns of each character, which allow more reliable evaluation at each base position in an alignment on account of all other base positions. The most common character-based methods are maximum parsimony (MP) and maximum likelihood (ML) methods. Maximum parsimony method requires the least assumptions, whereby the concept of the best explanation of the data is the simplest is exerted. Practically, MP tree is derived from the one with the fewest parallel changes and is the shortest. On the other hand,

maximum likelihood method takes the tree with the highest likelihood of producing the observed data. It is calculated based on the probability of the variation that would be produced at a site by a particular substitution process. The probabilities of each possible reconstruction of substitutions will then be summed up subsequently, and the likelihoods for all the sites are multiplied to obtain an overall 'likelihood of the tree' (Baxevanis & Ouellette, 2004).

Conclusively, the neighbor-joining method (distance matrix method) merely calculates the number of differences between two sequences, provided that the most similar sequences (those with the shortest distance) minimize the length of the tree. The maximum parsimony method employs the principle of having a tree with the smallest number of changes to describe the observed differences among the group or taxa. The maximum likelihood method uses the probability of nucleotide or amino acid changes to be generated from the observed data. The sum up of the probabilities for all possible reconstruction is performed to get the likelihoods for all alignment positions in the data set (Baxevanis & Ouellette, 2004).

2.9.4 Evolutionary Concepts in Malaria

Genetic diversity and selection vary among the vaccine candidate proteins in the *Plasmodium* species. Different demographic areas, time, hosts and species may also shape the variation in population diversity. In *P. falciparum*, the genetic polymorphism is distributed unequally among 10 proteins studied (Escalante *et al*., 1998). Some of these proteins, AMA-1, CSP, LSA-1, MSP-1 and PFS-48/45, have definite evidence for positive selection, while EBA-175, MSP-2, MSP-3 and RAP-1 have limited evidence. Similarly, AMA-1 in *P. vivax* is under positive selection, whereby high genetic polymorphism is found at domain I (Moon *et al*., 2009).

In another study of *P. vivax* in Thailand, MSP-5 exon 1 exhibits higher genetic diversity at both nonsynonymous and synonymous sites than exon 2 (Putaporntip *et al.*, 2010). A signature of positive selection is seen in *P. vivax* MSP-5 exon 1. Interestingly, intragenic recombination has taken place in most populations, but not in southern Thailand, where it is believed to have gone through recent bottleneck events. Spatio-temporal surveillance contributes to the genetic differentiation in *P. vivax* MSP-5.

On the other hand, the *P. knowlesi* AMA-1 gene is not under diversifying selection and is low polymorphic (Faber *et al.,* 2015). This suggests AMA-1 in *P. knowlesi,* unlike in *P. falciparum* and *P. vivax,* is not a main target for protective humoral immune response and may present as a good candidate for a single allele-based vaccine. In a recent investigation, genomic dimorphism was found in more than half of the *P. knowlesi* genome, suggesting two different types of *P. knowlesi* (Pinheiro *et al*., 2015). This evidence is in concordant to other studies of *P. knowlesi* genes including *P. knowlesi* DBPαII (Fong *et al.*, 2014), *P. knowlesi* NBPXα (Pinheiro *et al.,* 2015), *P. knowlesi* AMA-1 domain I (Fong e*t al.,* 2015), and *P. knowlesi* MSP-1 (Putaporntip *et al.,* 2013). Separation of gene haplotypes into two main clusters in the phylogenetic trees supports the postulation of the existence of two distinct *P. knowlesi* types or lineages in Southeast Asia (Muehlenbein *et al.,* 2015).

Further studies on the genetic diversity of *P. knowlesi* RAP-1 may provide valuable information and extend the knowledge on population genetics of malaria antigens. Elucidating the possible selective forces that might contribute to the diversity and grouping of the RAP-1 of *P. knowlesi* provides practical importance for the understanding of epidemic status and vaccine development (Cui *et al*., 2003).

CHAPTER 3: MATERIALS AND METHODS

3.1 Sterilisation

Glassware, instruments, and reagents were sterilized or disinfected before carrying out the experiments. Workspace and PCR laminar hood were sprayed with 70% alcohol. Preparation of PCR reagents and mixes was performed in the PCR laminar hood. Preparation of other reagents including media and competent cells was carried out in a sterilized area near the flame.

3.1.1 Moist heat

Micropipette tips, microcentrifuge tubes, centrifuge tubes (polycarbonate and polypropylene), and Schott bottles with plastic caps, and reagents such as distilled and deionised water (ddH_2O) and media were autoclaved at 15 pound-force per square inch (lbf/in^2) at 121°C for 15 min.

3.1.2 Dry heat

Dry heat sterilisation was performed on glassware such as measuring cylinders and flasks by heating in a hot air oven at 180°C for 1 h.

3.1.3 Membrane filtration

Thermo labile solutions, antibiotics and media were filtered on disposable syringe filters with pore size of 0.22 μm.

3.2 Reagents and chemicals

The highest grade or Analar grade reagents and chemicals were used, which were available from Amresco Inc., U.S.A.; Amersham Pharmacia Biotech Inc., Sweden;

APS Finechem, Australia; Difco, U.S.A.; BDH Ltd, England; Gibco BRL, Life Technologies Inc., U.S.A.; Invitrogen Corp, U.S.A.; MBI Fermentas, U.S.A.; Promega Corporation, U.S.A.; Qiagen, U.S.A.; Sigma Chemical CO., U.S.A. and Laboratories CONDA S.A., Spain.

The commonly used reagents and chemicals were glacial acetic acid, ethylenediaminetetraacetic acid (EDTA), tryptone, yeast extract, dimethyl formamide (DMF), isopropyl β-D-1-thiogalactopyranoside (IPTG), magnesium sulphate ($MgSO_4$), glucose, polyethylene glycerol (PEG) 8000, glycerol, phosphate buffered saline (PBS), acetone, ethanol, methanol, hydrochloric acid (HCl), sodium hydroxide (NaOH), magnesium chloride ($MgCl_2$), sodium chloride (NaCl), tris and tris hydrochloride.

3.3 Stock and working solutions

3.3.1 Gel electrophoresis

Preparation of 1 L 50X tris-acetate-EDTA (TAE) buffer stock solution

Tris base	242.0 g
Glacial acetic acid	57.1 ml
0.5M EDTA (pH 8.0)	100 ml
ddH_2O	up to 1000 ml

To prepare 1X TAE buffer, 20 ml of 50X TAE buffer was added to 980 ml of ddH_2O.

Preparation of the 1% agarose gels

Casting tray size	Weight of agarose (g)	Volume of 1X TAE buffer (ml)
1-comb tray	0.25	25
2-comb tray	0.35	35

3.3.2 Growth medium

Preparation of 1 L Luria- Bertani (LB) medium

Bacto tryptone	10 g
Yeast extract	5 g
NaCl	10 g
ddH_2O	up to 1000 ml

The mixture was autoclaved in moist heat. Fifteen g agar was added before autoclaving to make LB solid medium, and the medium was poured evenly (20-25 ml) onto agar plates.

Preparation of 100 mg/ml ampicillin stock solution

Ampicillin powder	1 g
ddH_2O	10 ml

The stock solution was filter sterilized, aliquoted into 100 µl volumes in microcentrifuge tubes and stored at -20°C.

3.3.3 Blue-white screening

Preparation of 3% X-Gal in DMF stock solution

X-Gal powder	0.15 g
DMF solution	5 ml

The stock solution was stored at -20°C.

Preparation of 100 mM IPTG stock solution

IPTG powder	0.12 g
ddH_2O	5 ml

The stock solution was filter sterilized and stored at -20°C.

3.3.4 *Escherichia coli* competent cells media

Preparation of 50 ml Medium A

1M sterile magnesium sulfate ($MgSO_4.7H_2O$)	500 µl
Glucose	0.1 g
LB broth	up to 50 ml

The medium was prepared freshly before use.

Preparation of 50 ml Medium B

1M sterile $MgSO_4.7H_2O$	6 ml
PEG 8000	6 g
100% glycerol	19 ml
LB broth	up to 50 ml

The medium was filter sterilized and kept at -20°C.

3.4 PCR thermal cycling profiles and reaction volumes

P. knowlesi RAP-1 PCR reaction volume

Component	Final concentration	Component volume
2X GoTaq® Long PCR Mastermix	1X	12.5 µl
*Pk*RAP-1 forward primer (10 µM)	0.4 µM	1 µl
*Pk*RAP-1 reverse primer (10 µM)	0.4 µM	1 µl
Total genomic DNA	100-500 ng	1-4 µl
Nuclease free water	-	up to 25 µl

P. knowlesi RAP-1 PCR thermal cycling profile

Initial denaturation	95°C	2 min
Denaturation	94°C	30 s
Annealing/ extension	63°C	2 min 30 s
Final extension	72°C	10 min

The denaturation and annealing/extension steps were repeated for 35 times.

Colony PCR reaction volume

2X DreamTaq buffer	5.0 µl
M13 (-40) forward primer (10 µM)	0.2 µl
M13 (-48) reverse primer (10 µM)	0.2 µl
ddH_2O	up to 10.0 µl

Colony PCR thermal cycling profile

Initial denaturation	95°C	10 min
Denaturation	95°C	30 s
Annealing	50°C	30 s
Extension	72°C	2 min 30 s
Final extension	72°C	10 min

The denaturation, annealing and extension steps were repeated 30 times.

3.5 Sample collection

Between 2010 and 2014, blood samples of 40 patients with *P. knowlesi* infection were collected from the University of Malaya Medical Centre and a few private clinics in Peninsular Malaysia. Ethics approval for the use of the blood samples was granted by the University of Malaya Medical Centre Ethic Committee (MEC No. 817.18). Two of the samples were from patients who acquired the infection in Malaysian Borneo. *P. knowlesi* infection in each patient was confirmed by microscopic examination of Giemsa-stained thin and thick blood smears and polymerase chain reaction (PCR) amplification using M13 diagnostic primers (Singh *et al.,* 2004).

3.6 Extraction of DNA

P. knowlesi genomic DNA was extracted using the QIAGEN Blood DNA Extraction Kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. Twenty µl proteinase K and 100 µl PBS were pipetted into a 1.5 ml microcentrifuge tube filled with 100 µl of blood sample. The mixture was thoroughly vortexed after the addition of 200 µl Buffer AL and incubated for 10 min at 56°C. Two hundred µl ethanol (96-100%) was then added and the tube was mixed thoroughly by vortexing.

