DIVERSITY OF MEROZOITE SURFACE PROTEIN-1, 2, 3, 4 and 5 (*msp*-1, 2, 3, 4 and 5) GENES AMONG *Plasmodium falciparum* COLLECTED FROM BORDER AREAS OF THAILAND

Mr. Phumin Simpalipan

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Zoology
Department of Biology
Faculty of Science
Chulalongkorn University
Academic Year 2010
Copyright of Chulalongkorn University

ความหลากหลายของยีนเมโรซอยต์เซอร์เฟซโปรตีน 1, 2, 3, 4 และ 5

(*msp*-1, 2, 3, 4 และ 5) ของเชื้ อมาลาเรีย*Plasmodium falciparum*

ที่เก็บจากพื้ นที่ชายแดนของประเทศไทย

นายภูมินทร์ สิมพลิพันธ์

วิทยานิพนธ์นี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสัตววิทยา ภาควิชาชีววิทยา

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	DIVERSITY OF MEROZOITE SURFACE PROTEIN-1, 2, 3, 4 and 5 (*msp*-1, 2, 3, 4 and 5) GENES AMONG *Plasmodium falciparum* COLLECTED FROM BORDER AREAS OF THAILAND
By	Mr. Phumin Simpalipan
Field of Study	Zoology
Thesis Advisor	Assistant Professor Pongchai Harnyuttanakorn, Ph.D.
Thesis Co-advisor	Naowarat Kanchanakhan, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.... Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr. rer. nat.)

THESIS COMMITTEE

.... Chairman
(Associate Professor Kumthorn Thirakhupt, Ph.D.)

.... Thesis Advisor
(Assistant Professor Pongchai Harnyuttanakorn, Ph.D.)

.... Thesis Co-advisor
(Naowarat Kanchanakhan, Ph.D.)

.... Examiner
(Chutaphant Pinswasdi, Ph.D.)

.... External Examiner
(Ms. Saowanit Vijaykadga)

ภุมินทร์ สิมพลิพันธ์ : ความหลากหลายของยีน เมโรซอยต์เซอร์เฟซโปรตีน 1, 2, 3, 4 และ 5 (*msp*-1, 2, 3, 4 และ 5) ของเชื้ออมาลาเรีย *Plasmodium falciparum* ที่เก็บจากพื้นที่ชายแดนของประเทศไทย (DIVERSITY OF MEROZOITE SURFACE PROTEIN-1, 2, 3, 4 and 5 (*msp*-1, 2, 3, 4 and 5) GENES AMONG *Plasmodium falciparum* COLLECTED FROM BORDER AREAS OF THAILAND) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร. พงชัย หาญยุทธนากร, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: อ.ดร. เนาวรัตน์ กาญจนาคาร, ... หน้า.

โปรตีนในกลุ่มเมโรซอยต์เซอร์เฟซโปรตีน (MSP) ของเชื้ออมาลาเรีย*P. falciparum* เป็นกลุ่มโปรตีนที่ได้รับการยอมรับว่ามีศักยภาพในการนำมาพัฒนาเป็นวัคซีนสำหรับป้องกันการติดเชื้ออมาลาเรียที่ดีกลุ่มหนึ่ง โดยเฉพาะอย่างยิ่งโปรตีน MSP-1, 2, 3, 4 และ 5 ดังนั้นความหลากหลายของยีนเมโรซอยต์เซอร์เฟซโปรตีน (*msp*) ที่มีรหัสในการสังเคราะห์โปรตีนทั้ง 5 จึงอาจส่งผลต่อการพัฒนาวัคซีน แต่ข้อมูลความหลากหลายของยีน *msp* ในประเทศไทยได้มาจากการศึกษาตัวอย่างของเชื้ออมาลาเรียในบางพื้นที่เท่านั้น การศึกษาในครั้งนี้จึงได้ทำการศึกษาความหลากหลายของยีน *msp-1, 2,* 3, 4 และ 5 ของเชื้ออมาลาเรียชนิด *P. falciparum* ที่เก็บได้จากจังหวัดกาญจนบุรี แม่ฮ่องสอน ระนอง อุบลราชธานี และตราดในช่วงปี ค.ศ. 2002 -2009.

ลำดับนิวคลีโอไทด์ของยีน *msp*-1 บริเวณที่ 17 จำนวน 61 ตัวอย่าง และยีน *msp*-2 จำนวน 50 ตัวอย่าง ถูกนำมาศึกษาและวิเคราะห์ความเกี่ยวข้องทางสายพันธุกรรมโดยใช้โปรแกรมชื่อ PHYLIPS ผลการศึกษาพบว่ายีน *msp*-1 บริเวณที่ 17 ของเชื้ออมาลาเรียในประเทศไทยมีการเปลี่ยนแปลงของลำดับเบสในยีนอยู่เพียง 5 ตำแหน่ง ทำให้เกิดอัลลีลที่แตกต่างกัน 5 อัลลีลซึ่งมีรหัสในการสังเคราะห์ MSP-1_{19} การวิเคราะห์ด้วยสถิติ Chi-square test พบว่ารูปแบบการกระจายของอัลลีลทั้ง 5 แบบในแต่ละจังหวัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p<0.001$) จากการศึกษายีน *msp*- 2 พบกลุ่มอัลลีล 2 กลุ่ม คือ 3D7 และ FC27 โดยเมื่อเปรียบเทียบลำดับเบสของยีน *msp*-2 จากเชื้อ 50 ตัวอย่าง (ทั้ง 2 กลุ่มอัลลีล) พบว่ามีรูปแบบของอัลลีลแตกต่างกันทั้งหมด 37 อัลลีล ซึ่งส่วนใหญ่มีความแตกต่างกันบริเวณที่ลำดับเบสเรียงตัวซ้ำกัน การวิเคราะห์สายพันธุกรรมของยีน *msp*-2 แสดงให้เห็นว่าลำดับเบสของเชื้ออมาลาเรียจากทั้ง 5 จังหวัด มีความใกล้เคียงกัน และอัลลีลที่อยู่ในกลุ่มอัลลีลเดียวกันจะถูกจัดอยู่ในกลุ่มพันธุกรรมใกล้เคียงกัน ส่วนการหาลำดับเบสของยีน *msp*-3, 4 และ 5 พบว่าไม่สามารถหาลำดับเบสด้วยวิธี nested-PCR และการหาลำดับเบสโดยตรง อาจต้องมีการปรับเปลี่ยนใช้เทคนิคที่มีความซับซ้อนมากขึ้น

ภาควิชา: ชีววิทยา	ลายมือชื่อนิสิต ....
สาขาวิชา: สัตววิทยา .	ลายมือชื่ออ. ที่ปรึกษาวิทยานิพนธ์หลัก.....................................
ปีการศึกษา: 2553	ลายมือชื่ออ. ที่ปรึกษาวิทยานิพนธ์ร่วม......................................

4972605523: MAJOR ZOOLOGY
KEYWORDS : *Plasmodium falciparum*/ MEROZOITE SURFACE PROTEIN/ PHYLOGENETIC
PHUMIN SIMPALIPAN : DIVERSITY OF MEROZOITE SURFACE PROTEIN-1, 2, 3, 4 and 5 (*msp*-1, 2, 3, 4 and 5) GENES AMONG *Plasmodium falciparum* COLLECTED FROM BORDER AREAS OF THAILAND. THESIS ADVISOR : ASSISTANT. PROF. PONGCHAI HARNYUTTANAKORN, Ph.D., THESIS CO-ADVISOR : NAOWARAT KANCHANAKHAN, Ph.D., pp.

The *Plasmodium falciparum* merozoite surface proteins (MSPs) had been considered as potential proteins to develop into an effective vaccine against malaria infection. Among those MSPs, the considerable proteins included MSP-1, 2, 3, 4 and 5. The variation of the merozoite surface protein (*msp*) genes, which encoded those MSPs, would have an influenced on the vaccine development. In Thailand, the *msp* genes diversity had been studied only in a few areas. The sequence diversity of *msp*-1, 2, 3, 4 and 5 genes of *P. falciparum* from 5 Thailand provinces, including Kanchanaburi, Mae Hong Son, Ranong, Ubonratchathani and Trat collected during 2002-2009, was determined in this study.

The block 17 of *msp*-1 gene sequences from 61 isolates and *msp*-2 gene sequences from 50 isolates had been studied and analyzed their phylogenetic relationship by the PHYLIPS program. The results indicated that the sequence of *msp*-1 block 17 gene in Thailand had only 5 limited substitution positions and these produced 5 restricted alleles of encoded MSP-1_{19}. The Chi-square test showed that the allele distribution patterns in each province were statistically different ($p<0.001$). The *msp*-2 gene sequence analysis results revealed 2 allele families, 3D7 and FC27 families. The full length comparison of 50 *msp*-2 sequences (from both families) showed 37 alleles variants most of which were different in their repetitive regions. The phylogenetic analysis of *msp*-2 gene indicated that all sequences from 5 locations were closely related and the members of the same gene family were clustered on the same clade. The DNA sequences of other genes, including *msp*-3, 4 and 5 genes, could not be obtained by nested-PCR and direct sequencing. More complicated techniques may be needed to complete those gene sequences.

Department: Biology	Student's Signature ....
Field of Study: Zoology	Advisor's Signature..
Academic Year: 2010	Co-advisor's Signature..

ACKNOWLEDGEMENTS

This thesis could not be completed without Assist. Dr. Pongchai Harnyuttanakorn, my supervisor. I would like to give my gratitude and appreciation to him for his invaluable support such as his guidance and the opportunity to pursue my research generously. In addition, I would like to sincere my special appreciation to my co-advisor Dr. Naowarat Kanchanakhan, who always provides her invaluable advices and teaching many basic molecular techniques, and Dr. Robert Bucher who give many advises of phylogenetic tree analysis. I would like to express my special thanks to Mrs. Aree Seugorn and Miss Napaporn Siripoon for their helpfulness and preparing the parasite samples for my research. I would like to acknowledge to Dr. Chutaphant Pinswasdi and Dr. Tepanata Pumpaiboon for their kindness support and many practical applied suggestions.

Furthermore, I would like to give my gratitude to these people, Dr. Daungkae Sitticharoenchai, Dr. Chatchawan Chaisuekul, Mr. Ierson Vasinopas, Mr. Thiti Saisue and Mr. Thapana Choicharoen, for their suggestions, cheerfulness and encouragement. Special thanks to Malaria Research Program, College of Public Health Sciences, Chulalongkorn University for Laboratory supports.

This paper is supported by Office of The National Research Council of Thailand, College of Public Health Sciences and The Thai government budget 2010, under the Research Program on Conservation and Utilization of Biodiversity and the Center of Excellence in Biodiversity, Faculty of Science, Chulalongkorn University, CEB_M_49_2009.

CONTENTS

	PAGE
ABSTRACT IN THAI	IV
ABSTRACT IN ENGLISH	V
ACKNOWLEDGEMENTS	VI
CONTENTS	VII
LIST OF TABLES	VIII
LIST OF FIGURES	IX
CHAPTER I INTRODUCTION AND MOTIVATIONS	11
1.1 Motivation	11
1.2 Objectives	12
1.3 Scope of the study	12
CHAPTER II LITERATURE REVIEW	13
2.1 Global situation of malaria	13
2.2 Biological characters of *Plasmodium falciparum*	16
2.3 The *Plasmodium* antigens	19
CHAPTER III MATERIALS AND METHODS	22
3.1 Materials	22
3.2 Methods	25
CHAPTER IV THE MEROZOITE SURFACE PROTEIN 1 GENE	35
4.1 Introduction	35
4.2 Results	36
4.3 Discussions and Conclusions	40
CHAPTER V THE MEROZOITE SURFACE PROTEIN 2 GENE	44
5.1 Introduction	44
5.2 Results	46
5.3 Discussions	61
CHAPTER VI THE MEROZOITE SURFACE PROTEIN 3 GENE	67
6.1 Introduction	67
6.2 Results	68
6.3 Discussion	71
CHAPTER VII THE MEROZOITE SURFACE PROTEIN 4 GENE	73
7.1 Introduction	73
7.2 Results	74
7.3 Discussions	77
CHAPTER VIII THE MEROZOITE SURFACE PROTEIN 5 GENE	78
8.1 Introductuion	78
8.2 Results	80
8.3 Discussion	84
CHAPTER IX CONCLUSION	85
REFERRENCES	87
APPENDICES	93
BIOGRAPHY	105

LIST OF TABLES

TABLE		PAGE
3.1	The profile of 64 parasite samples used in this thesis were shown here. The samples were collected from Kanchanaburi, Ranong, Mae Hong Son, Ubonratchathani and Trat in three sequential periods; 2002-2003, 2005-2006 and 2008-2009.	23
3.2	The data show the basic information of designed primer, including primer names, sequence of each primers, their melting temperatures and amplicon sizes.	27
3.3	The information of the outgroups which serve as references for the evolutionary relationship analysis of the merozoite surface protein 2 gene.	33
4.1	The table showed the distribution of *msp1*$_{19}$ alleles among falciparum malaria population in Thailand and also their frequencies.	38
4.2	The table showed the distribution pattern of MSP1$_{19}$ alleles in 5 provinces of Thailand in this study and the data of previous study in Tak province were included.	39
4.3	The table showed infected cases which reported between 2007-2009.	40
4.4	The table shows the conclusion of MSP-1$_{19}$ haplotype of all studies isolates.	43
5.1	The table shows the 3D7 and FC27 allele families frequencies in each studied period.	56
5.2	The table shows the allele frequency of 3D7 and FC27 family of MSP-2 in 5 studied areas including Kanchanaburi (k), Ranong (RN), Trat (TD), Ubonratchathani (UB) and Mae Hong Son (MH).	56
5.3	The table shows 3D7 and FC27 alleles frequencies in 3 periods of year.	57
5.4	The table shows summary of block haplotype of FC27 isolates.	64
5.5	The table shows summary of block haplotype of 3D7 isolates.	66
A.1	The 61 sequences of MSP-1$_{19}$ are shown below. The non-synonymous mutation positions are located at position 1 (E/Q), position 48 (K/T), position 57 (N/G), position 58 (G/R) and position 73 (L/F).	94
A.2	The 33 sequences of MSP-2 3D7 family are shown below.	96
A.3	The 17 sequences of MSP-2 FC27 family are shown below.	102

LIST OF FIGURES

FIGURE		PAGE
2.1	Global distribution of malaria. High prevalence occurs in tropical zone including Thailand and South-east Asia countries (World Health Organization, 2004).	14
2.2	Ten provinces with highest malaria infection, Thailand in 2009.	15
2.3	Life cycle of malaria parasites. CDC websites, http://www.dpd.cdc.gov/dpdx, 2010).	16
2.4	Re-orientation mechanism of the merozoite.	18
2.5	The localization of the surface antigens and associated proteins in merozoite stage.	21
4.1	A photograph of an agarose gel electrophoresis showed some amplified products from malaria parasite samples.	36
4.2	Five genotypes of the block 17 are aligned and compared with allele A (top line). Conserved sequences of other alleles are represented as dot below allele A sequence. Point mutations in other alleles are replaced with alternative nucleotide letters.	37
5.1	The picture shows putative structure of MSP2.	45
5.2	The picture shows showed some multiple bands products and size variation of *msp*-2 gene of Maehongson and Trat parasites.	47
5.3	The picture shows the MSP-2 amino acid sequences of FC27 (A) and 3D7 (B) alleles.	48
5.4	The picture shows 3 alleles of C-terminus sequences.	49
5.5	The picture shows MSP-2 block 2 sequence of FC27 family.	50
5.6	The picture shows the N-terminus (A.) and C-terminus (B.) of MSP-2 repetitive block 3 sequences of FC27 family.	51
5.7	The picture shows the MSP-2 block 4 sequence of FC27 family.	51
5.8	The pictures show the N-terminus (A.) and C-terminus (B.) non-repetitive regions of MSP-2 block 2 sequences of 3D7 family.	52
5.9	The picture shows the N-terminus of repetitive region in 3D7 block 3 sequence.	54
5.10	The picture shows the non-repetitive region of block 3 sequences of 3D7 family.	54