The mixture was transferred into a DNeasy Mini spin column placed in a 2 ml collection tube. Centrifugation was carried out at 8000 revolutions per minute (rpm) for 1 min to separate the DNA from the waste product (Eppendorf microcentrifuge model 5424, 24 × 1.5/2.0 ml capacity rotor). The flow-through and collection tube were discarded. In order to wash the DNA, washing step was performed twice with 500 µl Buffer AW1 and 500 µl Buffer AW2 subsequently. The first washing step was done by centrifugating the spin column at 8000 rpm for 1 min while the second wash was completed with a centrifugation at 14000 rpm for 3 min. The flow-through and collection tube were disposed.

The first DNA elution was performed by transferring the spin column into a new 1.5 ml microcentrifuge tube and adding 30-100 µl Buffer AE into the column. The tube was let to stand for 1 min at room temperature (15-25°C) and then centrifuged at 8000 rpm for 1 min. The elution step was repeated for the second elution, using another new 1.5 ml centrifuge tube with 30-100 µl Buffer AE.

3.7 Amplification by PCR of *P. knowlesi* RAP-1

Amplification of the *P. knowlesi* RAP-1 gene was conducted by PCR using specific oligonucleotide primers *Pk*RAP-1F: 5'-CGT TGA GCA GGA AAT GCC TAC TCC AAT C-3' and *Pk*RAP-1R: 5'-ATG ATA ACG TAC GCA AGT TCT CTG CTG G-3' (Figure 3.1). These primers (nucleotide positions 1782248-1782275 and 1784654-1784681) were designed based on the RAP-1 gene sequence of *P. knowlesi* strain H (GenBank Accession No. AM910995). The high fidelity DNA polymerase GoTaq® Long PCR Mastermix (Promega, Madison, WI, USA) was used to provide a proofreading activity and an efficient longer DNA amplification in the PCR. PCR was

Figure 3.1: Schematic diagram of the *P. knowlesi* RAP-1 gene

Locations of exon 1, intron and exon 2 are shown. Locations of internal sequencing primers are also shown. IntF sequencing primer annealed at nucleotide positions 286- 305, while IntR annealed at positions 2002-2021.

conducted in a total volume of 25 µl, according to the protocol in section 3.4. The PCR product with an expected size of ~2400 base pairs was detected following electrophoresis on 1% agarose gels.

3.8 Agarose gel electrophoresis

Amount of electrophoresis-grade agarose needed to run the gel depends on the size of casting tray (section 3.3.1). The agarose and 1X TAE buffer were added into a 100 ml conical flask and microwaved until the agarose melted. The flask was swirled for even mixing. Approximately 1-1.5 µl SYBR ® Safe DNA Gel Stain (Invitrogen™, U.S.A) was added to the gel mixture before it was hardened. Then the gel was poured onto the casting tray with the gel comb inserted on it. The gel was left to stand for 20-30 min to polymerize.

The polymerized gel was submerged into a gel tank filled with sufficient amount of the 1X TAE buffer. The gel comb was withdrawn from the gel. Five µl of the PCR products and GeneRuler 1kb DNA Ladder (Thermo Fisher Scientific, Waltham, M.A, U.S.A) were loaded into the wells accordingly. The gel electrophoresis was conducted for 30 min or more at 100 V. The DNA bands were visualized under ultraviolet (UV) light at 302 nm excitation wavelength using Molecular Imager® Gel Doc™ XR+ Imaging System (Bio-Rad Laboratories, U.S.A).

3.9 Purification of PCR product

The PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions. The PCR products with nonspecific bands were cut out on the gel to obtain the band of interest only. Five volumes of Buffer PB were added to one volume of the PCR reaction. The mixture was

pipetted up and down to mix until the colour turned to yellow. The mixture was then transferred into a QIAquick column placed in a 2 ml collection tube to bind the DNA. The tube was centrifuged at 13000 rpm for 1 min. The flow-through was discarded and the column was placed back in the same tube.

Seven hunded and fifty μl Buffer PE was added into the column to wash the DNA, and the tube was centrifuged at 13000 rpm for 1 min twice to remove residual wash buffer. The column was placed in a clean 1.5 ml microcentrifuge tube. Thirty μl Buffer EB was added to the center of the QIAquick membrane and the column was centrifuged at 13000 rpm for 1 min to elute the DNA. To confirm the presence of the DNA after purification, the purified DNA was analysed on a gel by adding one volume (1 μl) of 6X loading dye to five volumes (5 μl) of the purified DNA.

3.10 DNA cloning

3.10.1 Preparation of competent *E. coli* cells

Competent *E. coli* cells were prepared before performing the DNA cloning. A single colony of *E. coli* TOP10 cells was picked and inoculated into 5 ml of LB broth. The cells were grown overnight at 37°C for approximately 16 h in a shaking incubator set up at 250 rpm. Then, 0.5 ml overnight culture was inoculated in a 50 ml Medium A conical flask. The culture was shaken at 37°C for 1.5 to 2 h until it reached OD_{600} 0.4-0.6. The culture was then put on ice for 10 min to stop cell growth. The cells mixture was centrifuged at 6000 rpm for 10 min and the supernatant was discarded. Three ml of Medium B were added to the pellet and pipetted up and down gently to avoid cell lysis. One hundred μl of the cell suspension was aliquoted in each 30 fresh 1.5 ml microcentrifuge tube. The aliquoted cells were then stored in -80°C freezer.

3.10.2 DNA ligation

The concentration and purity of the purified PCR product were first determined using NanoDrop 2000 (Thermo Fisher Scientific). The PCR products were then ligated overnight into the pGEM-T vector (Promega,) at 16°C following the manufacturer's protocols. Ligation reactions of 10 µl contained 5 µl of 2X Rapid Ligation Buffer, 1 µl pGEM®-T Vector, 1 µl T4 DNA Ligase and 3 µl PCR product.

3.10.3 Transformation of *E. coli* TOP10 competent cells

IPTG/ X-Gal stock solutions were prepared earlier in order to perform a blue-white screen. Twenty to twenty-five µl of 1:1 IPTG:X-Gal ratio was spread on a pre-warmed agar plate (section 3.3.3). The plate was placed and dried in the 37°C incubator. The competent cells were thawed on ice following the transformation. The 10 000 µl ligation product was pipetted into the 100 000 µl competent cells tube and stirred slowly using the pipette tip to mix the reaction. The mixture was allowed to stand for 30 min on ice. The mixture was heat-shocked at 42°C for 1 min using a heat block. The cells were then recovered on ice for 5 min and added with 900 000 µl LB broth. The transformation reaction was incubated for 1 h at 250 rpm, 37°C in the shaking incubator. The transformants were centrifuged at 8000 rpm for 1 min and approximately 100 000 µl supernatant was left in the tube. The pellet and supernatant were mixed gently. The mixture was plated on IPTG/X-Gal LB plate with ampicillin (100 µg/ml). The agar plate was incubated overnight at 37°C.

3.10.4 Colony PCR

To confirm for positive recombinant plasmid, colony PCR was performed according to the reaction volume and thermal cycling profile in section 3.4. Single white colony was picked and added into the PCR reaction mixture. The gel electrophoresis

was conducted on 1% agarose gel. Positive clones were cultured in the 5 ml LB ampicillin broth and shaken overnight at 37°C prior to plasmid extraction.

3.11 Plasmid extraction

At least two positive recombinant clones from each of transformation mixture were selected for the plasmid extraction. The QIAprep® Spin Miniprep Kit (Qiagen) was used to obtain plasmid DNA from the overnight culture. A centrifugation to separate the pellet from the supernatant was carried out at 8000 rpm for 3 min at room temperature. The pellet was resuspended in 250 µl Buffer P1 and transferred into a clean 1.5 ml microcentrifuge tube. Two hundred and fifty µl Buffer P2 was added and the tube was inverted several times until the mixture turned into blue. Then, 350 µl Buffer N3 was pipetted into the tube to neutralise the mixture until white clumps were observed. The mixture was centrifuged for 10 min at 13000 rpm to separate the plasmid DNA from the lysed cells.

The supernatant was transferred into a QIAprep spin column and centrifuged for 1 min. The flow through was removed. A washing step was performed with 750 µl Buffer PE and the column was centrifuged twice for 1 min to remove the residual wash buffer. At the elution step, 50 µl Buffer EB was added into the spin column and incubated for 1min before centrifugation to elute the purified plasmid.

3.12 DNA Sequencing

Purified recombinant plasmid was sequenced by a commercial laboratory (MyTACG Bioscience Enterprise, Malaysia). Sanger dideoxy sequencing was performed using the M13 forward (-20) and reverse (-24) universal sequencing primers. In addition, *Pk*RAP-1 IntF: 5'-ATG AGC AAA CCG TTC GTG TG-3' and *Pk*RAP-1

IntR: 5'-GTG CAT ACT GGA AAG CAT GG-3' were used for DNA sequencing to obtain the full-length *P. knowlesi* RAP-1 gene sequence.

3.13 Phylogenetic analyses

Each clone's sequence was trimmed, joined and aligned using the AliView program. The *P. knowlesi* RAP-1 sequences were obtained and aligned together with sequences of the Nuri strain (GenBank Accession No. GQ2816500, as the reference sequence), Hackeri strain (GenBank Accession No. GQ281651), Malayan strain (GenBank Accession No. GQ281648) and Philippines strain (GenBank Accession No. GQ281652). Both the nucleotide and deduced amino acid sequences were analysed using the CLUSTAL-Omega program (Sievers *et al.*, 2011). Phylogenetic tree was constructed using the neighbour-joining method implemented in MEGA6 (Tamura *et al.*, 2013). When constructing the tree, bootstrap proportions of 1,000 replicates were utilized to verify the robustness of the tree. *P. coatneyi* RAP-1 isolate (GenBank Accession No. GQ281653) was used as outgroup.

3.14 RAP-1 sequence polymorphism analysis

The number of segregating sites (S), the number of haplotypes (H), haplotype diversity (Hd) and nucleotide diversity (π) were calculated using DnaSP version 5.10.00 (Librado & Rozas, 2009). To estimate the step-wise diversity across *P. knowlesi* RAP-1, π was established on a sliding window of 100 bases, with a step size of 25 bp. The Z-test ($P < 0.05$) in MEGA6, employing the Nei and Gojobori method and the Jukes and Cantor correction, was used to estimate and compare the rates of synonymous (d_S) and non-synonymous (d_N) substitutions. d_N will be lower than d_S ($d_N/d_S < 1$) when the gene is under negative (purifying) selection, while d_N will be greater than d_S ($d_N/d_S > 1$) when the positive selection is more advantageous. Tajima's D (Tajima, 1989) and Fu and Li's

D (Fu & Li, 1993) tests were examined for evidence of departure from the predictions of the neutral theory of molecular evolution using DnaSP version 5.10.00.

CHAPTER 4: RESULTS

4.1 *P. knowlesi* RAP-1 amplification, cloning and sequencing

The RAP-1 in *P. knowlesi* contains two exons and one intron (Figure 3.1). *P. knowlesi* RAP-1 intron starts from nucleotide position 278 to 483 or 484.The total coding region consists of 2206 base pairs. The PCR amplification of the RAP-1 gene spanning a 2433 or 2434 bp fragment was successfully carried out in 34 isolates (Figure 4.1). The difference in the fragment size was due to the presence of an additional nucleotide in the intron of some of the isolates. Six of the 40 isolates' RAP-1 gene was not successfully amplified (Table 4.1). After confirming for DNA purity ($A_{260/280}$) and concentration (ng/µl), the PCR products were cloned and the recombinant clones were visible on the LB plates (Figure 4.2). The purified PCR products of 4 isolates were not successfully cloned into the pGEM-T vectors (Table 4.1). The colony PCR amplification of positive transformants *P. knowlesi* RAP-1 produced a ~2.8 kbp amplicon (Figure 4.3).

After sequencing, the sequences were trimmed to obtain the full-length *P. knowlesi* RAP-1 (2411 or 2412 bp). Thirty sequences of *P. knowlesi* RAP-1 were obtained (Table 4.2). These sequences were aligned (Figure 4.4) and analysed for the diversity and natural selection. A comparison was also made between these sequences and the *P. knowlesi* RAP-1 of old strains including Nuri, Hackeri, Malayan and Philippines (isolated in 1953, 1960, 1962, and 1961, respectively).

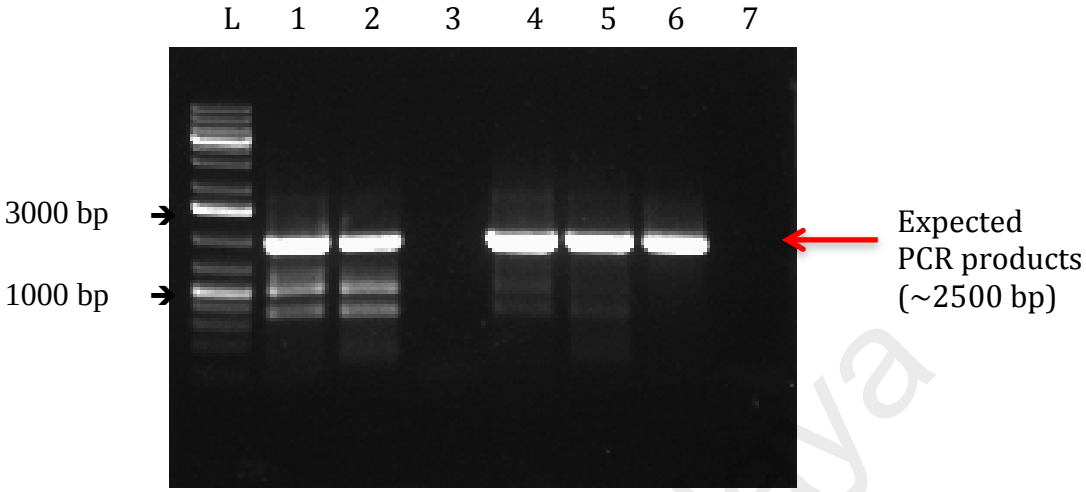

Figure 4.1: Gel picture showing PCR products of *P. knowlesi* RAP-1 gene
The amplification of *P. knowlesi* RAP-1 from 6 isolates is shown. Lane L contained DNA ladder of 1000 bp. The ~2500 bp *P. knowlesi* RAP-1 amplicons (lanes 1-2, lanes 4-7) are shown. Lanes 1, 2, 4, and 5 displayed multiple bands. Lane 3 indicates unsuccessful amplification of *P. knowlesi* RAP-1. Lane 7 represents negative template control.

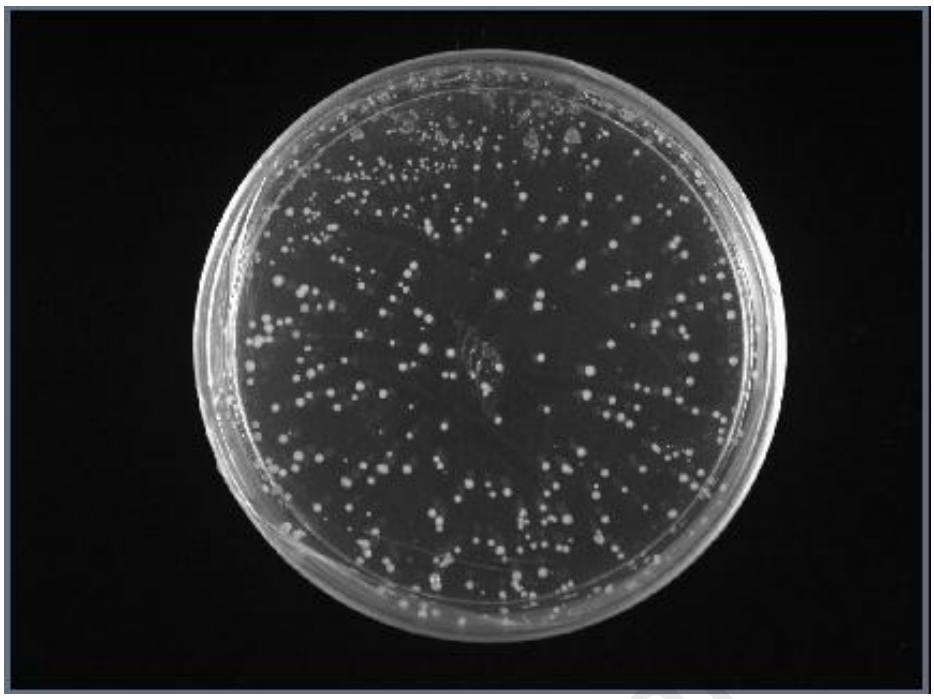

Figure 4.2: Recombinant clones of *P. knowlesi* RAP-1 on an LB plate

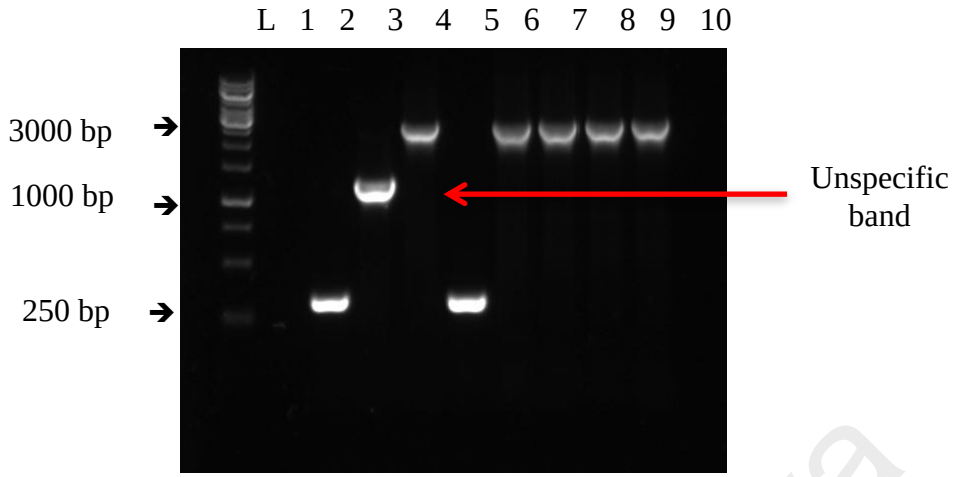

Figure 4.3: Gel of colony PCR products for the *P. knowlesi* RAP-1 gene
Eight single white colonies were picked and streaked from each LB ampicillin plate. Lane L contained DNA ladder of 1000 bp. Positive recombinant clones (lane 4, lanes 6-9) produced PCR products of 2500-3000 bp in size. PCR products from plasmids without inserts are at ~250 bp in size (lanes 2 and 5). Lane 3 showed a nonspecific band (plasmid with incorrect insert) at ~1000 bp. Lane 1 is an empty lane and lane 10 is the negative template control.