FIGURE		PAGE
5.11	The picture shows C-terminus repetitive region, including TPA region (A.) and T region (B.), of 3D7 family.	55
5.12	The pictures show the MSP-2 block 4. The dot (.) symbol refers to identical amino acid of allele D4D1.	55
5.13	This picture shows the maximum likelihood reconstructed tree.	59
5.14	This picture shows the most parsimony consensus tree. The FC27 clade of our samples is enclosed by a dash line box.	60
6.1	This picture shows putative structure of MSP-3.	68
6.2	This picture shows single band PCR products of *msp*-3 gene from 10 different isolates, using msp3F and msp3R primers as PCR primers.	69
6.3	The picture shows the signal interruption in electropherogram of forward strand of *msp*-3 gene from TD529 isolate.	70
6.4	Both gel pictures show the first round (a.) and second round (b.) of nested-PCR products which are given by gene amplification.	71
7.1	The picture shows putative structure of the *msp*-4 gene sequence.	74
7.2	This picture reveals signal interruption of the UB85 isolate forward strand sequence.	75
7.3	This picture shows both of first round (upper row) and second round (lower row) *msp*-4 gene product of nested-PCR technique.	76
8.1	The general features of the *msp*-5 gene.	79
8.2	The picture shows the second round PCR product of 9 Kanchanaburi isolates and 1 Tak isolate (served as positive control).	80
8.3	The picture shows an electropherogram with signal interference from sequencing reaction of K195 isolate.	81
8.4	The picture shows the PCR product from RN19 and RN26 isolates.	82
8.5	The picture shows the putative structure of the partial chromosome 2 of *Plasmodium falciparum* which contains *msp*-2, *msp*-4 and *msp*-5 genes.	83

CHAPTER I

INTRODUCTION AND MOTIVATIONS

1.1 Motivations

Malaria remains as one of the major global health problems. There are approximately 2-3 millions death from malaria infection each year. The transmission of the disease occurs in tropical regions including Thailand. Although the malaria eradication program launched by Thai government, successfully eradicate this disease in the central part of Thailand, people lived near the border regions of Thai-Myanmar, Thai-Laos, Thai-Cambodia and Thai-Malaysia are still under the influence of malaria disease. Four species of human malaria, *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*, are found in Thailand, with a newly identified species in human, *P. knowlesi*. However, the widely spread species of human malaria are *P. falciparum* and *P. vivax*.

The use of antimalaria drug is the most popular approach to prevent and control the malaria infection because of their effectiveness and low-cost. Additional strategies for malaria prevention are increasing after the drug resistant parasite emerged worldwide. Malaria vaccine has been considered as a hope for malaria eradication strategy. Unfortunately, there is no effective vaccine available. The major problem of the malaria vaccine development is the diversity of target genes and proteins of the parasites in different parts of the world. Thailand is, now, known as a low transmission area, but many reports indicate that there is high variation of vaccine candidate protein sequences among parasites in this country, especially those proteins on the surface of merozoite stage. The merozoite surface protein (MSP)-1, 2, 3, 4 and 5 are the most investigated vaccine candidate proteins of *P. falciparum*. Although many research works had determined the polymorphism status of these proteins, the data cannot be related. Moreover, the parasite collection in previous studies are collected only from the North-western

region of Thailand, Tak province, thus it may not be the good representative of polymorphism status of each proteins in Thailand. In this study, the leading vaccine candidate proteins of *P. falciparum*, merozoite surface protein (MSP) 1, 2, 3, 4 and 5, are studied to reveal the diversity of their nucleotide and amino acid sequences. The parasite isolates in our studies are collected from 5 border areas of Thailand, where high incident of infected cases are reported each year.

1.2 Objectives

1.2.1 To reveal the partial gene sequence of merozoite surface protein (*msp*)-1, 2, 3,4 and 5 genes of *P. falciparum* collected from Thailand

1.2.2 To study the diversity of *msp*-1, 2, 3, 4 and 5 genes of *P. falciparum* collected from Thailand

1.3 Scope of the study

1.3.1 To reveal the gene sequence by using PCR and sequencing technique and analyze their polymorphism status by appropriate statistical method.

CHAPTER II

LITERATURES REVIEW

2.1 Global situation of malaria

Malaria remains an important infectious disease. The transmission of the disease is worldwide especially in tropical region of the world (Figure 2.1). The number of infected case each year is around 200-300 millions. Among them, 2-3 million clinical cases die from severe symptom. Approximately 60% of the clinical cases and 80% of the deaths are occurred in Africa (World Health Organization, 2005). However, the real number of infected cases may be higher than the report of World health Organization (Hay *et al.*, 2007). In human, the malaria infection causes by *Plasmodium* parasite involving *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. The most malignant species among 5 human malaria parasites is *P. falciparum* which may lead to death by severe symptom known as cerebral malaria. Malaria has not only been the cause of health problem but it has also been the cause of economic and social burden problems. There are many ways which the disease affects economically and socially to human such as reducing rate of population growth, low productivity, losing saving and investment, adolescent mortality and medical costs for disease prevention (Sachs and Malaney, 2002).

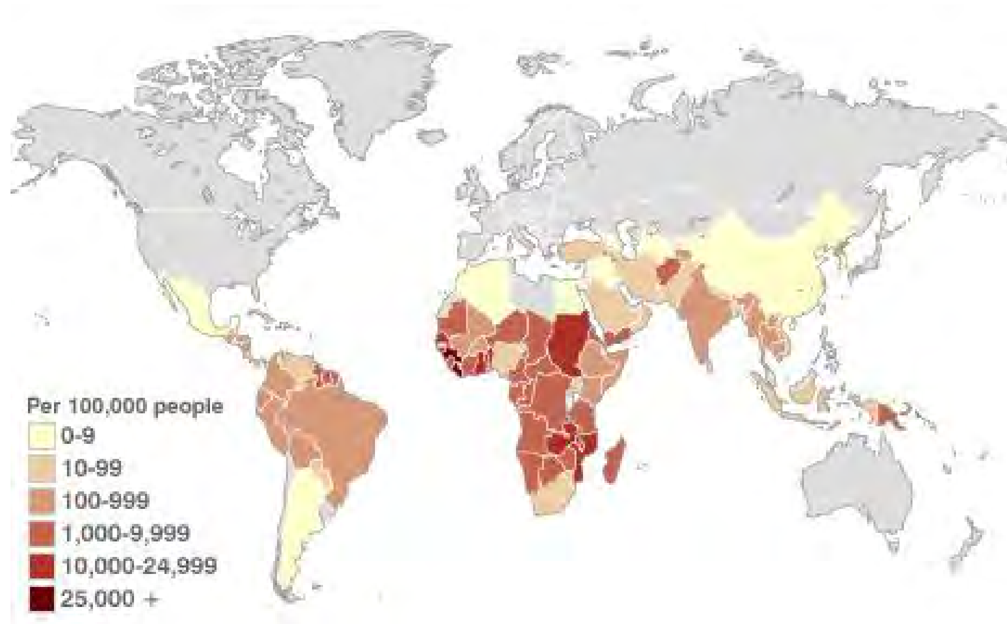

Figure 2.1 Global distribution of malaria. High prevalence occurs in tropical zone including Thailand and South-east Asia countries (World Health Organization, 2004).

Although, malaria cases, in Thailand, are still found along the border areas of Thai-Myanmar, Thai- Cambodia, Thai-Laos and Thai- Malaysia, the number of infection has declined annually. For example, malaria incidents during 2007-2009 periods decrease from 33,178 cases, 25,969 cases and 23,328 cases respectively. The top ten provinces where malaria infected cases are found in 2009 are Tak, Yala, Mae Hong Son, Kanchanaburi, Srisaket, Surin, Narathiwat, Chumphon, Chanthaburi and Prachuapkhirikhan as described in figure below (Department of Disease Control, Ministry of Public Health, Thailand, 2009).

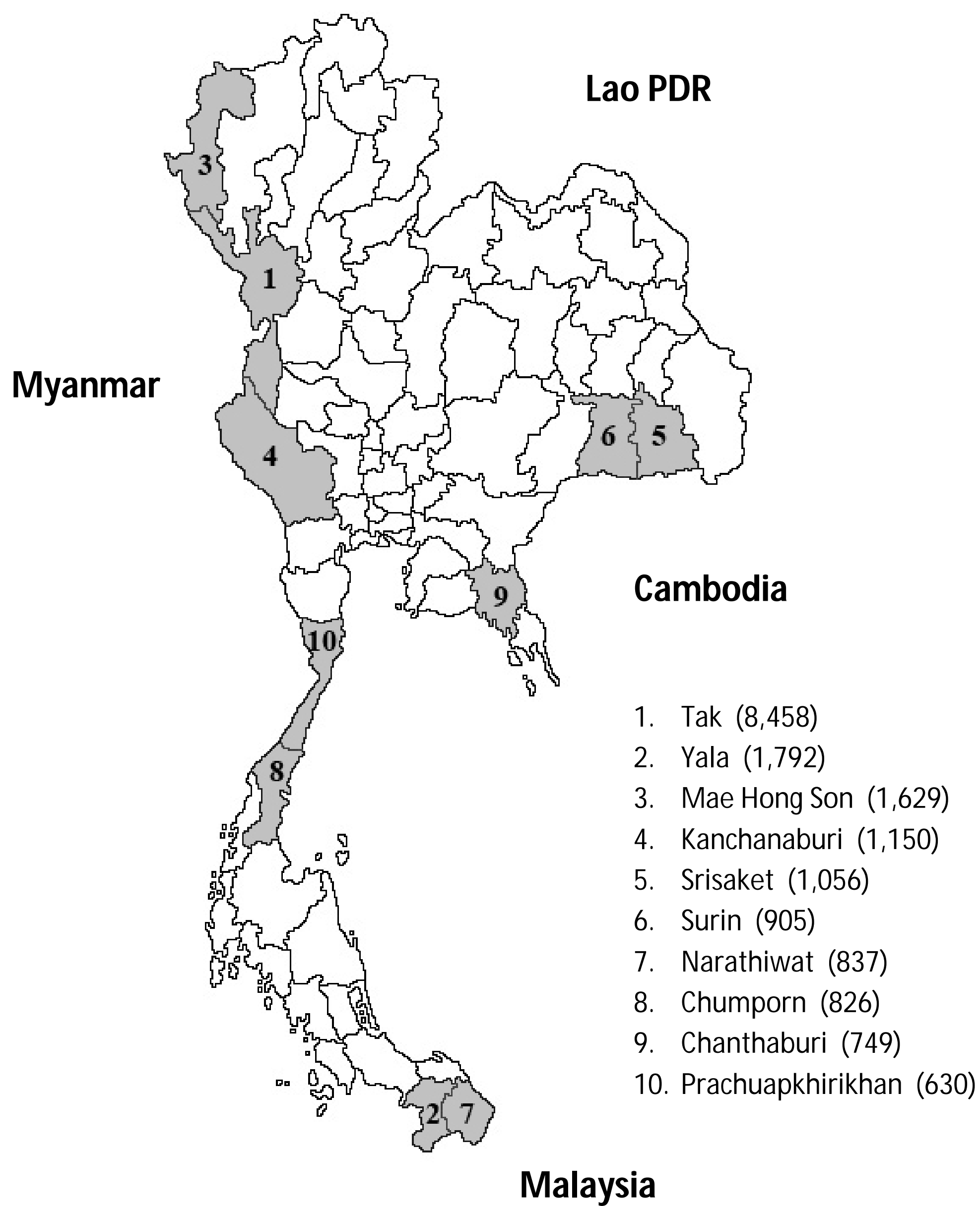

Figure 2.2 Ten provinces with highest malaria infection, Thailand in 2009

2.2 Biological characters of *Plasmodium falciparum* (Knell, 1991)

Malaria is caused by parasite in genus *Plasmodium.* These parasites are transmitted into human body by mosquito bites. The *Plasmodium* parasites are unicellular organisms which have complex life cycle in both human and mosquito vector.

2.2.1 Life cycle of *Plasmodium falciparum*

The life cycle of this parasite can be divided into 2 major stages: asexual stage in human host and sexual stage in female mosquito vector (Figure 2.3).

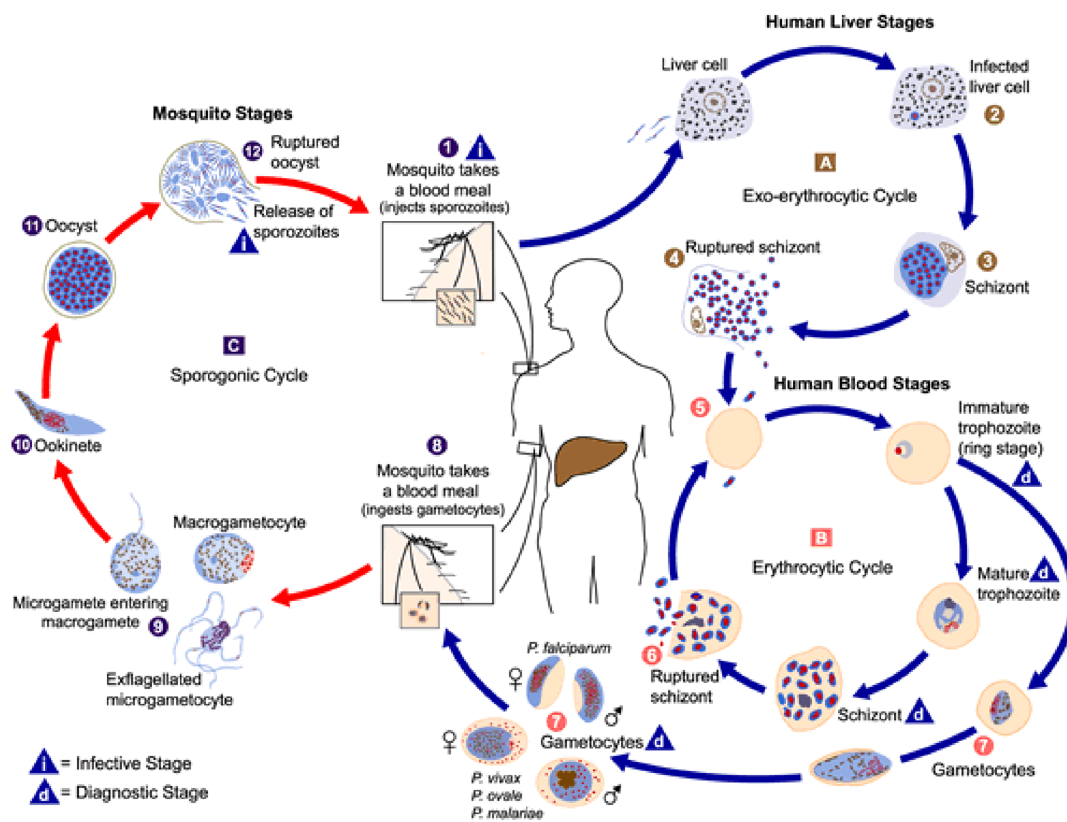

Figure 2.3 Life cycle of malaria parasites.

(CDC websites, http://www.dpd.cdc.gov/dpdx, 2010)

2.2.1.1 Asexual stage in human

Pre-erythrocytic schizogony

This stage starts with the inoculation of the parasite into human blood by the bite of mosquito vector. The sporozoites in mosquito saliva mobilize to the liver, then, rapidly penetrate into the hepatic parenchyma cells. The sporozoites develop to trophozoites, then, trophozoites initiate their intracellular asexual cell division. This process produces a large number of merozoites which are released from the liver cells at the completion of this stage. The number of merozoites and time for completion the process are depended on species of parasite.