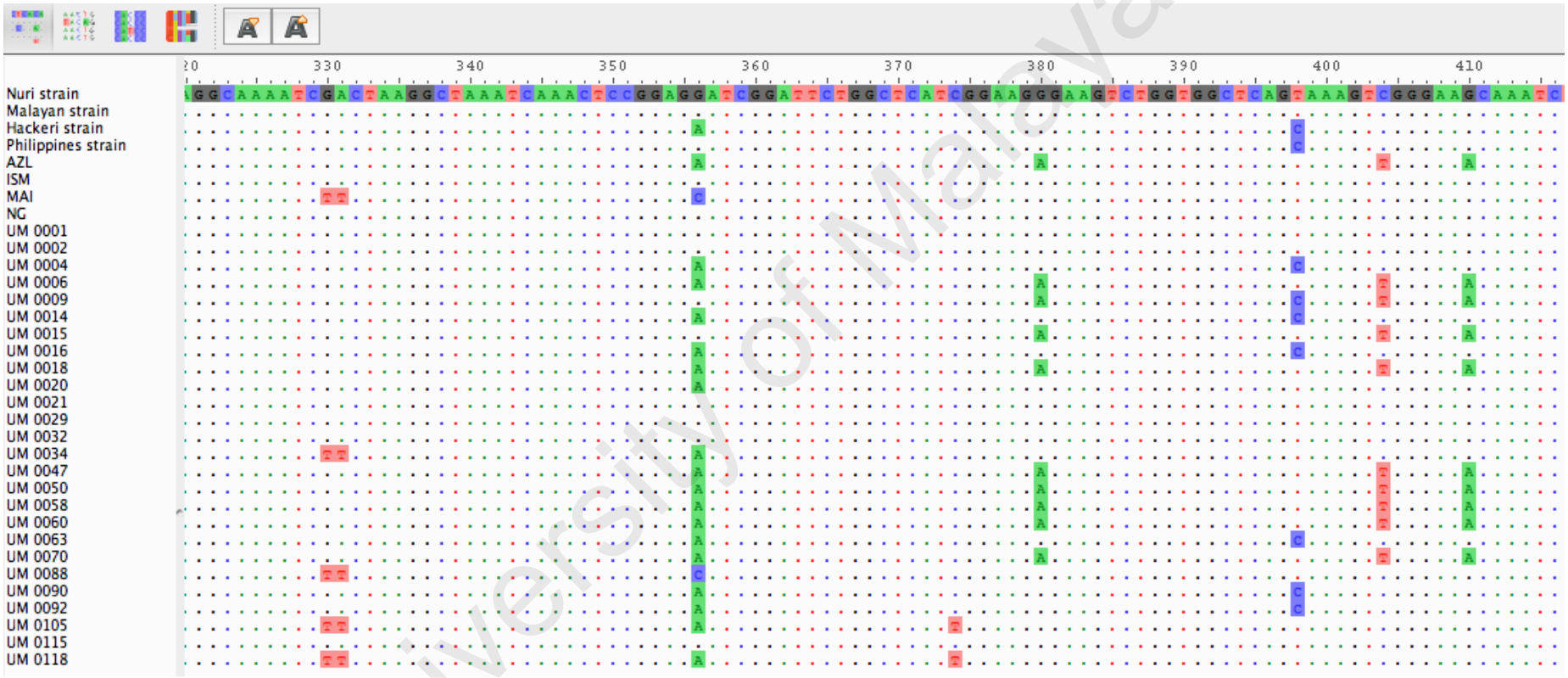

Figure 4.4: Alignment of 30 nucleotide sequences of the recent *P. knowlesi* RAP-1 isolates and 4 old strains

The sequences were aligned using the AliView program. Adenines are marked in *green*, cystosines are marked in *blue*, guanines are marked in *grey*, and thymines are marked in *red*. Nucleotides identical to those of Nuri strain are marked by dots.

Table 4.1: *P. knowlesi* RAP-1 isolates from Malaysia

No	Isolates with *P. knowlesi* infection	DNA amplification	DNA cloning
1	AZL	✓	✓
2	CHO	x	Not applicable
3	ISM	✓	✓
4	MAI	✓	✓
5	NAA	x	Not applicable
6	NEL	✓	x
7	NG	✓	✓
8	RAY	✓	x
9	ROS	x	Not applicable
10	PAU	x	Not applicable
11	WAF	x	Not applicable
12	UM 0001	✓	✓
13	UM 0002	✓	✓
14	UM 0004	✓	✓
15	UM 0006	✓	✓
16	UM 0009	✓	✓
17	UM 0014	✓	✓
18	UM 0015	✓	✓
19	UM 0016	✓	✓
20	**UM 0018**	✓	✓
21	**UM 0020**	✓	✓
22	**UM 0021**	✓	✓
23	**UM 0029**	✓	✓
24	**UM 0032**	✓	✓
25	**UM 0034**	✓	✓
26	**UM 0047**	✓	✓
27	**UM 0050**	✓	✓
28	**UM 0058**	✓	✓
29	**UM 0060**	✓	✓
30	**UM 0063**	✓	✓
31	**UM 0070**	✓	✓
32	**UM 0077**	✓	**x**
33	**UM 0079**	✓	**x**
34	**UM 0088**	✓	✓
35	**UM 0090**	✓	✓
36	**UM 0092**	✓	✓
37	**UM 0093**	**x**	**Not applicable**
38	**UM 0105**	✓	✓
39	**UM 0115**	✓	✓
40	**UM 0118**	✓	✓
	Total	**34**	**30**

Table 4.2: Genbank accession numbers of *P. knowlesi* RAP-1 sequences

Isolate	Gene Accession No.
AZL	KR259544
ISM	KR259548
MAI	KR259552
NG	KR259557
UM 0001	KR259562
UM 0002	KR259565
UM 0004	KR259567
UM 0006	KR259570
UM 0009	KR259572
UM 0014	KR259574
UM 0015	KR259578
UM 0016	KR259580
UM 0018	KR259583
UM 0020	KR259584
UM 0021	KR259587
UM 0029	KR259589
UM 0032	KR259592
UM 0034	KR259594
UM 0047	KR259597
UM 0050	KR259598
UM 0058	KR259601
UM 0060	KR259604
UM 0063	KR259608
UM 0070	KR259614
UM 0088	KR259615
UM 0090	KR259617
UM 0092	KR259620
UM 0105	KR259623
UM 0115	KR259628
UM 0118	KR259632

4.2 Nucleotide diversity and genetic differentiation

The results of the genetic diversity and neutrality tests of *P. knowlesi* RAP-1 are presented in Table 4.3. The Hd for exon 1, exon 2 and the total coding region was 0.818, 0.993 and 0.995, respectively. Additionally, the nucleotide diversity (π) of exon 1, 2 and the total coding region was 0.00915, 0.01353 and 0.01298, respectively. Higher π values were observed in exon 2 and total coding region of the recent isolates compared to the corresponding π values of the old strains (exon 2: 0.0076; total coding region: 0.0082) (Pacheco *et al.*, 2010). However, there was not much difference between the π values of exon 1 of the old strains (0.0123) and recent isolates (0.00915). Interspecies comparison (Table 4.4) showed that the nucleotide diversity of *P. knowlesi* RAP-1 was three-fold higher than of *P. falciparum* RAP-1 (Pacheco *et al.*, 2010) and 14-fold higher than of *P. vivax* RAP-1 (Garzon-Ospina *et al.*, 2010).

The sliding window plot (window length 100 bp, step size 25 bp) revealed that exon 2 contained both the highest and lowest polymorphic regions (Figure 4.5). The greatest diversity was observed within nucleotide positions 250-500 of the coding region, while the most conserved region was seen at nucleotide positions 1,800-1,950. The overall nucleotide diversity ranged from 0.003 to 0.033.

Table 4.3: Estimates of DNA diversity, selection, and neutrality tests of *P. knowlesi* RAP-1 in Malaysia

***P. knowlesi* RAP-1**	**n**	**S**	**Hd ± SD**	**π ± SD**	**d_N ± SE**	**d_S ± SE**	**d_N/d_S**	**Z-test**	**Tajima's D**	**Fu and Li's D**
Exon 1	34	276	0.818 ± 0.054	0.00915 ± 0.00089	0.00574 ± 0.00352	0.02253 ± 0.01104	0.25477	$d_N=d_S$	-0.44307 (P>0.10)	-0.47531 (P>0.10)
Exon 2	34	1929	0.993 ± 0.009	0.01353 ± 0.00102	0.00894 ± 0.00145	0.03274 ± 0.00591	0.27306	$d_N<d_S$ (P<0.05)	-0.20877 (P>0.10)	-0.22130 (P>0.10)
Total CDS	34	2205	0.995 ± 0.009	0.01298 ± 0.00091	0.00854 ± 0.00126	0.03137 ± 0.00483	0.27223	$d_N<d_S$ (P<0.05)	-0.23957 (P>0.10)	-0.26919 (P>0.10)

Note: n= number of sequences; S= number of sites; Hd= haplotype diversity; π= observed average pairwise nucleotide diversity; d_N= rate of non-synonymous substitutions per non-synonymous site; d_N= rate of synonymous substitutions per synonymous site

Table 4.4: Nucleotide diversity among the RAP-1 of *Plasmodium* species

Species	n	S	π	Reference
P. falciparum	32	2346-2349	0.0041	Pacheco *et al*., 2010
P. vivax	29	2413	0.00088	Garzon-Ospina *et al*., 2010
P. knowlesi	34	2411-2412	0.01298	Present study

Note: n= number of isolates; S= number of sites; π= nucleotide diversity

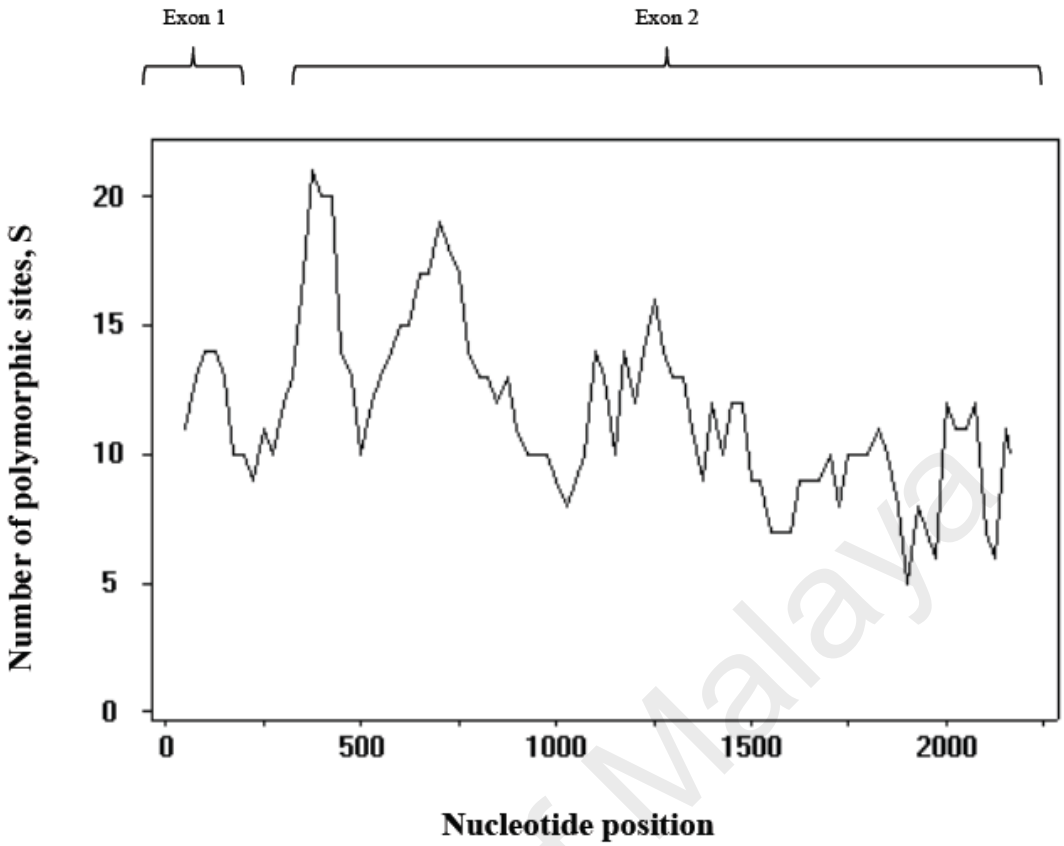

Figure 4.5: Nucleotide polymorphism of *P. knowlesi* RAP-1

Sliding window plot of number of polymorphic sites (S) in the *P.knowlesi* RAP-1 coding regions. The S values were calculated using DnaSP ver. 5.10.00 with a window length of 100 bp and a step size of 25 bp.

4.3 Amino acid changes and phylogenetic analysis

A total of 735 amino acid residues were deduced from the *P. knowlesi* RAP-1 total coding region. Using the Nuri strain sequence as reference, 61 segregating sites were identified. Singleton sites were found to be lower in frequency (23/61) than the parsimony-informative sites (38/61). From these variable sites, 54 of them were dimorphic and seven were trimorphic changes (85= R, M; 119= E, A; 140= L, S; 292= G, S; 320= S, T; 555= G, A; 682= N, Q) (Figure 4.7).

Twenty-two haplotypes were deduced from the amino acid sequences (Figure 4.6). A set of isolates or blood samples with identical amino acid sequences is grouped as one haplotype. Haplotype H7 had the highest frequency (7/34, 20.5%), followed by haplotype H3 (4/34, 11.76%), and haplotypes H2, H5 and H6 (each 2/24, 5.88%). It is interesting to note that some haplotypes consisted of old and recent isolates (Table 4.5). For instance, haplotype H2 contained the Malayan strain (1962) and isolate NG (2011). The Hackeri strain (1960) and three recent isolates (UM 0004, UM 0016 and UM 0092; isolated 2012-2013) were of haplotype H3. Phylogenetic tree analysis revealed that the haplotypes could be clustered into two main groups: A and B (Figure 4.7). Group A consisted of 19 haplotypes, whereas Group B had three haplotypes. The haplotypes (H1-H4) of the four old strains were grouped together with those of the recent isolates in Group A.

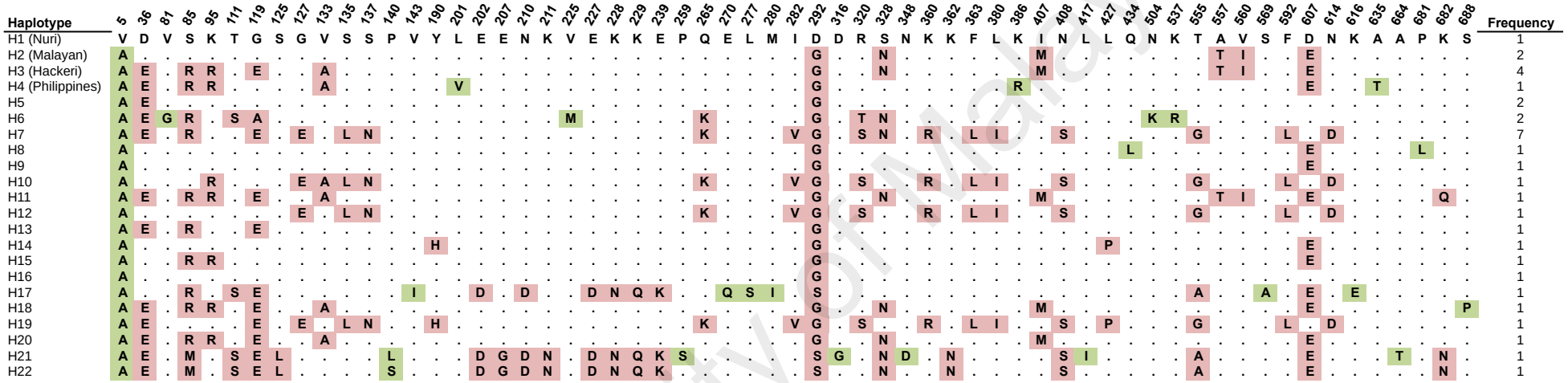

Figure 4.6: Amino acid sequence polymorphism in *P. knowlesi* RAP-1

Alignment of polymorphic amino acid residues showing 22 *P. knowlesi* RAP-1 haplotypes of the recent Malaysian *P. knowlesi* isolates and old strains. Haplotypes H1, H2, H3 and H4 are of the Nuri, Malayan, Hackeri and Philippines strains, respectively. The singleton sites are marked in green and the parsimony-informative sites are marked in red. Amino acid residues identical to those of Nuri strain are marked by dots. Frequency of each haplotype is listed in the right panel.

.

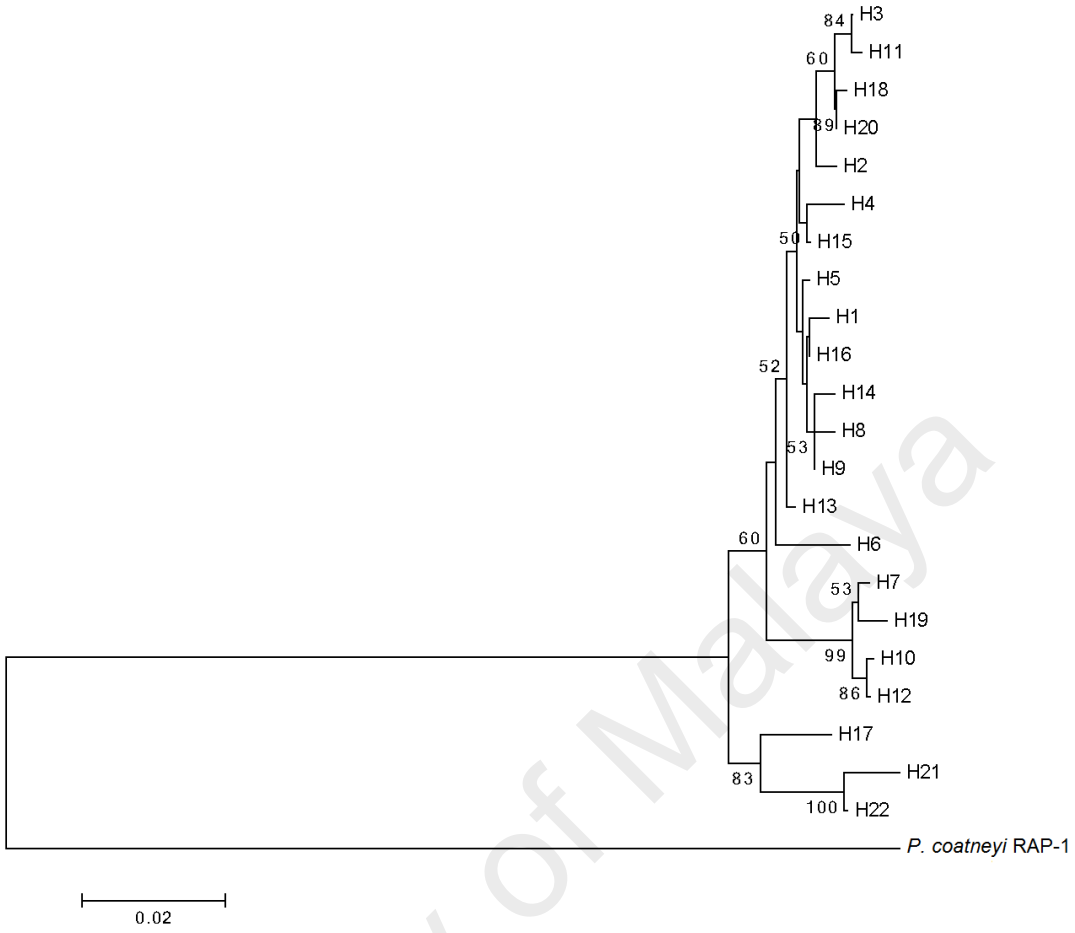

Figure 4.7: Phylogenetic tree of *P. knowlesi* RAP-1 haplotypes in Malaysia
Neighbour-joining phylogenetic tree of 22 haplotypes of *P. knowlesi* RAP-1, showing with two distinct groups, A and B. Numbers at nodes indicate percentage support of 1,000 bootstrap replicates. *Plasmodium coatneyi* RAP-1 is used as outgroup.

Table 4.5: RAP-1 haplotypes of *P. knowlesi* strains and isolates

Haplotypes	Strain/isolate (year isolated)
H1	Nuri (1953)
H2	Malayan (1962), NG (2011)
H3	Hackeri (1960), UM 0004 (2012), UM 0016 (2012), UM 0092 (2013)
H4	Philippines (1961)
H5	UM 0002 (2012), UM 0115 (2014)
H6	MAI (2010), UM 0088 (2013)
H7	AZL (2011), UM 0006 (2012), UM 0018 (2012), UM 0047 (2013), UM 0050 (2013), UM 0058 (2013), UM 0060 (2013)
H8	ISM (2011)
H9	UM 0001 (2012)
H10	UM 0009 (2012)
H11	UM 0014 (2012)
H12	UM 0015 (2012)
H13	UM 0020 (2012)
H14	UM 0021 (2012)
H15	UM 0029 (2012)
H16	UM 0032 (2012)
H17	UM 0034 (2012)
H18	UM 0063 (2013)
H19	UM 0070 (2013)
H20	UM 0090 (2013)
H21	UM 0105 (2014)
H22	UM 0118 (2014)

4.4 Natural selection in the *P. knowlesi* RAP-1 gene

A significant excess of synonymous substitutions was seen in the *P. knowlesi* RAP-1. The calculated ratios d_N/d_S for exon 1, exon 2 and total coding region were less than 1 (Table 4.1). This was indicative of negative selection of *P. knowlesi* RAP-1. Detailed analysis using the Z-test revealed negative selection in exon 2, but neutral selection in exon 1. In the Tajima's D and Fu and Li's D tests, all values obtained for *P. knowlesi* RAP-1 were negative, but did not differ statistically ($P>0.10$) significantly from zero. Therefore, Tajima's D and related statistics did not detect departure from neutrality.