Erythrocytic schizogony

The merozoites released from the liver cells attach to the red blood cell by multiple associated proteins on the surface of merozoites. The mechanism of merozoites invade into the red blood cell is described as re-orientation mechanism (Figure 2.4). This process start when the lateral position of merozoites attach with red blood cell membrane. Many surface proteins of the parasite, such as merozoite surface proteins (MSPs) and erythrocyte binding antigens (EBAs), serve as parasite-red blood cell attachment tools. Then, the attachment position gradually changes from the lateral to anterior position by re-orientation of the merozoite. The penetration of merozoite into the red blood cell begins when anterior path of parasite attach the target membrane. In this step, the parasite invasion uses apical complex organelle such as rhoptries and micronemes by discharge their contents through apical region. These contents believe to help the merozoite invasion into the red blood cell and form parasitophorous vacuole. The merozoites in the red blood cell grow and develop through the following stages; ring forms, trophozoites, early shizonts and mature shizonts. At the end of the

development, the merozoites are released into human blood circulation by breakage of the red blood cell membrane. After the releasing, the merozoites immediately invade to the other new red blood cell. These steps of the development, invasion-multiplication-releasing-invasion, are continued repeatedly . A few numbers of merozoites in red blood cell may differentiate into gametocytes, male and female. These gametocytes are taken by mosquito vector when mosquito bites and feeds the blood of infected host.

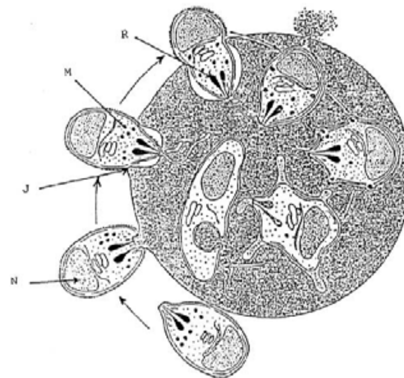

Figure 2.4 Re-orientation mechanism of the merozoite.

2.2.1.2 Sexual stage in mosquito

Sporogony

The intake gametocytes complete their development to be the flagellated microgametocytes and macrogametocytes. After their fertilization, the zygote is formed in the gut of mosquito. Then, the zygote differentiates to the ookinete which penetrates the epithelial layer of gut wall and transforms to the oocyst. The asexual division of an oocyst produces numerous sporozoites which are released to the haemolymph and moved to mosquito salivary gland. The new asexual stage in human starts when the sporozoites are inoculated to human blood stream through the mosquito bite.

2.3 The *Plasmodium* antigens

The sporozoites and merozoites are important stages of the parasite that are exposed to human immune system. The other forms, in human body, are intracellular living stage which is protected by host cell membrane from host immune system. As the result, many of the surface proteins, occur in sporozoites and merozoites, may be use as protective vaccine.

2.3.1 The merozoites antigens (Cowman *et al.*, 2002)

Merozoite invasion starts at the interaction between this parasite and red blood cell membrane, followed by the re-orientation which leaves the apical part of the parasite to interact with the membrane of the target cell. This host cell invasion process is achieved by various components such as surface antigens, apical complex, rhoptries and micronemes (Figure 2.5). The numerous surface proteins are responsible for initiation binding interaction while the ultrastructures show the tight junction forming between parasite's apical complex and host cell membrane. This tight junction can be moved along the surface of the merozoite by the actin-myosin motor. At the end of invasion steps, the merozoite surface coat proteins are shreded before the red blood cell membrane fuses at the posterior end of the merozoite. As the results, the parasitophorous vacuole is formed. ,Many surface merozoite proteins involve in the complex invasion process but the knowledge about the particular role of each individual protein is unclear. The initial interaction of the merozoite surface and erythrocyte membrane involves numerous of surface proteins and some of those have been characterized. The proteins which serve as the initial interaction can be simply divided into 2 groups. The first group is anchored protein on the merozoite surface and that interact with the red blood cell during invasion process. The other group is peripheral proteins which localize in parasitophorous vacuole. These proteins associate with the anchored proteins on the merozoite surface.

The merozoite surface proteins (MSPs) are one of the protein families which attach on the surface of the merozoite as its name implies. This protein family consists of many different forms of the peptide and they also have protein size variations. The most considerable protein of this family is merozoite surface protein (MSP) 1; the glycosylphosphatidylinositol (GPI)-anchored protein, because it may be one of the best vaccine candidate polypeptide (Holder and Freeman, 1984). Moreover, the MSP-1 is a largest protein among this protein family and it also have high variation in either protein or gene sequences. The other GPI-anchored proteins, including MSP-2 (Smythe *et al.*, 1988), MSP-4 (Marshall *et al.*, 1997) and MSP-5 (Marshall *et al.*, 1998), share the localization on the merozoite surface and they are also considered to be a vaccine candidate polypeptide. On the other hand, some proteins, which belong to the MSPs family, in merozoite stage, display the extrinsically associated protein. They associate to other surface proteins during invasion process. The MSP3 (McColl *et al.*, 1994) is one of the extrinsic protein which localize in the parasitophorous vacuole. This protein is synthesized in shizont-stage parasite, then, it is secreted in to the parasitophorous vacuole.

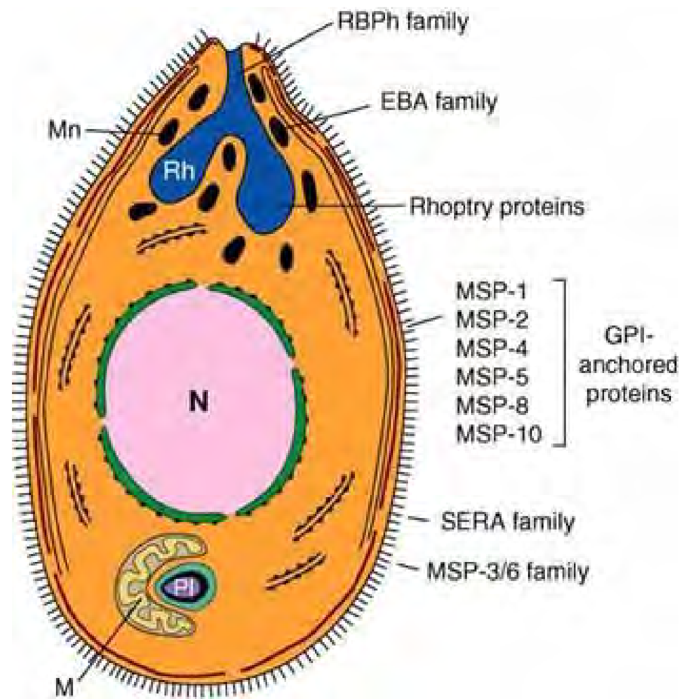

Figure 2.5 The localization of the surface antigens and associated proteins in merozoite stage

(Cowman *et al.*, 2000).

CHAPTER III

MATERIALS AND METHODS

3.1 Materials

3.1.1 Isolates of *Plasmodium falciparum*

All 64 parasite samples were received from Malaria Research Program, College of Public Health Sciences, Chulalongkorn University, Thailand. These blood samples were collected from 5 provinces of Thailand during 2002-2009. The collected areas and periods of parasite samples were shown in Table 3.1. Prior to this study, the parasites were culture-adapted and preserved in liquid nitrogen and then they were revived by the candle-jar method (Trager and Jensen, 1976).

Provinces	years	isolates	No. of
Kanchanaburi (K)	2002-2003	K58, K60, K64, K66, K74	5
	2005-2006	K165, K185, K195, K205, K215	5
	2008-2009	K386, K389, K391, K392, K397	5
Ranong (RN)	2002-2003	RN19, RN26, RN28, RN31, RN36	5
	2005-2006	RN63, RN66, RN68, RN70, RN72	5
	2008-2009	RN122, RN129, RN130, RN131, RN133	5
Mae Hong Son (MH)	2002-2003	MH6, MH7, MH9, MH10, MH11	5
	2005-2006	MH18, MH20, MH24, MH28, MH32	5
	2008-2009	-	0
Ubonratchathani (UB)	2002-2003	UB7, UB14, UB22, UB27, UB28	5
	2005-2006	UB50, UB51, UB52, UB58, UB59	5
	2008-2009	UB82, UB84, UB85	3
Trat (TD)	2002-2003	TD504, TD508, TD510, TD515	4
	2005-2006	TD529, TD530, TD531, TD533, TD542	5
	2008-2009	TD554, TD556	2
		total	64

Table 3.1, the profile of 64 parasite samples used in this thesis were shown here. The samples were collected from Kanchanaburi, Ranong, Mae Hong Son, Ubonratchathani and Trat in three sequential periods; 2002-2003, 2005- 2006 and 2008-2009.

3.2 Enzymes and buffers

The protenase K and *Taq* DNA polymerase used in this study were purchased from commercial source (Roche and Fermentas). The buffers and other chemical reagents for gene amplification namely 10X *Taq* DNA polymerase buffer without magnesium, 25 mM dNTPs and 25 mM magnesium chloride were provided by both manufacture. The commercial purchased reagents were kept in -20 °C until needed.

3.3 Reagents and chemicals

In this study, the reagents were used for genomic DNA preparation and gel electrophoresis. All reagents were sterilized by autoclaving before used.

3.3.1 Genomic DNA preparation reagents

Reagents	Compositions
0.05% saponin	0.05 g of saponin final volume 100 ml
6X phosphate buffer saline (PBS)	48.6 g of NaCl 3.06 g of KH_2PO_4 9.585 g of Na_2HPO_4 adjusted pH to 7.2 with NaOH final volume 1,000 ml
Lysis buffer	40 mM of Tris-HCl pH 8.0 80 mM of EDTA pH 8.0 2% SDS
1 M Tris-HCl buffer	121.1 g of Tris adjusted pH to 8.0 with HCl final volume 1,000 ml
TE buffer	10 mM of Tris-HCl pH 8.0 0.1 mM EDTA pH 8.0
Phenol-chloroform	50% saturated phenol 48% chloroform 2% isoamyl alcohol
3 M sodium acetate	40.8 g of Na acetate 3 H_2O adjusted pH to 5.2 with glacial acetic final volume 100 ml
10% sodium dodecyl sulfate (SDS)	10 g of SDS final volume 100 ml
0.5 M disodium ethylene diamine tetra actate (EDTA)	181.6 g of EDTA adjusted pH to 8.0 with NaOH pellet final volume 1,000 ml

3.3.2 Gel electrophoresis reagents

Gel electrophoresis reagents	compositions
10X TBE buffer	121.14 g of Tris-HCl 55 g of boric acid 40 ml of 0.5 M EDTA pH 8.0 Final volume 1,000 ml
6X loading dye	0.4 mg of bromophenol blue 0.4 mg of xylene cyanol 3 ml of glycerine final volume 10 ml
100 bp standard DNA ladder marker	5:1 ratio of marker and 6X loading dye
Ethidium bromide (EtBr) staining solution	0.5 µg/ml EtBr

3.4 Methods

3.4.1 *Plasmodium falcipaum* genomic DNA preparation

In this study, the parasite genomic DNA was extracted by phenol-chloroform method as described by Snounou (Snounou *et al.*, 1993). Firstly, two hundred microlitre (µl) of blood contained cultured parasites was mixed with 500 µl of 0.05% saponin and incubated at room temperature for 10 min or until the solution was transparent. The mixture was centrifuged at 13,000 rpm for 10 min. The supernatant was removed and the parasite pellet was washed twice with 1,000 µl of cold 1X Phosphate buffer saline (PBS). The pellet was lysed with 200 µl of the Proteinase K solution (Proteinase K, 0.5mg/ml) overnight at 37 ºC. After incubation, the volume was adjusted to 500 µl by autoclaved distilled water. The sample was, then, extracted with the phenol chloroform extraction procedure. The Tris-EDTA (TE) saturated phenol, 500 µl, was added and mixed by vortex for 1 min. After centrifugation at 13,000 rpm for 10 min, the upper layer was transferred to a new tube and combined with 500 µl of phenol/chloroform/isoamyl alcohol. The aqueous phase was separated by centrifuged at 13,000

rpm for 10 min. Sodium acetate (3 M, pH 5.2) 50 µl and 1,000 µl of cold absolute ethanol was mixed with the upper layer in a new tube for DNA precipitation. Solution was kept in -20ºC overnight. DNA pellet were obtained by centrifugation at 13,000 rpm for 10 min and washed with 70% ethanol. The pellet was air dried for 20-30 min. Finally, 100 µl of TE buffer was added to dissolve DNA. The DNA stocked solution was preserved in -20ºC until used.

3.4.2 Primer design

P. falciparum 3D7 strain chromosome sequences, from Genbank databases, were used for primer design. The sequence numbers included NC000910 (*msp* -2, 4 and 5 genes), NC004314 (*msp*-1 gene) and NC004330 (*msp*-3 gene) which were located on the *P. falciparum* chromosome 2, 9 and 10 respectively. Melting temperature (T_m) of each primer was predicted by a simplified T_m equation.

$$T_m = 2(A+T) + 4(G+C)$$

A, T, G and C factors mean the total number of A, T, G and C nucleotides in the primer sequence. All designed primers were checked for their specificity with Basic Local Alignment Search Tool (BLAST) at National Center of Biotechnology Information (NCBI) website. The sequences of 5 merozoite surface protein genes primers were shown in the following table (table 3.2). Primers were synthesized and their T_m was calculated by Sigma-Aldrich Pte.Ltd, Singapore.

Gene	Primer name	Nucleotide sequence	Tm (°C)	Product size (bp)
msp-1 block 17	msp-1/17 forward	TCA CAA CAC CAA TGC GTA AAA	64	312
	msp-1/17 Reverse	CTT AGG AAT ATC ATT CTT ATT AAT ACT C	55	
msp2	msp-2 Forward	ATG AAG GTA ATT AAA ACA TTG TCT A	57	819
	msp-2 Reverse	AAG TCA ATG TTT TAA AAT GAA GAG AA	60	
msp-3	new_outM3F	TGC ACA CAT TAT TAC GTC TAA AT	58	1,553
	new_outM3R	GGA TTT ATA GAA TAT AAA AGG AAT C	54	
	msp3F	ATG AAA AGT TTT ATA AAT ATT ACT CTT TC	56	1,101
	msp3R	CAT GTT ATG AAT ATA AAT TAT GTT CA	56	
msp-4	msp4_primerF	TTT AAG TAA TCA TGC TTT TTC AAC AC	51	1,795
	msp4_primerR	TGG TTC CAT TCA ATG ACA AAA	50	
	Int_msp4F	TAT GAA TTC ATA TTT TAT ATT TTC AGA TGT	51	1,405
	Int_msp4R	GTA TAC ATC CCT TTA AGT TTT CGA ACA	54	
	newmsp4F	GAA TTC ATA TTT TAT ATT TTC AGA TGT	57	1,086
	newmsp4R	GGT AAG ATT TAA ACA TAT AAA ATT ATG	55	
	new_intM4F	GTG GAT AGT TAA ATT TTA ATA GTA G	51	1,276
	new_intM4R	AGG TTG TAG ATA TAA AGA ATT ATG GA	57	
msp-5	Ext_msp5F	GTA ACA TGC ATC TGA TTC G	54	1,856
	Ext_msp5R	GTG TGC ATT TGA TTT ATT ATA TGA AAA G	52	
	msp5_primerF	TGA CTT ATG AAT GGT CAC ACA CA	50	1,308
	msp5_primerR	TCC GAG GGA TAA TAA GTG TG	58	
	new_outM5F	TGA ATG GTC ACA CAC ACA TGT	63	1,036
	newmsp5R	TTA AAC GAA TAA TAA CAA AAA CAA TGT A	60	
	newmsp5F	ATG AAT ATA TTA TGT ATT CTA TCA TAT A	51	954
	newmsp5R	TTA AAC GAA TAA TAA CAA AAA CAA TGT A	60	

Table 3.2 The data show the basic information of designed primers, including primer names, sequence of each primers, their melting temperatures and amplicon sizes.