CHAPTER 5: DISCUSSION

5.1 *P. knowlesi* RAP-1 amplification, cloning and sequencing

The primer design and amplification of the RAP-1 gene in *P. knowlesi* was partly challenging due to the length of the gene. A set of degenerate primers was designed initially according to the previous studies of RAP-1 in non-human primate malaria species (Pacheco *et al*., 2010). These primers were intended to amplify DNA regions of the orthologue in which codons code for more than one amino acid. However, these degenerate primers nonspecifically amplified the DNA of *P. knowlesi* RAP-1. This nonspecific priming was resolved by designing the *Pk*RAP-1F and *Pk*RAP-1R primers that specifically annealed to the DNA template with the annealing temperature of more than 60°C.

A short-fragment amplification generally requires only *Taq* polymerase, a thermostable DNA polymerase that lacks of 3' to 5' exonuclease proofreading activity. Commercial *Taq* polymerases from several producers did not amplify the *P. knowlesi* RAP-1 gene. Non-specific bands were observed on the gel. The use of *Pfu* DNA polymerase with proofreading activity was able to amplify ~2.5 kbp fragment. However, *Pfu* DNA polymerase does not generate A (Adenine) overhangs at the 3' ends of the PCR products, therefore, hinders ligation into the pGEM-T vector, which has T overhangs The use of high fidelity DNA polymerase GoTaq® Long PCR Mastermix solved this problem by providing the A overhangs for a proficient cloning and having proofreading activity for an efficient long DNA amplification in the PCR.

The unsuccessful amplification of the 6 isolates' RAP-1 gene could be due to the instability of the DNA samples. The DNA samples, which were extracted a few years

ago, might have degraded as the samples were thawed and used for a number of times. A fluctuation in temperature from thawing the samples would lead to a fragmentation of DNA. As a result, the gene could not be amplified from the genomic DNA. Apart from that, the DNA degradation would cause in low yield of amplicons, which could result in unsuccessful cloning of the PCR products into the vectors.

5.2 Pyhlogenetic analyses of *P. knowlesi* RAP-1

A study has been carried out previously on the diversity and natural selection of *P. knowlesi* RAP-1, albeit using a small sample size (n=5) of old *P. knowlesi* strains (Pacheco *et al.*, 2010). The present study was carried out using the same approach, but using a larger sample size (n=30) consisting of recent isolates. Unlike the findings on the old strains [π: 0.0082 (total coding region), 0.0123 (exon 1), 0.0076 (exon 2)], the present study found relatively higher diversity among the *P. knowlesi* RAP-1 of the recent isolates [π: 0.01298 (total coding region)], and diversity was much higher in exon 2 (π: 0.01353) than in exon 1 in exon 1(π: 0.00915). However, both the old strains and recent isolates showed negative selection in exon 2 and neutral selection in exon 1. The *P. knowlesi* RAP-1 (π: 0.01298) was observed to be relatively more diverse than *P. falciparum* RAP-1 (π: 0.0041) (Pacheco *et al.*, 2010) and *P. vivax* RAP-1 (π: 0.00088) (Garzon-Ospina *et al.*, 2010). A similar finding was reported for rhoptry bulb proteins (Kats *et al.*, 2008). It has been suggested that such contrasting level of polymorphism in rhoptry-related proteins is expected because these proteins are distinct across the *Plasmodium* species, presumably for adaptation in their respective target host cells (Kats *et al.*, 2008).

Merozoite surface protein-8 (MSP-8), MSP-9, apical membrane antigen-1 (AMA-1) and Duffy binding protein (DBPαII) are among the widely studied proteins

known to be potential vaccine candidates. For *P. knowles*i, the MSP-8 (Pacheco *et al.*, 2012), MSP-9 (Chenet *et al.*, 2013) and AMA-1 (Faber *et al.*, 2015) expressed lower genetic diversity (π: 0.0008 and 0.00501, respectively) than *P. knowlesi* RAP-1. *P. knowlesi* DBPαII (π: 0.013) (Fong *et al.*, 2014), however, has almost similar diversity level with *P. knowlesi* RAP-1. Similar to *P. knowlesi* RAP-1, these proteins also appear to be under negative selection.

The sliding window plot analysis showed that *P. knowlesi* RAP-1 was more conserved at the C-terminal region. This is most likely due to the role of this region in a key binding activity. The RAP-1 is known to bind to RAP-2 or RAP-3 via its C-terminal region (Baldi *et al.*, 2000). Furthermore, deletion of the RAP-1 C-terminus leads to RAP-1 mislocalization to the rhoptry neck instead of the bulb (Counihan *et al.*, 2013), suggesting the importance of this region in protein targeting. In contrast, the N-terminal of *P. knowlesi* RAP-1 exhibited genetic diversity and this may be due to the presence of T-cell epitopes. It has been observed that lymphocytes gave response to the N-terminus of *P. falciparum* RAP-1 (Stowers *et al.*, 1997; Fonjungo *et al.*, 1998).

Many of the malaria parasite blood stage antigens, such as the merozoite surface proteins, display polymorphism as a result of positive selection (Weedall & Conway, 2010). This is said to be an escape mechanism for the parasite to evade the immune responses of the host. Antigenic polymorphism involving the expression of different alleles of the gene would hamper the host's immune system to recognize the protein (Escalante *et al.*, 2004). Immune defences, such as antibodies and T cells, will not be able to identify antigenically different epitopes, and these mutated alleles will then be selectively expanded. Negative selection usually minimizes genetic variants, therefore leading to low frequency rare alleles in the population. Low frequency rare haplotypes were evident among the *P. knowlesi* RAP-1 in the present study (Figure 4.7).

Interestingly, negative selection is also seen in the RAP-1 gene of several non-human primate malarial parasites such as *P. cynomolgi, P. inui* and *P. fieldi* but not in human parasites such as *P. falciparum* and *P. vivax* (Pacheco *et al.,* 2010). For *P. knowles*i, this negative selection may be due to a bottleneck event that drives population expansion or growth. Mitochondrial DNA analysis have shown that *P. knowlesi* in Southeast Asia underwent significant population expansion approximately 30,000-40,000 years ago (Lee e*t al.*, 2011). An alternative explanation for the negative selection is that *P. knowlesi* RAP-1, being an important protein in erythrocyte invasion, has functional constraints that limit polymorphism, and any variant form of *P. knowlesi* RAP-1 will be disadvantageous to the parasite.

The phylogenetic tree in this present study also showed separation of the *P. knowlesi* RAP-1 haplotypes into two groups (Figure 4.8). This separation of *P. knowlesi* RAP-1 haplotypes groups may indicate dimorphism of the gene. Similar observations have been reported in *P. knowlesi* genes such as *P. knowlesi* DBPαII (Fong *et al.,* 2014), *P. knowlesi* NBPXα (Pinheiro *et al.,* 2015), *P. knowlesi* AMA-1 domain I (Fong e*t al.,* 2015) and *P. knowlesi* MSP-1 (Putaporntip *et al.,* 2013). These findings provide support to the postulation of the existence of two distinct *P. knowlesi* types or lineages in Southeast Asia (Muehlenbein *et al.,* 2015). Microsatellite genotyping data revealed admixture of two highly divergent *P. knowlesi* populations, and each population is associated with different forest-dwelling macaque reservoir host species (Divis e*t al*., 2015). Recently, a whole-genome population study showed two major sub-groups of *P. knowlesi* clinical isolates (Assefa *et al.*, 2015)

5.3 The importance of *P. knowlesi* RAP-1

The present study has shown evidence of negative selection. This finding concurs with the findings of a previous study of *P. knowlesi* RAP-1 of the old isolates (Pacheco *et al.,* 2010). The relatively low polymorphism level and negative selection of *P. knowlesi* RAP-1 is a good indicator of the potential of this protein as a vaccine for knowlesi malaria.

An effective vaccine candidate should be accessible to the host's immune system and able to induce protection (Richie & Saul, 2002). The RAP-1 protein has been shown to be capable of inducing immune responses (Harnyuttanakorn *et al*., 1992; Schofield *et al*., 1986) and providing protection against *Plasmodium* infection (Baldi *et al*., 2000; Collins *et al*., 2000; Perrin *et al*., 1985). Future studies on the immunogenicity of *P. knowlesi* RAP-1 are crucial to determine the protective capability of this protein as a vaccine agent for knowlesi malaria.

CHAPTER 6: CONCLUSION

The present study found higher genetic polymorphism in the *P. knowlesi* RAP-1 gene than the polymorphism level reported in the previous study. This observation may stem from the difference in sample size between the present (n=30) and the previous (n=5) study. Synonymous and nonsynonymous mutation analysis indicated purifying (negative) selection of the gene. Change of hosts from macaques to humans has lead to a bottleneck event that drives population expansion or growth of *P. knowlesi* species. The separation of *P. knowlesi* RAP-1 haplotypes into two groups is further evidence to the existence of two distinct *P. knowlesi* types or lineages, indicating dimorphism of the gene.

Despite the limitations faced during the amplification and cloning of the samples, 30 RAP-1 gene sequences were successfully obtained by utilizing the high fidelity enzyme and designing new primers. The use of next-generation sequencing (NGS) rather than single gene sequencing would be advantageous. It enables whole genome sequencing, which does not require cloning that could save time and cost. Further characterization of this protein is crucial to obtain insight into amino acid conservation and protein folding. Also, analyses on more Malaysian Borneo's clinical isolates are important to further determine phenomenon of geographical separation of haplotypes.

REFERENCES

Anders, R. F., & Saul, A. J. (1994). Candidate antigens for an asexual blood stage vaccine against falciparum malaria. In M. F. Good & A. J. Saul (Eds.), *Molecular Immunological Considerations in Malaria Vaccine Development* (pp. 169-208). Boca Raton, Florida,: CRC Press.

Baldi, D. L., Andrews, K. T., Waller, R. F., Roos, D. S., Howard, R. F., Crabb, B. S., & Cowman, A. F. (2000). RAP1 controls rhoptry targeting of RAP2 in the malaria parasite *Plasmodium falciparum*. *EMBO Journal, 19*(11), 2435-2443.

Baum, J., Gilberger, T. W., Frischknecht, F., & Meissner, M. (2008). Host-cell invasion by malaria parasites: insights from *Plasmodium* and *Toxoplasma*. *Trends in Parasitology, 24*(12), 557-563.

Baxevanis, A. D., & Ouellette, B. F. (2004). *Bioinformatics: a practical guide to the analysis of genes and proteins* (Vol. 43). John Wiley & Sons.

Boothroyd, J. C., & Dubremetz, J. F. (2008). Kiss and spit: the dual roles of *Toxoplasma* rhoptries. *Nature Reviews Microbiology, 6*(1), 79-88.

Bozdech, Z., Llinas, M., Pulliam, B. L., Wong, E. D., Zhu, J., & DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biology, 1*(1), E5.

Calderaro, A., Piccolo, G., Gorrini, C., Rossi, S., Montecchini, S., Dell'Anna, M. L., . . . Arcangeletti, M. C. (2013). Accurate identification of the six human *Plasmodium* spp. causing imported malaria, including *Plasmodium ovale wallikeri* and *Plasmodium knowlesi*. *Malaria journal*, *12*(1), 1.

Centers for Disease Control and Prevention (2015). *Malaria.* Retrieved 11 November, 2015, from http://www.cdc.gov/malaria/

Chenet, S. M., Pacheco, M. A., Bacon, D. J., Collins, W. E., Barnwell, J. W., & Escalante, A. A. (2013). The evolution and diversity of a low complexity vaccine candidate, merozoite surface protein 9 (MSP-9), in *Plasmodium vivax* and closely related species. *Infection, Genetics, and Evolution, 20*, 239-248.

Chin, W., Contacos, P. G., Coatney, G. R., & Kimball, H. R. (1965). A Naturally Acquited Quotidian-Type Malaria in Man Transferable to Monkeys. *Science, 149*(3686), 865.

Coatney, G. R. (1968). Simian malarias in man: facts, implications, and predictions. *American Journal of Tropical Medicine and Hygiene, 17*(2), 147-155.

Cole-Tobian, J., & King, C. L. (2003). Diversity and natural selection in *Plasmodium vivax* Duffy binding protein gene. *Molecular and Biochemical Parasitology, 127*(2), 121-132.

Collins, W. E., Walduck, A., Sullivan, J. S., Andrews, K., Stowers, A., Morris, C. L., . . . Saul, A. (2000). Efficacy of vaccines containing rhoptry-associated proteins RAP1 and RAP2 of *Plasmodium falciparum* in *Saimiri boliviensis* monkeys. *American Journal of Tropical Medicine and Hygiene, 62*(4), 466-479.

Cooke, G. S., & Hill, A. V. (2001). Genetics of susceptibility to human infectious disease. *Nature Reviews Genetics, 2*(12), 967-977.

Coppens, I., & Joiner, K. A. (2003). Host but not parasite cholesterol controls *Toxoplasma* cell entry by modulating organelle discharge. *Molecular Biology of the Cell, 14*(9), 3804-3820.

Counihan, N. A., Kalanon, M., Coppel, R. L., & de Koning-Ward, T. F. (2013). *Plasmodium* rhoptry proteins: why order is important. *Trends in Parasitology, 29*(5), 228-236.

Cowman, A. F., Berry, D., & Baum, J. (2012). The cellular and molecular basis for malaria parasite invasion of the human red blood cell. The *Journal of Cell Biology, 198*(6), 961-971.

Cowman, A. F., & Crabb, B. S. (2006). Invasion of red blood cells by malaria parasites. *Cell, 124*(4), 755-766.

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites and Vectors, 3*(1), 5.

Cox-Singh, J., Davis, T. M., Lee, K. S., Shamsul, S. S., Matusop, A., Ratnam, S., . . . Singh, B. (2008). *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clinical Infectious Diseases, 46*(2), 165-171.

Cui, L., Escalante, A. A., Imwong, M., & Snounou, G. (2003). The genetic diversity of *Plasmodium vivax* populations. *Trends in Parasitology, 19*(5), 220-226.

Escalante, A. A., Cornejo, O. E., Rojas, A., Udhayakumar, V., & Lal, A. A. (2004). Assessing the effect of natural selection in malaria parasites. *Trends in Parasitology, 20*(8), 388-395.

Escalante, A. A., Grebert, H. M., Chaiyaroj, S. C., Magris, M., Biswas, S., Nahlen, B. L., & Lal, A. A. (2001). Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology, 113*(2), 279-287.

Escalante, A. A., Lal, A. A., & Ayala, F. J. (1998). Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics, 149*(1), 189-202.

Faber, B. W., Kadir, K. A., Rodriguez-Garcia, R., Remarque, E. J., Saul, F. A., Vulliez-Le Normand, B., . . . Singh, B. (2015). Low levels of polymorphisms and no evidence for diversifying selection on the *Plasmodium knowlesi* apical membrane antigen 1 gene. *PLoS One, 10*(4).

Figtree, M., Lee, R., Bain, L., Kennedy, T., Mackertich, S., Urban, M., . . . Hudson, B. J. (2010). *Plasmodium knowlesi* in human, Indonesian Borneo. *Emerging Infectious Diseases, 16*(4), 672-674.

Fong, M. Y., Lau, Y. L., Chang, P. Y., & Anthony, C. N. (2014). Genetic diversity, haplotypes and allele groups of Duffy binding protein (PkDBPalphaII) of *Plasmodium knowlesi* clinical isolates from Peninsular Malaysia. *Parasites and Vectors, 7*, 161.

Fong, M. Y., Rashdi, S. A., Yusof, R., & Lau, Y. L. (2015). Distinct genetic difference between the Duffy binding protein (PkDBPalphaII) of *Plasmodium knowlesi* clinical isolates from North Borneo and Peninsular Malaysia. *Malaria Journal, 14*, 91.

Fonjungo, P. N., Stuber, D., & McBride, J. S. (1998). Antigenicity of recombinant proteins derived from rhoptry-associated protein 1 of *Plasmodium falciparum*. *Infection and Immunity, 66*(3), 1037-1044.

Fu, Y. X., & Li, W. H. (1993). Statistical Tests of Neutrality of Mutations. *Genetics, 133*(3), 693-709.

Garzon-Ospina, D., Romero-Murillo, L., & Patarroyo, M. A. (2010). Limited genetic polymorphism of the *Plasmodium vivax* low molecular weight rhoptry protein complex in the Colombian population. *Infection, Genetics, and Evolution, 10*(2), 261-267.

Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. New York: Oxford University Press.

Harnyuttanakorn, P., McBride, J. S., Donachie, S., Heidrich, H. G., & Ridley, R. G. (1992). Inhibitory monoclonal antibodies recognise epitopes adjacent to a proteolytic cleavage site on the RAP-1 protein of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology, 55*(1-2), 177-186.

Hisaeda, H., Yasutomo, K., & Himeno, K. (2005). Malaria: immune evasion by parasites. *International Journal of Biochemistry and Cell Biology, 37*(4), 700-706.

Holder, A. A. (1994). Proteins on the surface of the malaria parasite and cell invasion. *Parasitology, 108*(S1), S5-S18.

Howard, R. F. (1992). The sequence of the p82 rhoptry protein is highly conserved between two *Plasmodium falciparum* isolates. *Molecular and Biochemical Parasitology, 51*(2), 327-330.

Howard, R. F., & Peterson, C. (1996). Limited RAP-1 sequence diversity in field isolates of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology, 77*(1), 95-98.

Howard, R. F., Jensen, J. B., & Franklin, H. L. (1993). Reactivity profile of human anti-82-kilodalton rhoptry protein antibodies generated during natural infection with *Plasmodium falciparum*. *Infection and Immunity, 61*(7), 2960-2965.

Howard, R. F., Narum, D. L., Blackman, M., & Thurman, J. (1998). Analysis of the processing of *Plasmodium falciparum* rhoptry-associated protein 1 and localization of Pr86 to schizont rhoptries and p67 to free merozoites. *Molecular and Biochemical Parasitology, 92*(1), 111-122.

Howard, R. F., & Schmidt, C. M. (1995). The secretary pathway of *Plasmodium falciparum* regulates transport of p82/RAP1 to the rhoptries. *Molecular and Biochemical Parasitology, 74*(1), 43-54.

Hughes, M. K., & Hughes, A. L. (1995). Natural selection on *Plasmodium* surface proteins. *Molecular and Biochemical Parasitology, 71*(1), 99-113.

Jakobsen, P. H., Moon, R., Ridley, R. G., Bate, C. A., Taverne, J., Hansen, M. B., . . . McBride, J. S. (1993). Tumour necrosis factor and interleukin-6 production induced by components associated with merozoite proteins of *Plasmodium falciparum*. *Parasite Immunology, 15*(4), 229-237.

Jeslyn, W. P., Huat, T. C., Vernon, L., Irene, L. M., Sung, L. K., Jarrod, L. P., . . . Ching, N. L. (2011). Molecular epidemiological investigation of *Plasmodium knowlesi* in humans and macaques in Singapore. *Vector-Borne and Zoonotic Diseases, 11*(2), 131-135.

Jiang, N., Chang, Q., Sun, X., Lu, H., Yin, J., Zhang, Z., . . . Chen, Q. (2010). Co-infections with *Plasmodium knowlesi* and other malaria parasites, Myanmar. *Emerging Infectious Diseases Journal, 16*(9), 1476-1478.

Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Sata, T., & Kanbara, H. (2004). Naturally acquired *Plasmodium knowlesi* malaria in human, Thailand. *Emerging Infectious Diseases, 10*(12), 2211-2213.

Kats, L. M., Cooke, B. M., Coppel, R. L., & Black, C. G. (2008). Protein trafficking to apical organelles of malaria parasites - building an invasion machine. *Traffic, 9*(2), 176-186.

Kats, L. M., Black, C. G., Proellocks, N. I., & Coppel, R. L. (2006). *Plasmodium* rhoptries: how things went pear-shaped. *Trends in Parasitology, 22*(6), 269-276.

Khim, N., Siv, S., Kim, S., Mueller, T., Fleischmann, E., Singh, B., . . . Menard, D. (2011). *Plasmodium knowlesi* infection in humans, Cambodia, 2007-2010. *Emerging Infectious Diseases, 17*(10), 1900-1902.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature, 267*(5608), 275-276.

Kimura, M. (1984). *The neutral theory of molecular evolution*. Cambridge University Press.

Knowles, R., & Gupta, D. (1933). BM (1932) A study of monkey-malaria, and its experimental transmission to man. *Indian Medical Gazette, 67*, 301-320.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948.