3.4.3 Gene amplification

The Polymerase chain reaction (PCR) was used to amplify the 5 interested genes. In addition, the *msp*-3, 4 and 5 genes were amplified by the nested-PCR method. The method of PCR was described by Mullis (Mullis, 1990). For PCR and nested-PCR, the total volume for 1 reaction was 20 μl. The genomic DNA templates were diluted with distilled water into 1:10 dilution, then, they were incubated at 60°C for 10 min prior to use. In this study, 1 μl of 1:10 dilution DNA template was added into the PCR mixture. The thermocycler in this study was the GeneAmp PCR system 9700 machine, Applied Biosystems. The concentration of *Taq* DNA polymerase and other PCR reagents in the mixture were adjusted as shown below.

Reagents	Final volume (μl)
10X *Taq* DNA polymerase buffer	2
25 mM magnesium chloride	1.6
2.5 mM dNTPS	2
20 μM forward primer	0.5
20 μM forward primer	0.5
5 unit/μl Taq DNA polymerase	0.5

The specific amplification cycle of each studied genes was described below.

3.4.3.1 *msp*-1 block 17 gene amplification cycle

Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	40 sec
Annealing	53ºC	40 sec
Extention	72ºC	40 sec
Last extention	72ºC	3 min
Number of cycle: 30 cycles		

3.4.3.2 *msp*-2 gene amplification cycle

Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	30 sec
Annealing	50ºC	30 sec
Extention	72ºC	1 min
Last extention	72ºC	3 min
Number of cycle: 30 cycles		

3.4.3.3 *msp*-3 gene amplification cycle

The condition of *msp*-3 gene amplification was varied because the final result could not be obtained. The latest nested-PCR profile of *msp*-3 gene was described below.

First round amplification		
Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	40 sec
Annealing	55ºC	40 sec
Extention	72ºC	2 min
Last extention	72ºC	7 min
Number of cycle: 25 cycles		
Second round amplification		
Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	30 sec
Annealing	52ºC	30 sec
Extention	72ºC	1.20 min
Last extention	72ºC	3 min
Number of cycle: 30 cycles		

3.4.3.4 *msp*-4 gene amplification cycle

The condition of *msp*-4 gene amplification was varied because the final result could not be obtained. The latest nested-PCR profile of *msp*-4 gene was described below.

First round amplification		
Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	40 sec
Annealing	61ºC	40 sec
Extention	72ºC	1.50 min
Last extention	72ºC	3 min
Number of cycle: 25 cycles		
Second round amplification		
Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	40 sec
Annealing	60ºC	40 sec
Extention	72ºC	1.30 min
Last extention	72ºC	5 min
Number of cycle: 30 cycles		

3.4.3.5 *msp*-5 gene amplification cycle

The condition of *msp*-5 gene amplification was varied because the final result could not be obtained. The latest nested-PCR profile of *msp*-5 gene was described below.

First round amplification		
Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	40 sec
Annealing	60ºC	40 sec
Extention	72ºC	2 min
Last extention	72ºC	5 min
Number of cycle: 25 cycles		
Second round amplification		
Steps	Temperature	Time
Initial denaturation	95ºC	5 min
Denaturation	95ºC	40 sec
Annealing	57ºC	40 sec
Extention	72ºC	1.30 min
Last extention	72ºC	5 min
Number of cycle: 30 cycles		

3.4.4 PCR product detection

In this step, the amplified products were detected with agarose gel electrophoresis method (Sambrook et al., 1989). Five microlitres of PCR product were mixed with 1 µl of 1X loading dye before and loaded into 1.8% (w/v) agorose gel using 100 base pair ladder from Bioexcellent™ as a DNA size standard. The gel electrophoresis was run at 100 volts for 30 min in 1X Tris-Boric acid-EDTA (TBE) buffer with Mupid-ex intelligent power supply, Advance Co., Ltd. After stained by 0.5 µg/µl ethidium bromide (EtBr) in 1X TBE buffer for 5-10 min, the gel was washed in distilled water, then, it was illuminated by ultraviolet light. Graphic information was captured by the Autochemi™ system.

3.4.5 DNA sequencing

After the studied gene, from each isolate, was amplified in 3-5 reaction tubes, then, they were pooled into a 1.5 ml microfuge tube. All samples, at least 50 μl, were purified prior to direct sequencing. Both strands of all amplified products were sequenced. Sequencing reactions were performed in a MJ Research PTC-225 Peltier Thermal Cycler using a ABI PRISM BigDyeTM Terminator Cycle Sequencing Kits with AmpliTaq DNA polymerase (FS enzyme) (Applied Biosystems), following the protocols supplied by the manufacturer. Single-pass sequencing was performed on each template using our designed primer. The fluorescent-labeled fragments were purified from the unincorporated terminators with an ethanol precipitation protocol. The samples were resuspended in distilled water and subjected to electrophoresis in an ABI 3730xl sequencer, Applied Biosystems (Macrogen Inc., Korea).

3.4.6 Sequence alignment

The sequences from both strands of a PCR product were aligned with BioEdit program. The input file format of this program was sequence.ab1 file which obtained from Macrogen Inc., Korea. The sequence of each isolate was deduced from the matched sequences of both strands. Later, all sequences were compared and analyzed for the diversity.

3.4.7 Phylogenetic tree analysis

The PHYLIPS program version 3.2 (Felsenstein, 1989), free license, was used as a phylogeny analysis tool. This program was run on Mac Book, Mac OS X version 6.9. The input file format of this analysis program was PHYLIPS file format (sequence.phy).

3.4.7.1 The input outgroup

The outgroup is a group of organisms which serves as a reference for determination of the evolutionary relationship among monophyletic groups of organisms. The outgroup is hypothesized to be rather closely related to the interested groups, but less closely than any single one of those. The evolutionary conclusion is that the outgroup separated from the parent group before the other two groups separated from each other. The outgroup sequences of 5 studied merozoite surface protein genes were pulled out from Genbank databases. Those outgroup sequences consisted of several *P. falciparum msp*-1, 2, 3, 4 and 5 genes collecting from many regions of the world. The information of outgroups were shown in the following table.

Gene	Strain	Origin	Accession number
msp-2	IRI18	Iran	DQ338451.1
	RO33	Ghana	U91649.1
	VN14	Vietnam	AF104717.1
	Wos42	Papua New Guinea	AJ318753.1

Table 3.3 The information of the outgroups which serve as references for the evolutionary relationship analysis of the merozoite surface protein 2 gene.

3.4.3.2 Program application

This program provided several application modules. The steps of analysis were processed by following the program manuscript. The sequence dataset, including the outgroups of *msp*-1 block 17, msp-2, 3, 4 and 5 genes, was random by Bootstrapping method. The output files, from previous re-sampling step, was used as an input file for the next processes, such as

phylogenetic tree analysis with maximum likelihood (DNAml module) and the most parsimony methods (DNApenny module). For those tree analyses, they were commanded to reveal 10,000 possible trees and, then, collected the best 1,000 trees.

3.4.3.3 Consensus the tree

The multiple trees from a tree construction module, such as DNApenny, could be summarized with the Consense program.

3.4.8 Statistical analyses

The independent of distribution pattern among parasite populations was tested by Chi-square (χ^2) test. The formula of the test was shown below:

$$\chi^2 = \Sigma\frac{(fo-fe)^2}{fe}$$

fo = frequency of observed value

fe = frequency of expected value

CHAPTER IV

THE MEROZOITE SURFACE PROTEIN 1 BLOCK 17 GENE

4.1 Introduction

The development of vaccine against the *P. falciparum* infection is an alternative choice to fight against the parasite. Blood stage parasite surface proteins such as merozoite surface proteins 1 (MSP-1) are considered as vaccine candidates. The MSP-1 is located on the surface of erythrocytic merozoite. It is cleaved twice during erythrocyte invasion process. The first cleavage produces 4 protein fragments, 82, 30, 38 and 42 kDa. The second cleavage cut the 42 kDa fragment into 33 kDa and 19 kDa fragments. After the red blood cell invasion, all fragments, except the 19 kDa fragment, are shreded from the parasite surface (Blackman *et al.*, 1990 and Blackman *et al.*, 1991). The 19 kDa ($MSP\text{-}1_{19}$) fragment, which containing of 2 epidermal growth factors (EGF) like domains, still remains on the parasite surface with glycosylphosphatidylinositol (GPI) anchored tail on C-terminus of the peptide (Blackman *et al.*, 1991). Several studies indicated that these EGF-like domains are the targets of protective immunity (Chang *et al.*, 1992 and Chappel and Holder, 1993). The 19 KDa fragment is encoded by the block 17 of *msp-1* ($msp\text{-}1_{17}$) gene which appears to be conserved with some point mutations (Ferreira *et al.*, 2003, Lalitha *et al.*, 1999 and Qari *et al.*, 1998). In Thailand, the $msp\text{-}1_{17}$ gene is studied only in a group of parasite population at Mae Sod district, Tak province (Sakihama *et al.*, 1999 and Sakihama *et al.*, 2006). No published sequence variation has been reported from other areas of Thailand. The aim of this project is to explore the polymorphic data of the $msp\text{-}1_{17}$ gene among *P. falciparum* population in Thailand.

In our study, the 64 isolates of *P. falciparum* which collected from 5 provinces of Thailand, during 2002-2009, were studied. The merozoite surface protein 1 block 17 ($msp\text{-}1_{17}$)

gene was amplified by our new designed primers, specific for this block region. Consequently, all amplified products were sent to purify and directly sequenced.

4.2 Results

4.2.1 *msp*-1 block 17 gene amplification product

After the block 17 fragment was amplified, all amplified products were checked by agarose gel electrophoresis with TBE buffer. Compared with DNA size marker, all amplified products were seen as a single band at 300 bp in size, similar to the calculated size from designed primers. No size variation was detected (Figure 1).

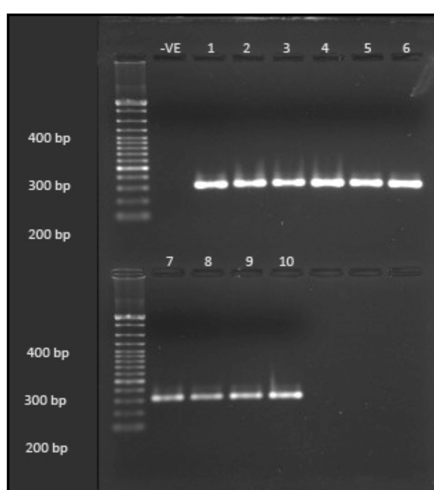

Figure 4.1 A photograph of an agarose gel electrophoresis showed some amplified products from malaria parasite samples. The gel electrophoresis condition was 1.5% (W/V) agarose in 1XTBE buffer processing under 80 voltages for 30 mins. After staining with ethidium bromide, the products were detected by UV transilluminator (UVP Bioimaging systems, AutoChemi system). The amplified product size was compared with 100 bp DNA standard ladder marker (BioExcellence, Thailand).

4.2.2 Diversity of *msp-1*$_{17}$ among *Plasmodium falciparum* population in Thailand

All 217 nucleotides of the block 17 from 61 samples were sequenced and further analyzed. A sequence from 2 samples indicated a multiple infection and was excluded from data analysis. Another 1 sample cannot be collected its genomic DNA. These block 17 fragments from Thailand is highly conserved. They showed 5 point mutation positions: CAA/GAA at 4858, AAA/ACA at 5000, AAC/AGC at 5027, GGA/AGA at 5029 and CTT /TTT at 5074 (Figure 2). All of the mutations occurred at the first (3/5) or second (2/5) positions of each codon (Table 1). These point mutations caused non-synonymous mutation at all 5 positions; E/Q at position 1644, K/T at 1691, N/S at 1700, G/R at 1701 and L/F at 1716 where they can be located on the first (1 amino acid change) and second (4 amino acid changes) EGF-like domain. Neither insertions nor deletions were detected in this gene.

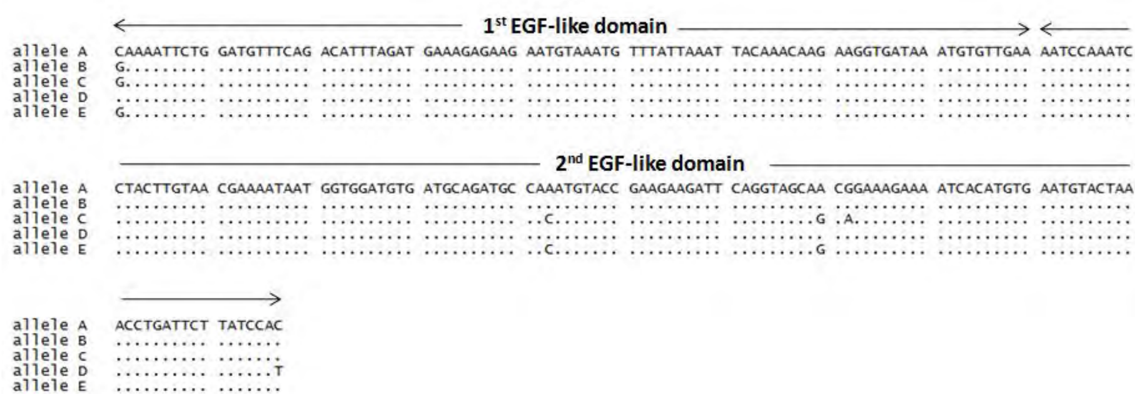

Figure 4.2 Five genotypes of the block 17 are aligned and compared with allele A (top line). Conserved sequences of other alleles are represented as dot below allele A sequence. Point mutations in other alleles are replaced with alternative nucleotide letters.

4.2.3 MSP-1_{19} alleles analysis

4.2.3.1 In Thailand population

The DNA sequences of 61 isolates were decoded and then analyzed. The amino acid change could be divided into 5 distinct alleles; E/KNG/L, Q/KNG/L, E/TSR/L, E/TSG/L and Q/KNG/F as previously reported (Sakihama *et al.*, 2006 and Sakihama *et al.*, 2007). The E/KNG/L allele was found in the majority of isolates (16/22 in 2002-2003, 13/24 in 2005-2006) of Thailand *P. falciparum* population whereas the result from last collection period, 2008-2009, was shown the share of majority between E/KNG/L and Q/KNG/L alleles. The remaining isolates were identified as other alleles (Table 1).

Allelic types	Collected years			Codon position changes				
	2002-2003 (n=22)	2005-2006 (n=24)	2008-2009 (n=15)	4858 (1644) C→G	5000 (1691) A→C	5027 (1700) A→G	5029 (1701) G→A	5074 (1716) C→T
E/KNG/L	16	13	5					
Q/KNG/L	4	5	5	**1**	**-**	**-**	**-**	**-**
E/TSR/L	2	2	2	**-**	**2**	**2**	**1**	**-**
E/TSG/L	0	3	1	**-**	**2**	**2**	**-**	**-**
Q/KNG/F	0	1	2	**1**	**-**	**-**	**-**	**1**

Table 4.1. The table showed the distribution of *msp*-1_{19} alleles among falciparum malaria population in Thailand and also their frequencies. The nucleotide changes at each codon position, in comparison with allele A, were shown. The positions of point mutations and amino acid changes, together with nucleotide changes, were indicated at the top of the column.