Lee, C. E., Adeeba, K., & Freigang, G. (2010). Human *Plasmodium knowlesi* infections in Klang Valley, Peninsular Malaysia: a case series. *Medical Journal of Malaysia, 65*(1), 63-65.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*(11), 1451-1452.

Liddell, H. G., Scott, R., & Jones, H. S. (1968). *A Greek-English lexicon*. Oxford: Clerendon Press.

Lobo, C. A., Rodriguez, M., Hou, G., Perkins, M., Oskov, Y., & Lustigman, S. (2003). Characterization of PfRhop148, a novel rhoptry protein of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology, 128*(1), 59-65.

Loewe, L. (2008). Negative selection. *Nature Education, 1*(1), 59.

Luchavez, J., Espino, F., Curameng, P., Espina, R., Bell, D., Chiodini, P., . . . Singh, B. (2008). Human Infections with *Plasmodium knowlesi*, the Philippines. *Emerging Infectious Diseases, 14*(5), 811-813.

Margos, G., Bannister, L. H., Dluzewski, A. R., Hopkins, J., Williams, I. T., & Mitchell, G. H. (2004). Correlation of structural development and differential expression of invasion-related molecules in schizonts of *Plasmodium falciparum*. *Parasitology, 129*(03), 273-287.

McCutchan, T. F., de la Cruz, V. F., Good, M. F., & Wellems, T. E. (1988). Antigenic diversity in *Plasmodium falciparum*. *Progress in Allergy, 41*, 173-192.

Mcdonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature, 351*(6328), 652-654.

Mitchell, G. H., Thomas, A. W., Margos, G., Dluzewski, A. R., & Bannister, L. H. (2004). Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells. *Immunity, 72*(1), 154-158.

Moon, S. U., Na, B. K., Kang, J. M., Kim, J. Y., Cho, S. H., Park, Y. K., . . . Kim, T. S. (2010). Genetic polymorphism and effect of natural selection at domain I of apical membrane antigen-1 (AMA-1) in *Plasmodium vivax* isolates from Myanmar. *Acta Tropica, 114*(2), 71-75.

Muehlenbein, M. P., Pacheco, M. A., Taylor, J. E., Prall, S. P., Ambu, L., Nathan, S., . . . Escalante, A. A. (2015). Accelerated diversification of nonhuman primate malarias in Southeast Asia: adaptive radiation or geographic speciation?. *Molecular Biology and Evolution, 32*(2), 422-439.

Napier, L. E., & Campbell, H. G. M. (1932). Observations on a *Plasmodium* infection which causes haemaglobinuria in certain species of monkey. *Indian Medical Gazette, 67*, 246-249.

Ng, O. T., Ooi, E. E., Lee, C. C., Lee, P. J., Ng, L. C., Pei, S. W., . . . Leo, Y. S. (2008). Naturally acquired human *Plasmodium knowlesi* infection, Singapore. *Emerging Infectious Diseases, 14*(5), 814-816.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics, 23*, 263-286.

Ohta, T. (1996). The neutralist-selectionist debate. *Bioessays, 18*, 673-677.

Pacheco, M. A., Ryan, E. M., Poe, A. C., Basco, L., Udhayakumar, V., Collins, W. E., & Escalante, A. A. (2010). Evidence for negative selection on the gene encoding rhoptry-associated protein 1 (RAP-1) in *Plasmodium* spp. *Infection, Genetics and Evolution, 10*(5), 655-661.

Pancer, Z., & Cooper, M. D. (2006). The evolution of adaptive immunity. *Annual Review of Immunology, 24*, 497-518.

Perez-Leal, O., Mongui, A., Cortes, J., Yepes, G., Leiton, J., & Patarroyo, M. A. (2006). The *Plasmodium vivax* rhoptry-associated protein 1. *Biochemical and Biophysical Research Communications, 341*(4), 1053-1058.

Perrin, L. H., Merkli, B., Gabra, M. S., Stocker, J. W., Chizzolini, C., & Richle, R. (1985). Immunization with a *Plasmodium falciparum* merozoite surface antigen induces a partial immunity in monkeys. *Journal of Clinical Investigation, 75*(5), 1718-1721.

Piotrowski, J. (2014). *Monkey malaria on the rise among humans in Malaysia*. Retrieved 20 November 2015, from http://www.scidev.net/global/malaria/news/monkey-malaria-rise-humans-malaysia.html

Polley, S. D., & Conway, D. J. (2001). Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics, 158*(4), 1505-1512.

Preiser, P., Kaviratne, M., Khan, S., Bannister, L., & Jarra, W. (2000). The apical organelles of malaria merozoites: host cell selection, invasion, host immunity and immune evasion. *Microbes and Infection, 2*(12), 1461-1477.

Prudencio, M., Rodriguez, A., & Mota, M. M. (2006). The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nature Reviews Microbiology, 4*(11), 849-856.

Richard, D., Kats, L. M., Langer, C., Black, C. G., Mitri, K., Boddey, J. A., . . . Coppel, R. L. (2009). Identification of rhoptry trafficking determinants and evidence for a novel sorting mechanism in the malaria parasite *Plasmodium falciparum*. *PLoS Pathogen, 5*(3), e1000328.

Richie, T. L., & Saul, A. (2002). Progress and challenges for malaria vaccines. *Nature, 415*(6872), 694-701.

Ridley, R. G., Takacs, B., Etlinger, H., & Scaife, J. G. (1990). A rhoptry antigen of Plasmodium falciparum is protective in *Saimiri* monkeys. *Parasitology, 101*(02), 187-192.

Saitou, N. (1996). Reconstruction of gene trees from sequence data. *Methods in Enzymology, 266*, 427-449.

Sam-Yellowe, T. Y., Shio, H., & Perkins, M. E. (1988). Secretion of *Plasmodium falciparum* rhoptry protein into the plasma membrane of host erythrocytes. The *Journal of Cell Biology, 106*(5), 1507-1513.

Schofield, L., Bushell, G. R., Cooper, J. A., Saul, A. J., Upcroft, J. A., & Kidson, C. (1986). A rhoptry antigen of *Plasmodium falciparum* contains conserved and variable epitopes recognized by inhibitory monoclonal antibodies. *Molecular and Biochemical Parasitology, 18*(2), 183-195.

Sermwittayawong, N., Singh, B., Nishibuchi, M., Sawangjaroen, N., & Vuddhakul, V. (2012). Human *Plasmodium knowlesi* infection in Ranong province, southwestern border of Thailand. *Malaria Journal, 11*, 36.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*, 539.

Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S. S., Cox-Singh, J., . . . Conway, D. J. (2004). A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet, 363*(9414), 1017-1024.

Sinton, J. A., & Mulligan, H. W. (1933). A critical review of the literature relating to the identification of the malarial parasites recorded from monkeys of the families *Circopithecidae* and *Colobidae. Records of the Malaria Survey of India III*, 357-380.

Stowers, A., Taylor, D., Prescott, N., Cheng, Q., Cooper, J., & Saul, A. (1997). Assessment of the humoral immune response against *Plasmodium falciparum* rhoptry-associated proteins 1 and 2. *Infection and Immunity, 65*(6), 2329-2338.

Sterkers, Y., Scheidig, C., da Rocha, M., Lepolard, C., Gysin, J., & Scherf, A. (2007). Members of the low-molecular-mass rhoptry protein complex of *Plasmodium falciparum* bind to the surface of normal erythrocytes. *The Journal of Infectious Diseases, 196*(4), 617-621.

Sulistyaningsih, E., Fitri, L. E., Loscher, T., & Berens-Riha, N. (2010). Diagnostic difficulties with *Plasmodium knowlesi* infection in humans. *Emerging Infectious Diseases, 16*(6), 1033-1034.

Tajima, F. (1989). Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution, 30*(12), 2725-2729.

Tetteh, K. K., Stewart, L. B., Ochola, L. I., Amambua-Ngwa, A., Thomas, A. W., Marsh, K., . . . Conway, D. J. (2009). Prospective identification of malaria parasite genes under balancing selection. *PLoS One, 4*(5), e5568.

Topolska, A. E., Lidgett, A., Truman, D., Fujioka, H., & Coppel, R. L. (2004). Characterization of a membrane-associated rhoptry protein of *Plasmodium falciparum*. *The Journal of Biological Chemistry, 279*(6), 4648-4656.

Van den Eede, P., Van, H. N., Van Overmeir, C., Vythilingam, I., Duc, T. N., Hung le, X., . . . Erhart, A. (2009). Human *Plasmodium knowlesi* infections in young children in central Vietnam. *Malaria Journal, 8*, 249.

Vlachou, D., Schlegelmilch, T., Runn, E., Mendes, A., & Kafatos, F. C. (2006). The developmental migration of *Plasmodium* in mosquitoes. *Current Opinion in Genetics and Development, 16*(4), 384-391.

Vythilingam, I. (2010). *Plasmodium knowlesi* in humans: a review on the role of its vectors in Malaysia. *Tropical Biomedicine, 27*(1), 1-12.

Vythilingam, I., Noorazian, Y. M., Huat, T. C., Jiram, A. I., Yusri, Y. M., Azahari, A. H., . . . Lokmanhakim, S. (2008). *Plasmodium knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. *Parasites and Vectors, 1*(1), 26.

Warren, M., Cheong, W. H., Fredericks, H. K., & Coatney, G. R. (1970). Cycles of jungle malaria in West Malaysia. *American Journal of Tropical Medicine and Hygiene, 19*(3), 383-393.

White, N. J. (2008). *Plasmodium knowlesi*: the fifth human malaria parasite. *Clinical Infectious Diseases, 46*(2), 172-173.

Wiser, M. F. (1999). Cellular and molecular biology of *Plasmodium*, from http://www.tulane.edu/~wiser/malaria/cmb.html

World Health Organization (2015). World Malaria Report. Geneva. Switzerland: World Health Organization.

Zhu, H. M., Li, J., & Zheng, H. (2006). [Human natural infection of *Plasmodium knowlesi*]. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi, 24*(1), 70-71.

LIST OF PUBLICATION

Rawa, M. S., Fong, M. Y., Lau, Y. L. (2016). Genetic diversity and natural selection in the rhoptry-associated protein 1 (RAP-1) of recent *Plasmodium knowlesi* clinical isolates from Malaysia. *Malaria Journal*, *15*(1), 62.