4.2.3.2 Allele frequencies in each population

The allele frequencies of MSP-1_{19} in all areas of Thailand, except Trat (TD) province, demonstrated that E/KNG/L allele was the majority allele and the other allelic types

frequency were identified as describe in the table 2. In the table below, the allele frequency from previous reports of this peptide were included. The E/TSG/L and Q/KNG/F alleles were not found in Mae Hong Son (MH) and Ubonratchathani (UB) provinces in all studied periods. Our data visually showed the different of allele frequency distribution in each area, then, we used the Chi-square (χ^2) test to detect our prediction. By the χ^2 test formula (described in Chapter 2), the χ^2 value showed the significantly higher than our data expected value, that illustrated the allele frequency of each locations were independent to each other.

Areas	Number of isolates	Allelic type frequencies (%)				
		E/KNG/L	Q/KNG/L	E/TSR/L	E/TSG/L	Q/KNG/F
K	15	53.33 (n=8)	13.33 (n=2)	6.7 (n=1)	13.33 (n=2)	13.33 (n=2)
RN	14	42.86 (n=6)	35.72 (n=5)	7.14 (n=1)	7.14 (n=1)	7.14 (n=1)
MH	9	77.78 (n=7)	11.11 (n=1)	11.11 (n=1)	-	-
UB	12	75 (n=9)	16.67 (n=2)	8.33 (n=1)	-	-
TD	11	36.36 (n=4)	36.36 (n=4)	18.18 (n=2)	9.1 (n=1)	-
Tak (Sakihama *et al.*, 1999)	72	54.17 (n=39)	18.06 (n=13)	19.44 (n=14)	2.78 (n=2)	5.56 (n=4)
Tak (Sakihama et al., 2006)	48	52.08 (n=25)	25 (n=12)	16.67 (n=8)	-	6.25 (n=3)

Table 4.2. The table showed the distribution pattern of MSP-1_{19} alleles in 5 provinces of Thailand in our study and the data of previous study in Tak province were included.

4.3 Discussions and conclusions

The data of genetic structure among *P. falciparum* population is the necessary information to develop fighting strategies to control the disease, for example, vaccine development, drug treatment. In this study, the genetic diversity of the *msp-1*$_{17}$ gene from Thailand was studied. The results showed that nearly all *P. falciparum* isolated collected during 2002 – 2009 may contain only one genotype of the *msp-1*$_{17}$ (monogenotype infection). These may reflect the reduction of malaria infection in the past ten years by the effective drug treatment and other malaria control program. The infected cases numbers in our studied locations, at last three years, were shown as described (Table. 3).

Areas	Number of infections		
	2007	2008	2009
K	1178	1210	333
RN	1538	1109	249
MH	1873	1453	489
UB	526	424	230
TD	335	248	201

Table 4.3. The table showed infected cases which reported during 2007-2009. The data indicated that number of infections, in all studied locations, were reduced. Especially, the reducing in Kanchanaburi (K), Ranong (RN) and Mae Hong Son (MH) was lower 50% from the first year, 2007, of the report (Department of Disease Control).

Although early studies of *msp-1* gene showed that the gene in *P. falciparum* may arrived from the recombination of two allelic types, K1 and MAD20 (Tanabe *et al.*, 1987), later studies on the block 2 of this gene revealed 4 distinct allelic groups; K1, MAD20, RO33 and MR (Tanabe *et al.*, 1987, Ferreira *et al.*, 1998, Takala *et al.*, 2002 and Happi *et al.*, 2004). On the other hand, our results and others suggested that the block 17 can be typed into 10 alleles worldwide (Sakihama *et al.*, 1999, Sakihama *et al.*, 2006, Sakihama *et al.*, 2007, Kaneko *et al*,. 1997, Ferreira *et al.*, 2003, Vijay Kumar *et al.*, 2005, Lalitha *et al.*, 1999 and Qari *et al.*, 1998) but only 5 alleles have been reported in Thailand (Sakihama *et al.*, 1999 and Sakihama *et al.*, 2006). Similarly, the *msp-1*$_{17}$ from GenBank showed that K1, MAD20 and RO33 isolates which have unique block 2 alleles, have Q/KNG/L, E/TSR/L and Q/KNG/L alleles, respectively, in their block 17.

In our study, five alleles in their MSP-*1*$_{19}$ sequences were found as previously reported in Tak province. Most of the malaria population in Tak showed the E/KNG/L as the dominant alleles, 54.17% (n=72), and also E/TSR/L (19.44%), Q/KNG/L (18.06%), Q/KNG/F (5.56%) and E/TSG/L (2.78%) [14]. However, another study of the same block of *msp-1* gene, in Tak province, revealed only 4 alleles (no E/TSG/L allele) while the E/KNG/L still remained the majority allele of MSP-*1*$_{19}$ (Sakihama *et al.*, 2007). From these reports and our finding indicated that E/KNG/L allele may be the most dominant allele of MSP-*1*$_{19}$ in Thailand.

Although, the E/KNG/L allele was marked as a dominant allele in Thailand, the difference of allele distribution in different areas and also in each year can be clearly seen. Our analysis of MSP-*1*$_{19}$allele distribution in each area with the χ^2 test indicated that the distribution pattern of all population were different. The uniqueness distribution pattern of each population should be actually collected from many distinct areas. One studied area dataset, which large sample size, was not good representative information of the population in large scale area such

as country and continent. The fluctuation of allele distribution should also be frequently determined.

Furthermore, a comparison with other parts of the world, the dominant allele of MSP-1_{19} in each regions are different. The major alleles in different geographic areas include; E/TSR/L (54.55%, n=77) in Solomon Islands (Sakihama *et al.*, 2007), E/KNG/L (58.06%, n=31) in Vietnam (Kaneko *et al.*, 1997) and E/KNG/L (65.22%, n=31) also in Vietnam (Ferreira *et al.*, 2003) E/TSG/L (81.25%, n=16) in India (Lalitha *et al.*, 1999) and E/TSR/L (33.33%, n=57) in Philippines (Sakihama *et al.*, 2007) and E/TSG/L (28.26%, n=92) in Iran (Mehrizi *et al.*, 2008). The major alleles in each areas appeared to exist in different proportion, from above 80% to as low as 33.33%. These data reflect different genetic background of falciparum malaria population. It also pointed out that a common vaccine, at the global level, may not be possible to design and to design an appropriate vaccine for a specific area; the genetic background of malaria population in the area must be thoroughly explored.

Isolates	Haplotype	Isolates	Haplotype
UB27	E/KNG/L	TD530	Q/KNG/L
UB7	E/TSR/L	TD529	E/KNG/L
UB14	E/KNG/L	TD531	Q/KNG/L
UB28	E/KNG/L	TD542	E/TSR/L
UB22	E/KNG/L	TD533	E/TSG/L
UB59	E/KNG/L	TD554	E/TSR/L
UB52	E/KNG/L	TD556	E/KNG/L
UB51	E/KNG/L	RN19	Q/KNG/L
UB50	E/KNG/L	RN31	E/KNG/L
UB85	Q/KNG/L	RN28	E/KNG/L
UB84	E/KNG/L	RN26	E/KNG/L

Isolates	Haplotype	Isolates	Haplotype
UB82	Q/KNG/L	RN72	E/KNG/L
K58	E/KNG/L	RN63	Q/KNG/L
K74	Q/KNG/L	RN70	E/KNG/L
K66	E/KNG/L	RN68	Q/KNG/L
K60	E/KNG/L	RN66	E/TSG/L
K64	E/KNG/L	RN130	Q/KNG/L
K165	Q/KNG/F	RN122	E/KNG/L
K205	E/KNG/L	RN129	Q/KNG/L
K215	E/KNG/L	RN133	Q/KNG/F
K185	E/TSG/L	RN131	E/TSR/L
K195	E/TSR/L	MH7	E/KNG/L
K397	E/KNG/L	MH11	E/KNG/L
K391	Q/KNG/F	MH6	E/KNG/L
K386	E/KNG/L	MH10	E/TSR/L
K392	Q/KNG/L	MH18	E/KNG/L
K389	E/TSG/L	MH20	E/KNG/L
TD510	Q/KNG/L	MH32	Q/KNG/L
TD504	Q/KNG/L	MH24	E/KNG/L
TD515	E/KNG/L	MH28	E/KNG/L
TD508	E/KNG/L		

Table 4.4. The table shows the conclusion of MSP-1_{19} haplotype of all studies isolates.

CHAPTER V

THE MEROZOITE SURFACE PROTEIN-2 GENE

5.1 Introduction

The MSP-2 sequence was first determined by the cDNA sequence of the parasite isolate namely FCQ27 (FC27) collecting from Papua New Guinea. The evidences from Triton X-114 and immunofluorescent microscopy indicated that this protein was on the merozoite membrane. The sequence comparison of C-terminus, uncharged hydrophobic region, between the MSP-2 and one of *Trypanosoma bruzi* glycoprotein showed high similarity between both sequences.,Because the *Trypanosoma bruzi* glycoprotein attached to its membrane by the glycosylphosphatidylinositol (GPI)-anchored tail, it had been proposed that the MSP-2 may also serve as an integral membrane protein which attaches to the merozoite surface using the same machanism (Smythe *et al*., 1988).

The dimorphic pattern of MSP-2 was reported after the sequence of 3D7 isolate has been determined (Smythe *et al*., 1990). The structure of MSP-2 could be divided to highly conserved and variable dimorphic regions (Figure 5.1). N-terminus and C-terminus were the conserved sequence of 15 hydrophobic amino acids and 17 hydrophobic amino acids regions, respectively. The central domain displayed repetitive amino acid sequence in dimorphic types, 3D7 and FC27 alleles. The FC27 allele demonstrated 32 amino acids repeat, while the 3D7 allele showed the Glysine, Serine and Alanine repeated sequence. The repetitive sequences of 3D7 and FC27 do not have either nucleotide or amino acid sequence relationship (Smythe *et al*., 1990)

Worldwide sequences of the MSP-2 showed some variations of the amino acid chain. The recombination, between FC27 and 3D7 alleles, was also found in Africa (Ferreira *et al.*, 2007) and Myanmar (Kang *et al.*, 2010). In Thailand, however, the malaria isolates from Tak (North-western part) and Trat (Eastern part) provinces demonstrated that no hybrid between 2 allelic variants occurred and most isolates were determined as 3D7 (Snounou *et al.*, 1999 and Putapornthip *et al.*, 2008). In another report, all MSP-2 of *Plasmodium falciparum* population in Narathiwat and Yala (Southern part of Thailand) were also 3D7 type (Jongwutiwes *et al.*, 2010).

Because the MSP-2 was able to recognize by the human immune system, especially at its repetitive unit (Smythe *et al.*, 1988, Smythe *et al.*, 1990), this part of MSP-2 was considered as one of notable *P. falciparum* candidate antigens for vaccine development (Lawrence *et al.*, 2000). Several reports indicated that (a) MSP-2 is recognized by naturally acquired immunity of malaria patients in Kenya, Africa (Polley *et al.*, 2006) and (b) this protein also bound to vaccine-induced antibodies (Fluck *et al.*, 2004).

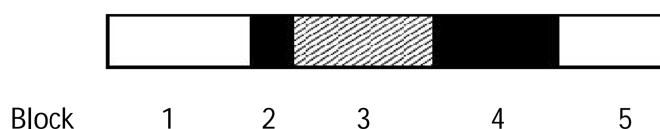

Figure 5.1. The picture shows putative structure of MSP-2. This protein contains 5 domains of 2 conserved blocks (block 1 and block 5) as open boxes. The central region composed of a polymorphic repeats sequence (block 3, stripped box), flanked with 2 non-repetitive dimorphic domains (block 2 and block 4, black boxes).

In this chapter, the *msp*-2 gene of 61 isolates from Thailand was amplified; then, their amplified products were checked by agarose gel electrophoresis after ethidium bromide staining. The sequences were analyzed. The proportion of 2 alleles and the diversity of variant among those allele families were identified. Moreover, the evolutionary trend of this gene among Thailand population and other region of the world were determined with the phylogenetic tree, constructed by the PHYLIPS program.

5.2 Results

5.2.1 *msp*-2 gene amplification

The multiple bands PCR products and size variation were detected by agarose gel electrophoresis (Figure 5.2). Eleven isolates of total 61 isolates gave the multiple bands PCR products. It suggested that these isolates may contain multiple clones. So, they were excluded from DNA sequencing process. The predicted size of our PCR product, from 3D7 sequence, was 819 bp but the results showed size variation of amplicon between 800-1,200 bp (comparison with 100 bp ladder markers, Bioexcellent).

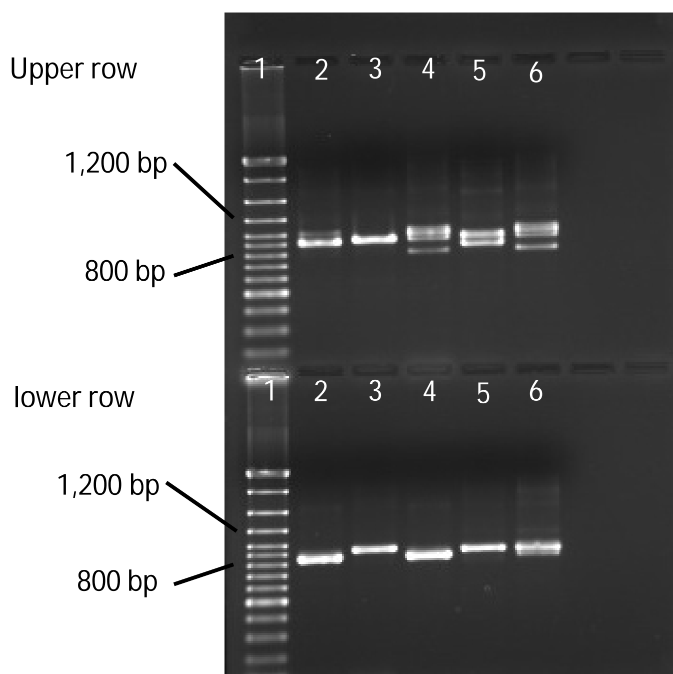

Figure 5.2. The picture shows some multiple band products and size variation of *msp*-2 genes of parasite isolates from Mae Hong Son and Trat. Lane 1, in upper and lower rows, are 100 bp ladder markers (Bioexcellent). The upper row reveals the PCR product of parasite isolates from Mae Hong Son, MH11 (lane2), MH24 (lane 3), MH9 (lane 4), MH6 (lane 5) and MH7 (lane 6). The lower row shows the PCR product of parasite isolates from Trat, TD508 (lane 2), TD530 (lane 3), TD529 (lane 4), TD504 (lane 5) and TD510 (lane 6).

5.2.2 MSP-2 sequence analysis

From the *msp*-2 sequence data, there were several types of mutation compared to the *msp*-2 sequence of the 3D7 isolate. Some of the mutations caused changes in amino acid residues while some mutations were synonymous mutations. For simplicity, the *msp*-2 gene sequences were translated into amino acid residue and then aligned by BioEdit program. The MSP-2 sequences revealed both conserved and highly variable regions. The N and C-terminus

sequences (block 1 and block 5) were mostly conserved with some amino acid changes. The 3D7 allele and FC27 allele sequences considerably diverged at the central region. The amino acid sequence of both MSP-2 allele families are shown in the picture below (Figure 5.3).

Figure 5.3 The picture shows the MSP-2 amino acid sequences of FC27 (A) and 3D7 (B) alleles. Block 1 and block 5 sequences are conserved in both allelic types. The central region, including block 2 – 4, displays dimorphic patterns.

5.2.2.1 The terminus sequence of MSP-2

The N and C- terminus sequences of 48 isolates were analyzed. Two samples were excluded from N-terminus sequence analysis because their N-terminus sequences were not completed. The 18 amino acids in N-terminus region (block 1), YSNTFINNAYNMSIRRSM, of the MSP-2 were highly conserved and were identical in all samples. The C-terminus region (block 5) consisted of 53 amino acids most of which were conserved. The 48 sequences of C-terminus showed 2 amino acid substitutions; T/P and N/S. The amino acid change produced 3 allelic

patterns; T-N, P-S and T-S, in this region (Figure 5.4). The frequencies of 3 alleles were 87.5% (5C1 allele, T-N, n=42), 10.42% (5C2 allele, P-S, n=5) and 2.08% (5C3 allele, T-S, n=1).

```
5C1     ----PENKGTGQHGHMHGSRNNHPQNTSDSQKECTDGNKENCGAATSLLNNSSNIAS----
5C2     ----.........................................P...S.......----
5C3     ----.............................................S.......----
```

Figure 5.4 the picture shows 3 alleles of C-terminus sequences. By comparison with allele 5C1, the identical amino acids to allele 5C1 are referred by dot (.) symbol. The dash (-) symbol refers to upstream and downstream sequences which do not show here.

The central region consisted of consequently block 2, block 3 and block 4, which were dimorphic. In 3D7 and FC27 families, the block 2 and block 4 were non-repetitive regions, interrupted by a highly variable region, block 3. The block 3 was composed of repetitive regions with a repeat unit or multiple sets of a repeat unit and a non-repetitive region.

5.2.2.2 The FC27 family

From 50 isolates, the MSP-2 sequences of 17 isolates were identified as FC27 allelic type. The dimorphic block 2 sequences, 16 amino acids, revealed amino acid substitutions at 5 positions. These substitutions resulted to 5 allele patterns in this region (Figure 5.5). The numbers of sample which identified to those alleles were 1 sample (allele F2A, NS-D-K-N), 4 samples (allele F2B, NR-G-N-K), 4 samples (allele F2C, TS-G-N-N), 1 sample (allele F2D, TS-G-N-K) and 6 samples (allele F2E, KS-G-N-K).

```
F2A    ----ANEGSNTNSVDAKAPN----
F2B    ----........R.G.N..K----
F2C    ----.......T..G.N...----
F2D    ----.......T..G.N..K----
F2E    ----.......K..G.N..K----
```

Figure 5.5 The picture shows MSP-2 block 2 sequence of FC27 family. By comparison with allele F2A, the identical amino acids to allele F2A are referred by dot (.) symbol. The dash (-) symbol refers to upstream and downstream sequences which do not present in the picture.

The highly variable region, block 3, of 17 samples were grouped into different alleles by their sequence similarities (Figure 5.7). The conserved non-repetitive sequence, ADTPTAT, located in the middle of this block. This non-repetitive unit was flanked by 2, N and C terminus, repetitive regions. The N terminus repetitive locus of block 3 had 1 or 3 repeats of 32 amino acids, ADTIASGSQSSTNSASTSTTNNGESQTTTPTA which can be grouped into 2 allelic type (F3N1 and F3N2). The F3N1 allele had 3 repeat units with an amino acid substitution from SQSS to SQRS, in both the second and third repeats. No multiple repeats were found in the F3N2 alleles. The C terminus repetitive region showed 1 to 5 repetitive units of ESNSPSPPITTT. Some amino acid substitutions, shown as picture below, are detected at this locus which can be grouped into 7 allelic types, F3C1 through F3C6.

```
A.
    F3N1    ADTIASGSQSSTNSASTSTTNNGESQTTTPTA ADTIASGSQRSTNSASTSTTNNGESQTTTPTA ADTIASGSQRSTNSASTSTTNNGESQTTTPTA (n=1)
    F3N2    ................................ -------------------------------- -------------------------------- (n=16)

B.
    F3C1    ESNSPSPPITTT ------------ ------------ ------------ ------------  (n=1)
    F3C2    K........... ESNSPSPPITTT ESNSPSPPITTT ------------ ------------  (n=9)
    F3C3    K........... ............ ............ ....P....... ------------  (n=1)
    F3C4    ....R....... ....R....... ....R....... ESNSRSPPITTT ------------  (n=1)
    F3C5    ....R....... ....R....... ....R....... ------------ ------------  (n=1)
    F3C6    ....R....... ....R....... ....R....... ............ ESNSRSPPITTT  (n=4)
```

Figure 5.6 The picture shows the N-terminus (A.) and C-terminus (B.) of MSP-2 repetitive block 3 sequences of FC27 family. By comparison with allele F3N1 or F3C1 sequence, the dot (.) symbol refers to the identical amino acid and the dash (-) symbol refers to missing sequence. The conserved non-repetitive sequence, ADTPTAT, is not shown in this figure.

The non-repetitive block 4 sequence of the FC27 family, 49 amino acids, demonstrated 2 positions of amino acid change; E to K and Q to K. These substitutions produced 2 allele patterns, F4A (E-Q) and F4B (K-K) as showed below (Figure 5.6). The proportion between 2 alleles was 12/17 and 5/17 for F4A and F4B, respectively.

```
F4A    ----ESSSSGNAPNKTDGKGEESEKQNELNESTEEGPKAPQEPQTAENENPAA----
F4B    ----...................K.K...........................----
```

Figure 5.7 The picture shows MSP-2 block 4 sequence of FC27 family. By comparison with allele F4A, the identical amino acids are referred by dot (.) symbol. The dash (-) symbol refers to upstream and downstream sequences which do not present in the picture.

5.2.2.3 The 3D7 family

The MSP-2 sequences of 33/50 isolates were identified as 3D7 family. The block 2 sequences of this group were characterized as non-repetitive sequence of 5 amino acids, followed by 1 to 3 repeat units of 2 amino acids (Figure 5.8). The non-repetitive loci of the block 2 sequences revealed 1 - 4 amino acid substitution positions which could be divided into 9 allelic types, D2N1 (EE-NP), D2N2 (TE-NP), D2N3 (TE-KP), D2N4 (TE-KT), D2N5 (AE-KT), D2N6 (AE-KP), D2N7 (AV-KP), D2N8 (AV-NP) and D2N9 (KE-KP) . The following repetitive locus contained 1 to 3 repeats of PT amino acid and the substitution mutation of PT to ST was detected in some isolates; D2C1 (PT), D2C2 (PTPT), D2C3 (PTPTPT) and D2C4 (ST).

A.

```
D2N1    EESNP
D2N2    T....
D2N3    T..K.
D2N4    T..KT
D2N5    A..KT
D2N6    A..K.
D2N7    AV.K.
D2N8    AV...
D2N9    K..K.
```

B.

```
D2C1    PT----
D2C2    ..PT--
D2C3    ....PT
D2C4    S.----
```

Figure 5.8. The pictures show the N-terminus (A.) and C-terminus (B.) non-repetitive regions of MSP-2 block 2 sequences of 3D7 family. The dot (.) symbol refers to identical amino acid. The dash (-) symbol refers to missing amino acid.

From 33 isolates, block 3 sequences of 3D7 family could be arranged into many distinct alleles. These block 3 sequences consisted of 3 distinct parts, including a non-repetitive region, flanking with repetitive regions (Figure 5.9, 5.10 and 5.11). The N terminus repetitive locus of block 3 revealed highly variable sequences most of which were glysine (G), serine (S) and alanine (A) amino acids. Some alleles also contained aspartic (D) amino acid in their repetitive

sequences. There were 21 distinct types in this N terminus region, from D3N1 through D3N21 (Figure 5.9). The 15 amino acids sequence of non-repetitive locus, GNGANPGADAERSPS, in the middle of block 3, seemed to be semi-conserved because some alleles showed amino acids substitutions and deletions. The substitutions of this 15 amino acids region occurred in 3 positions, including positions 1, 11, 12 and 14. Those amino acids changes were G/R at position 1, E/K at position 11, R/G at position 12 and P/S at position 14. These polymorphisms could be sorted into 8 allelic types, from D3M1 through D3M8 (Figure 5.10). At the C terminus of the repetitive region, the sequencing reaction revealed 1 or 8 repeat units of TPA sequence which was followed by 8, 11 or 14 Threonine (T) residues. These variants can be separated into different allelic types. Those allelic types in the TPA region were named as D3CR1 and D3CR2. Those in the T region were called D3CT1, D3CT2, and D3CT3 (Figure 5.11).

```
D3N1    GAGAVAGSGAGAVAGSGAGAVAGSGAGAGAGAVAGSGAGAGAGAVAGSGAGAGAVAGSGAGASA--------------------------------
D3N2    GAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGSGD------------------------------------------
D3N3    GASGSAGAGASGSAGASGSAGASGSAGASGSAGASGSAGASGSAGASGSAGASGSAGSGA-------------------------------------
D3N4    GASGSAGSGSGAVASA-----------------------------------------------------------------------------
D3N5    GASGSAGSGAVASAGSGAVASAGSGAVASA-------------------------------------------------------------------
D3N6    GAGAVAGSGAGAGAVAGSGAGAGAVAGSGAGAGAGAGAVAGSGAGAGAGAVAGSGAGAGAVASA--------------------------------
D3N7    GASGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGS
D3N8    GAGAGAGASGSAGSGDGASGSASGSAGASGSASGSAGASGSASGSAGASGSASGSAGASGSASGSAG-----------------------------
D3N9    GAGASGRAGSGDGAVASAGSGDGAVASAGSGDGAVASA-------------------------------------------------------
D3N10   GAGAVAGSGAGAVAGSGAVAGSGAGAGAVAGSGAGAGAGAVAGSGAGAGAVAGSGAGASA-------------------------------------
D3N11   GASGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGSGD--------------------------------------------------
D3N12   GAGAGASGRAGSGDGAVASAGSGDGAVASA-------------------------------------------------------------------
D3N13   GAVAGSGAGAGAVAGSGAGAGAVAGSGAGAGAGAGAVAGSGAGAGAVAGSGAGAVASA--------------------------------------
D3N14   GAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGGSGSAGSGA--------------------------------------------------
D3N15   GAGAVAGSGAGAVAGSGAGAVAGSGAGAGAGAVAGSGAGAGAVAGSGAGAGAVAGSGAGASA-----------------------------------
D3N16   GASGSAGAGASGSAGAGASGSAGASGSAGAGASGSAGAGASGSAGAGASGSAGAGASGSAGAGASGSAGASGSAGAGASGSAGSGA-----------
D3N17   GASGSAGAGASGSAGAGASGSAGASGSAGAGASGSAGAGASGSAGAGASGSAGAGASGSAGAGASGSAGAGASGSAGAGASGSAGSGA---------
D3N18   GASGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGSGD------------------
D3N19   GASGSAGAGASGSAGAGASGSAGASGSAGASGSAGASGSAGASGSAGASGSAGASGSAGASGSAGSGA----------------------------
D3N20   GTGASGSAGSGAGASGSAGSGDGAVASA---------------------------------------------------------------------
D3N21   GASGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGGSAGSGD-------
```

```
D3N1    --------------------  (n=1)
D3N2    --------------------  (n=2)
D3N3    --------------------  (n=1)
D3N4    --------------------  (n=1)
D3N5    --------------------  (n=1)
D3N6    --------------------  (n=2)
D3N7    AGGSAGGSAGGSAGGSAGGS  (n=1)
D3N8    --------------------  (n=4)
D3N9    --------------------  (n=1)
D3N10   --------------------  (n=2)
D3N11   --------------------  (n=1)
D3N12   --------------------  (n=1)
D3N13   --------------------  (n=5)
D3N14   --------------------  (n=1)
D3N15   --------------------  (n=1)
D3N16   --------------------  (n=1)
D3N17   --------------------  (n=1)
D3N18   --------------------  (n=1)
D3N19   --------------------  (n=1)
D3N20   --------------------  (n=3)
D3N21   --------------------  (n=1)
```

Figure 5.9 The picture shows the N-terminus of repetitive region in 3D7 block 3 sequence. The dot (.) symbol refers to identical amino acid. The dash (-) symbol refers to the missing amino acid in the sequence

```
D3M1    GNGANPGADAERSPS
D3M2    ....----..K....
D3M3    --------.......
D3M4    ...........G.S.
D3M5    ---------..G...
D3M6    R.........K....
D3M7    ..........K....
D3M8    R..........G.S.
```

Figure 5.10 The picture shows the non-repetitive region of block 3 sequences of 3D7 family. The dot (.) symbol refers to identical amino acid. The dash (-) symbol refers to the missing amino acid in the sequence.

```
A.                                              B.

  D3CR1    TPA---------------------                 D3CT1    TTTTTTTT------
  D3CR2    ...TPATPATPATPATPATPATPA                 D3CT2    ...........---
                                                    D3CT3    ........TTTTTT
```

Figure 5.11 The picture shows C-terminus repetitive region, including TPA region (A.) and T region (B.), of 3D7 family.

In block 4 of 3D7 family, the amino acid sequences showed conserved sequence with some amino acid substitutions, except in some isolates in which 11 amino acids, PHGNGGVQKPN, were deleted at the center of the block (Figure 5.12). This suggested that there may be at least 11 alleles in this region, from D4D1 through D4D11.

```
D4D1    NDAEASTSTSSENPNHNNAETN PKGKGEVQKPN QANKETQNNSNVQQDSQTKSNVPPTQDADTKSPTAQPEQAENSAPTAEQTESPELQSA  (n=3)
D4D2    ...................K.. ...N.G..E.. ..........................................................  (n=5)
D4D3    ................K..... ........... ..........................................................  (n=7)
D4D4    ...................K.. ...N.G..... ..........................................................  (n=2)
D4D5    ...................... .....Q..E.. ..........................................................  (n=4)
D4D6    .............S........ ........E.. ..........................................................  (n=1)
D4D7    ...................... ----------- ..........................................................  (n=4)
D4D8    ...................... ........EQ. ..........................................................  (n=1)
D4D9    ...................K.. ...N.G..E.. K..T......................................................  (n=1)
D4D10   ...................K.. ........... ..........................................................  (n=2)
D4D11   ...................... ...N.K..E.. ..........................................................  (n=3)
```

Figure 5.12. The pictures show the MSP-2 block 4. The dot (.) symbol refers to identical amino acid of allele D4D1. The gaps between sequences use to divide each repetitive and non-repetitive locus.

5.2.3 Allele frequency of MSP-2

The sequences of 50 isolates showed 2 distinct allele families. No hybrid between these 2 families was detected. The MSP-2 allele families of *Plasmodium falciparum* population of Thailand, collected during 2002-2009, were 3D7 (66%, n=33) and FC27 (34%, n=17). In comparison, the 3D7 frequencies in all 3 collecting periods, including 2002-2003, 2005-2006 and 2008-2009, were higher than FC27 family (Table 5.1). These results indicated that the dominant

allele family during studied periods is 3D7. It suggested that the 3D7 family was the major allele family of MSP-2 among the parasite population in Thailand.

Period	Number of isolates	Allele frequency (%)	
		3D7	FC27
2002-2003	16	68.75	31.25
2005-2006	20	65	35
2008-2009	14	64.29	35.71

Table 5.1 The table shows the 3D7 and FC27 allele families frequencies in each studied period.

Eventhough, the 3D7 allelic type seemed to be the dominant allele of MSP-2 in Thailand, the 3D7 and FC27 distribution pattern in 5 provinces were different. Paraiste samples from Ubonratchathani displayed FC27 as a major allele (72.73%) whereas the other areas (Kanchanaburi, Ranong, Mae Hong Son and Trat) had 3D7 allele as the majority with different frequency (Table 5.2). The χ^2 test value of the allele distribution showed the significant higher than critical value at $p < 0.001$ (degree of freedom = 4). This result indicated that the MSP-2 allele distribution, among the parasite populations in Thailand, was area specific. The allele frequencies of MSP-2 among *P. falciparum* population in 5 provinces were shown below.

Allele	Allele frequency (%)				
	K (n=13)	RN (n=12)	TD (n=9)	UB (n=11)	MH (n=5)
FC27	23.08	33.33	0.00	72.73	40.00
3D7	76.92	66.67	100.00	27.27	60.00

Table 5.2. The table shows the allele frequency of MSP-2 in 5 studied areas including Kanchanaburi (K), Ranong (RN), Trat (TD), Ubonratchathani (UB) and Mae Hong Son (MH).

Although the allele frequency between 2 families at each period seemed to be stable, the frequency in each area showed some fluctuation between 2 allele types (Table 5.3). In Kanchanaburi and Trat, the frequency of 3D7 family is still higher than FC27 in all periods. In contrast, the allele frequency in Ranong demonstrates that 3D7 was the major allele in 2002-2003 but FC27 was the dominant allele in 2005-2006. The frequency in Ubonratchathani also shows allele switching, similar to Ranong and Mae Hong Son in 2005-2006. More samples may be needed to concrete any hypothesis.

Areas	2002-2003		2005-2006		2008-2009	
	3D7	FC27	3D7	FC27	3D7	FC27
K (n=13)	80 (4)	20 (1)	75 (3)	25 (1)	75 (3)	25 (1)
RN (n=12)	100 (3)	0	25 (1)	75 (3)	80 (4)	20 (1)
TD (n=9)	100 (3)	0	100 (3)	0	100 (3)	0
UB (n=11)	25 (1)	75 (3)	50 (2)	50 (2)	0	100 (3)
MH (n=5)	0	100 (1)	75 (3)	25 (1)	0	0

Table 5.3 The table shows 3D7 and FC27 alleles frequencies in 3 interval periods .

5.2.4 Evolutionary relationship of *msp*-2 gene in Thailand population

The 3D7 sequence from Genbank database (accession number was described in chapterIII) served as an outgroup sequence in phylogenetic tree analysis. More *msp*-2 sequences from other 4 isolates, including Vietnam, Papua New Guinea, Ghana and Iran, were added to sequence input data. The result from DNAml (maximum likelihood method) showed the distinct 2 clades of 3D7 and FC27 families (Figure 5.10). The branch distance of FC27 family was far from another group (3D7). This clearly displayed the differences between these 2 families' sequences. Other 4 outgroups were also located in both of 3D7 and FC27 clades. The DNApenny (most

parsimony method) demonstrated 2 different clades of 3D7 and FC27 families, similar to DNAml results (Figure 5.11). Surprisingly, some isolates such as MH20 (3D7) and RN68 (FC27) were located apart from its own clades by the parsimony analysis. Both of maximum likelihood and most parsimony results indicated that there is no relationship between geographic pattern and distribution of these genes in Thalland.

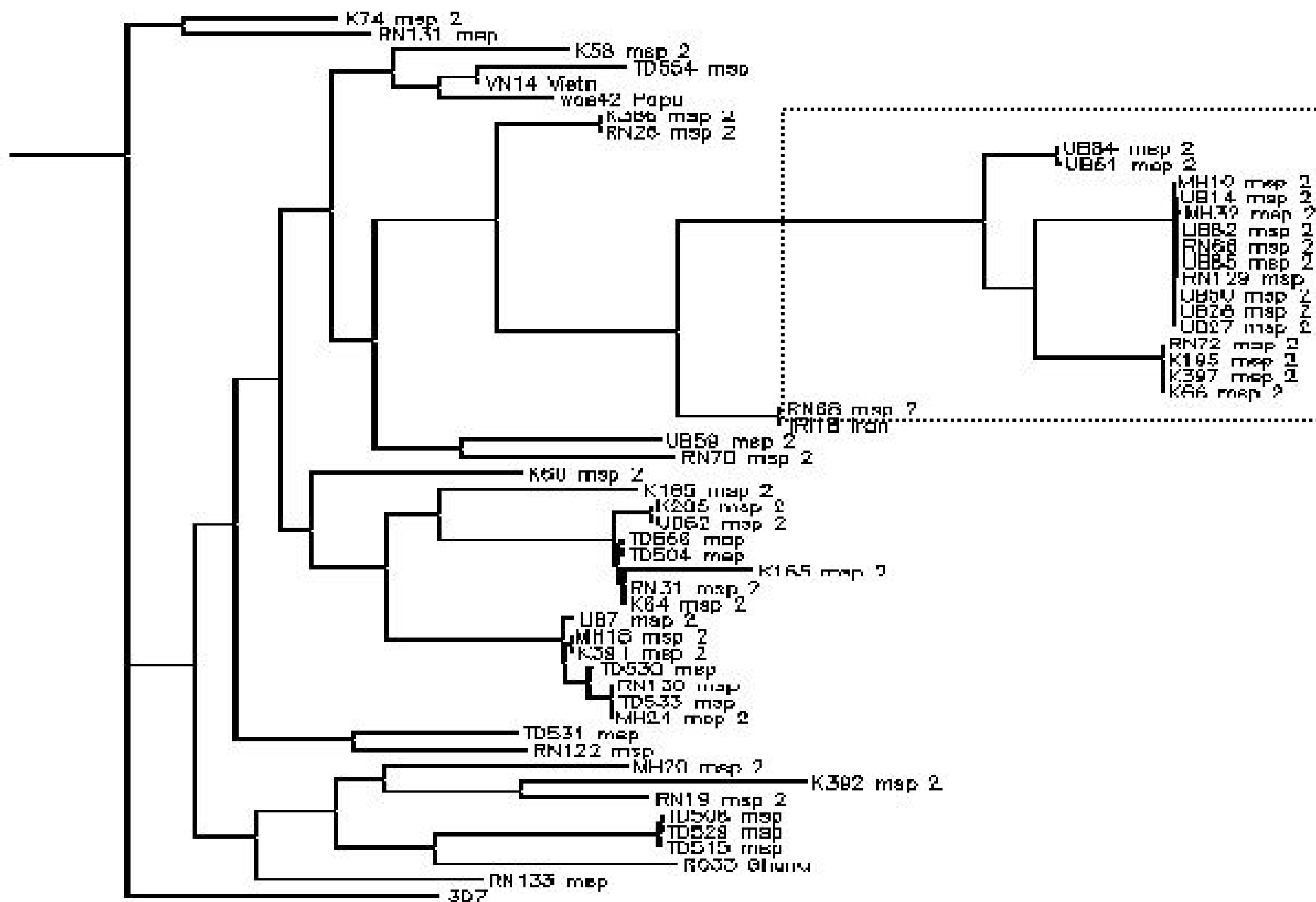

Figure 5.13 This picture shows the maximum likelihood reconstructed tree. The FC27 clade of our samples is enclosed by dash line square box.

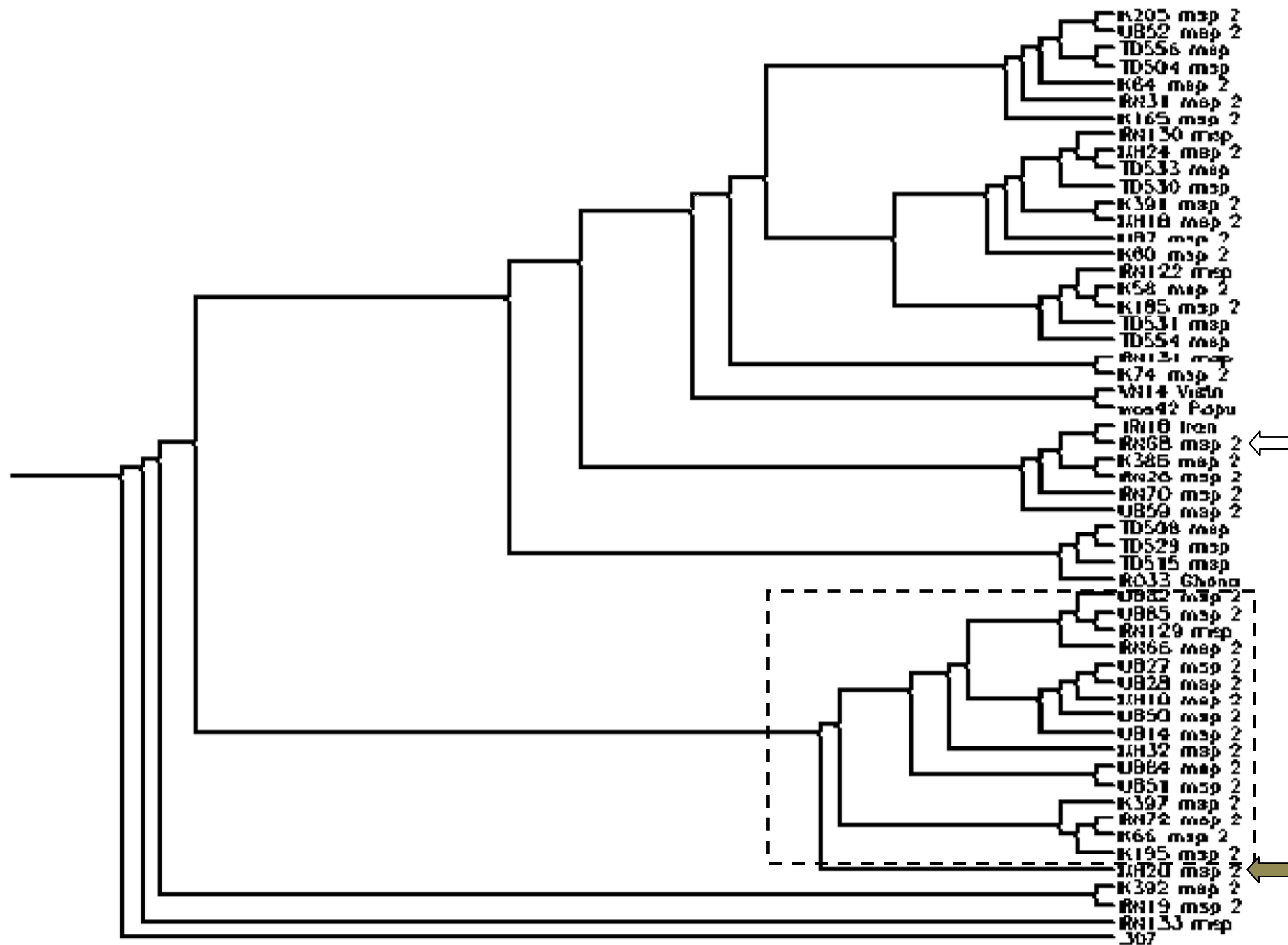

Figure 5.14 This picture shows the most parsimony consensus tree. The FC27 clade of our samples is enclosed by dash line square box. The blank arrow shows RN68 isolates (FC27) located on 3D7 clade, while, the dark arrow shows MH20 isolate (3D7) located on FC27 clade.

5.3 Discussions

Agarose gel analysis showed that the *msp*-2 gene of parasite isolates from Thailand carried size variation. Some isolates gave multiple band PCR products, resulted from *msp*-2 gene variation in each clone members. In this study, 50 isolates gave a single band of PCR product which varied in size from 800 to 1,000 bp. From the sequence results, using directly sequencing procedure, the electropherogram showed no signal interference in both forward and reverse strands. The full length sequence analysis of 3D7 and FC27 families revealed several allele variants which caused by substitution mutation, insertion or deletion in the gene structure. Thirty-three isolates of 3D7 family showed 26 variants and the 17 isolates of FC27 family revealed 9 variants. Most of these variants were found in a single isolate. The proportion of the variant from a single isolate in both of 3D7 and FC27 familes was 74.28% (n = 35 variants). This also supports our hypothesis that most of *P. falciparum* isolates in Thailand were clones (ref).

According to 26 variants of 3D7 family, only 6 variants were found in few isolates whereas other 20 variants were found in only one isolate. The 6 variants consisted of identical sequences which were 3D7-1 (MH24, RN130 and TD533), 3D7-2 (K386 and RN26), 3D7-3 (TD529 and TD508), 3D7-4 (RN31 and K165), 3D7-5 (TD504 and TD556) and 3D7-6 (MH18 and K391). Among these variants, the member of 3D7-2, 3D7-3, 3D7-4 and 3D7-5 variants were collected from the same or nearby provinces. On the other hand, the FC27 variant also showed similar results. Three variants of FC27 family were found in a few isolates, while in other 6 variants, each variants was found in an isolate. Those 3 variants were FC27-1 (K66, K195, K397 and RN72), FC27-2 (RN66, RN129 and UB85) and FC27-3 (UB27, UB50, UB28 and MH10) and. It should be pointed out that only the member of FC27-1 and FC27-3 variants were collected from the same or nearby provinces.

In this study, the 3D7 allele proportion was higher than FC27 allele during 2002-2009. This result indicated that the 3D7 was the major allele of *msp*-2 in Thailand. Although, the proportion of two MSP-2 alleles in Thailand (2002-2009) showed the dominance of 3D7, the allele proportion in each area was different. In the Western provinces, including Mae Hong Son, Kanchanaburi and Ranong, revealed the majority of 3D7 whereas the Eastern region such as Ubonratchathani showed large proportion of FC27. Surprisingly, in Trat province, located in the eastern part, demonstrated that the 3D7 was the dominant allele. This may cause by the human migration from Western regions to Trat province. Because of gem mining and gem trate in the past 20 years, many immigrants from Myanmar migrated to Trat for job carriers (Snounou *et al.*, 1999). These may explain the similarity of MSP-2 allele frequency in Trat province and other 3 western provinces.

The hybrid between 3D7 and FC27, as previously described (Ferreira et al., 2007), was not found in this study. Similar study in *P. falciparum* isolates from Asian countries such as Vietnam, India and China also showed no evidence of allele hybridization in these countries (Ferreira et al., 2007). On the contrary, the *msp*-2 gene study in Myanmar revealed large proportion of 3D7/FC27 hybrid (68.3%, n=63) sequence (Kang et al., 2010). The difference may cause by the distance between 2 studied locations. Those 63 isolates in the previous study were collected from malaria patients in Mandalay Division which located in the central part of Myanmar. On the other hand, the parasite samples in this study, which supported by College of Public Health Sciences, Chulalongkorn University, were collected from malaria foreigner patients in 3 western provinces of Thailand. These suggested that the human migration influences to geographic distribution pattern of *msp*-2 gene. Because of the distance between 2 human populations are so far, people cannot easily move across the distance to other locations. As a result, the gene pool of parasite cannot flow to other parasite populations via human migration.

So, the allelic pattern of the *msp*-2 gene of 2 parasite populations, Mandalay and 3 western provinces were different.

The variation in the msp-2 gene sequences was analyzed with the maximum likelihood method, using 4 isolates as the outgroups. From the sequences of 4 outgroups (VN14, Wos42, IRI18 and RO33), the VN14, Wos42 and RO33 were typed as the 3D7 family, whereas IRI18 sequence was FC27 family. The result showed the FC27 isolates were separated from 3D7 gene family. The outgroup sequences, including VN14 (Vietnam), Wos42 (Papua New Guinea), IRI18 (Iran) and RO33 (Ghana) were located in the branch of their similar alleles. Only the sequence of Iran isolate was in FC27 clade. These may cause by the outgroup sequence which used as the reference during analysis, the 3D7. This procedure separated between 3D7 clone similar sequences (3D7 isolates) and 3D7 clone dissimilar sequences (FC27 isolates). Both of the maximum likelihood and most parsimonious trees indicated that the *msp*-2 gene in 5 populations of Thailand was related and 2 allele families could be separated from each other.

Isolates	Allelic types					
	Block 1	Block 2	Block3		Block 4	Block5
			N-terminus	C-terminus		
RN68	1N1	F2A	F3N1	F3C1	F4A	5C1
UB85	1N1	F2B	F3N2	F3C2	F4A	5C1
UB27	1N1	F2C	F3N2	F3C2	F4B	5C1
UB14	1N1	F2D	F3N2	F3C2	F4A	5C1
UB28	1N1	F2C	F3N2	F3C2	F4B	5C1
UB82	1N1	F2B	F3N2	F3C2	F4A	5C2
UB50	1N1	F2C	F3N2	F3C2	F4B	5C1
UB51	1N1	F2E	F3N2	F3C4	F4A	5C1
UB84	1N1	F2C	F3N2	F3C3	F4B	5C1
MH32	1N1	F2E	F3N2	F3C5	F4A	5C1
MH10	1N1	F2C	F3N2	F3C2	F4B	5C1
RN72	1N1	F2E	F3N2	F3C6	F4A	5C1
RN66	1N1	F2B	F3N2	F3C2	F4A	5C1
RN129	1N1	F2B	F3N2	F3C2	F4A	5C1
K66	1N1	F2E	F3N2	F3C6	F4A	5C1
K397	1N1	F2E	F3N2	F3C6	F4A	5C1
K195	1N1	F2E	F3N2	F3C6	F4A	5C1

Table 5.4. The table shows summary of block haplotype of FC27 isolates.

Isolates	Allelic types								
	Block 1	Block 2		Block 3				Block 4	Block 5
		N-terminus	C-terminus	N-terminus	Middle	TPI region	T region		
UB7	1N1	D2N1	D2C1	D3N1	D3M2	D3CR1	D3CT1	D4D5	5C2
UB52	1N1	D2N1	D2C4	D3N2	D3M1	D3CR1	D3CT1	D4D1	5C1
UB59	1N1	D2N2	D2C1	D3N3	D3M3	D3CR1	D3CT1	D4D4	5C1
MH24	1N1	D2N4	D2C2	D3N8	D3M5	D3CR1	D3CT1	D4D2	5C1
MH20	1N1	D2N9	D2C1	D3N4	D3M1	D3CR2	D3CT2	D4D9	5C1
K392	1N1	D2N5	D2C1	D3N5	D3M4	D3CR1	D3CT3	D4D2	5C1
MH18	1N1	D2N8	D2C1	D3N6	D3M2	D3CR1	D3CT1	D4D3	5C1
RN122	1N1	D2N2	D2C1	D3N7	D3M1	D3CR1	D3CT1	D4D7	5C1
RN70	1N1	D2N4	D2C3	D3N8	D3M5	D3CR1	D3CT1	D4D2	5C1
RN130	1N1	D2N4	D2C2	D3N8	D3M5	D3CR1	D3CT1	D4D2	5C1
RN133	1N1	D2N5	D2C1	D3N9	D3M6	D3CR1	D3CT1	D4D6	5C1
RN26	1N1	D2N1	D2C1	D3N10	D3M2	D3CR1	D3CT1	D4D5	5C1
RN131	1N1	D2N2	D2C1	D3N11	D3M1	D3CR1	D3CT1	D4D7	5C1
RN19	1N1	D2N7	D2C1	D3N12	D3M4	D3CR1	D3CT1	D4D1	5C1
RN31	1N1	D2N8	D2C1	D3N13	D3M2	D3CR1	D3CT1	D4D3	5C1
K74	1N1	D2N5	D2C1	D3N14	D3M7	D3CR1	D3CT1	D4D8	5C1
K60	1N1	D2N1	D2C1	D3N15	D3M2	D3CR1	D3CT1	D4D5	5C2
K58	1N1	D2N2	D2C1	D3N16	D3M3	D3CR1	D3CT2	D4D10	5C1
K165	1N1	D2N8	D2C1	D3N13	D3M2	D3CR1	D3CT1	D4D3	5C1
K64	1N1	D2N8	D2C1	D3N13	D3M2	D3CR1	D3CT1	D4D3	5C1
K391	1N1	D2N8	D2C1	D3N6	D3M2	D3CR1	D3CT1	D4D3	5C1
K386	1N1	D2N1	D2C1	D3N10	D3M2	D3CR1	D3CT1	D4D5	5C1
K185	1N1	D2N2	D2C1	D3N17	D3M3	D3CR1	D3CT2	D4D10	5C1
K205	1N1	D2N1	D2C4	D3N2	D3M1	D3CR1	D3CT1	D4D1	5C1
TD554	1N1	D2N2	D2C1	D3N18	D3M1	D3CR1	D3CT1	D4D7	5C1
TD504	1N1	D2N8	D2C1	D3N13	D3M2	D3CR1	D3CT1	D4D3	5C2
TD556	1N1	D2N8	D2C1	D3N13	D3M2	D3CR1	D3CT1	D4D3	5C2
TD530	1N1	D2N2	D2C1	D3N19	D3M3	D3CR1	D3CT1	D4D4	5C1

TD515	1N1	D2N6	D2C1	D3N20	D3M8	D3CR1	D3CT3	D4D11	5C3
TD529	1N1	D2N3	D2C1	D3N20	D3M8	D3CR1	D3CT3	D4D11	5C1
TD508	1N1	D2N3	D2C1	D3N20	D3M8	D3CR1	D3CT3	D4D11	5C1
TD531	1N1	D2N2	D2C1	D3N21	D3M1	D3CR1	D3CT1	D4D7	5C1
TD533	1N1	D2N4	D2C2	D3N8	D3M5	D3CR1	D3CT1	D4D2	5C1

Table 5.5. The table shows summary of block haplotype of 3D7 isolates.

CHAPTER VI

THE MEROZOITE SURFACE PROTEIN 3 GENE

6.1 Introduction

The merozoite surface protein-3 (MSP-3) is one of many vaccine candidate proteins. This 48 kDa polypeptide was first described as a secreted polymorphic antigen associated with merozoite (SPAM) (McColl *et al.*, 1994). However, when the SPAM and MSP-3 proteins were compared, the result shown that both peptides were actually the same protein (Oeuvray *et al.*, 1994). Later, the SPAM had been referred as MSP3 (McColl and Anders, 1997). From further studies, it was found that the MSP-3 was located on the merozoite and was detected in the late schizont extract (McColl *et al.*, 1994 and Oeuvray *et al.*, 1994). The location on the merozoite surface of this protein had been supported by Triton X-114 extraction and immunoelectron microscopy evidences (Oeuvray *et al*,. 1994). The protein was synthesized as the precursor, then, it was secreted to the parasitophorous vacuole. In the parasitophorous vacuole, the MSP-3 was proteolytic processed and produced a 48 kDa antigen (McColl *et al.*, 1994). The function of MSP-3 was still unclear but it was believed to help holding the merozoite membrane structure, in associated with other surface proteins (McColl *et al.*, 1994).

The sequence of MSP-3 peptide consists of different amount of amino acids, depends on the parasite clone, e.g. 353 residues in 3D7, 380 residues in D10. Most of MSP-3 sequence is conserved, especially in the N and C-terminus. The sequences from 2 distinct locations (Papua New Guinea and Tanzania) show limit MSP-3 diversity (Huber *et al.*, 1997). The sequence analysis reveals dimorphic patterns, 3D7 and K1 types (McColl and Ander, 1997). The peptide has 3 blocks of tandemly-repeated heptads, AXXAXXX which are separated by short non-repeated sequence (Figure 6.1). Most sequence substitutions, insertions and deletions occurred

in the heptads repeat domains, only limit sequence variations occurr in conserved flanking regions (McColl and Ander, 1997 and Huber *et al.*, 1997).

Figure 6.1 This picture shows putative structure of MSP-3. The black boxes represent the conserved regions of N-terminus and C-terminus of the polypeptide. The stripped box is the variation region which consists of 4 alanine heptads repeats. The blank boxes represent conserved loci which are not served as a signaling sequence and a GPI-anchored tail.

6.2 Results

6.2.1 The gene amplification and sequencing

The *msp-3* gene was amplified by using both PCR and nested PCR technique. The primers, which used in this gene amplification, were described in previous chapter (Chapter III). The amplified products from each isolates, using msp3F and msp3R primers, gave a single band, with size variation in different isolates (Figure 6.2). These single band isolates were amplified in 3-5 reaction tubes, then, they were pooled into a 1.5 ml of microfuge tube. The sequencing processes were described in chapter III. This PCR product failed to give any readable sequencing result. The electropherogram of 30 isolates, which were sequenced with those 2 primers, showed high noise signals (Figure 6.3). As the result, the consensus *msp-3* sequences of each isolates cannot be completed.

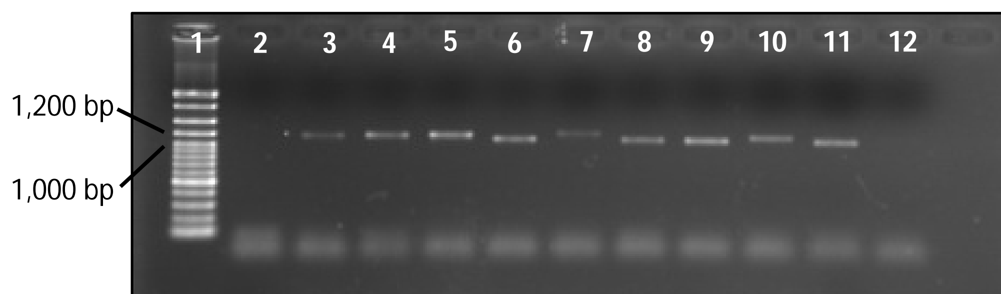

Figure 6.2 This picture shows a single band PCR products of *msp-3* gene from 10 different isolates, using msp3F and msp3R primers as PCR primers. At lane 1 is the 100 bp ladder marker (Bioexcellent). The negative control of amplification reaction is shown in lane 2. The parasite isolates, including TD510, TD515, TD529, TD504, TD530, TD508, UB27, UB14, UB7 and UB85, are shown in lane 3 – 12, respectively. The thermal cycles of this amplification include initiation denaturation step at 95°C for 5 min following with 30 cycles of 95°C for 30 sec, 52°C for 30 sec and 72°C for 1.20 min. The *msp-3* gene products are between 1,100-1,200 bp markers of the 100 bp ladder (Bioexcellent).

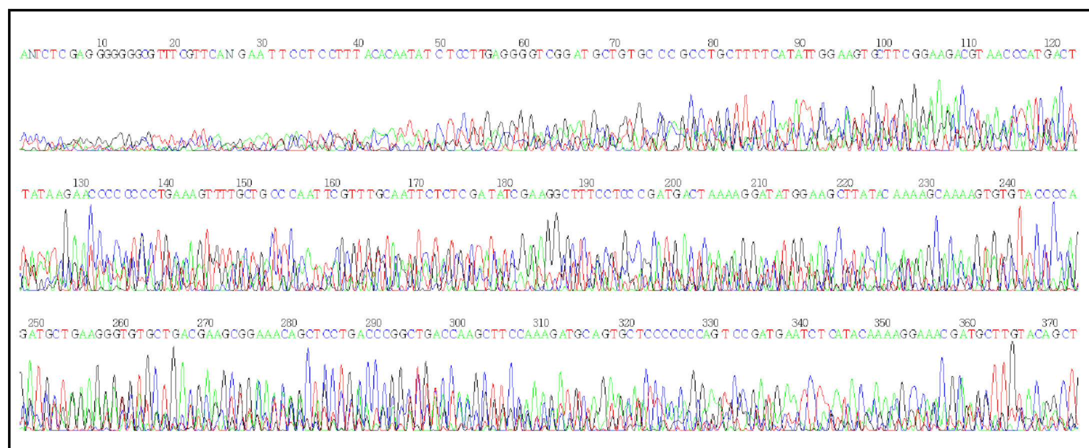

Figure 6.3 The picture shows the signal interruption in electropherogram of forward strand of *msp*-3 gene from TD529 isolate. In our study, both of forward and reverse sequences will be aligned. The sequence of each isolate is deduced from the complementary sequence of both strands.

The new primers were designed to solve the problem and the nested-PCR technique was used to enhance reaction specificity. The outer primers, namely new_outM3F and new_outM3R, which located at upstream region and downstream region of older primers were used in the first round of nested-PCR. The PCR product from this reaction was not detected by UV transillumination. In this nested reaction, the DNA was amplified by using the msp3F and msp3R primers. The nested PCR products of second round were single bands (Figure 6.4). Unfortunately, the electropherograms of nested-PCR products still showed signal interruption. Finally, the *msp*-3 gene sequence cannot be completed.

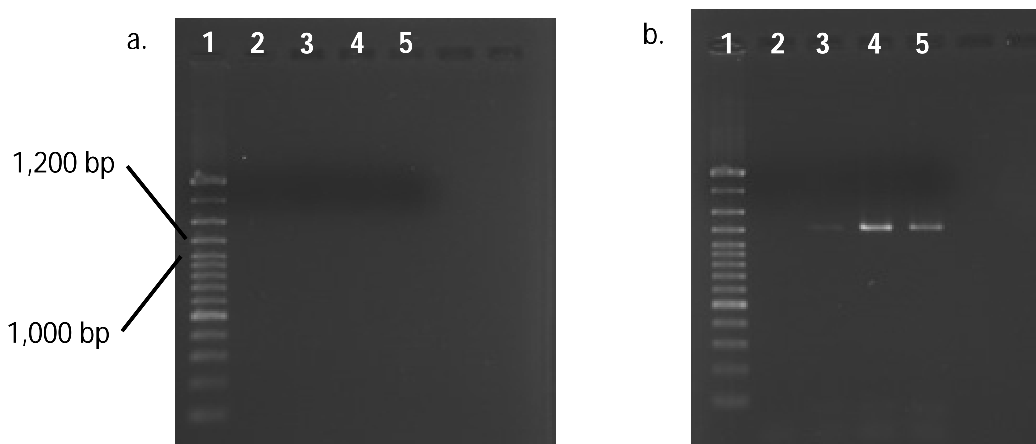

Figure 6.4 Both gel pictures show the first round (a.) and second round (b.) of nested-PCR products which are given by gene amplification. At lane 1 and lane 2 of both pictures show the 100 bp ladder marker and negative control results respectively. The first round of nested-PCR product of MH51, MH66 and MH50 parasite isolates cannot be detected by the agarose gel electrophoresis, in the lane 3-5 of picture a. The PCR products of the second round of the same parasite isolates are in the lane 3-5 of picture b.

6.3 Discussion

The DNA amplification reactions, both PCR and nested PCR, were able to amplify a DNA product from each isolates, about 1,200 bp in size, as expected. The agarose gel electrophoresis, from PCR reaction, can detect size variation among parasite isolates. This result suggested that these *msp3* genes also contained insertion-deletion variation, similar to previous reports. Because these parasite samples were collected from different areas of Thailand, it could be concluded that most isolates of *P. falciparum* parasites in Thailand were cloned and most of them are highly polymorphism.

The error of sequencing results may cause by several problems. Although the annealing temperature had been raised to increase specific binding between primer and DNA template until the electrophoresis results showed a single size of PCR products, the results still produced the signal interruptions of electropheroogram. It was possible that the binding between primers and DNA templates were not specific enough, there were other PCR products with the same size as required PCR product or other minor PCR products which cannot be detected by agarose gel electrophoresis. These non-specific PCR products may cause those noises as seen in electropherograms. New primers may require to solve this problem. However, the AT rich nature of plasmodium DNA, especially in the flanking regions of *msp-3* gene, is a major obstacle to primer design. More complicated molecular techniques should be explored to improve the sequencing result, such as genomic digestion with restriction fragment prior to PCR or DNA cloning.

CHAPTER VII

THE MEROZOITE SURFACE PROTEIN-4 (MSP-4)

7.1 Introduction

The merozoite surface protein-4 is an integral membrane protein on the merozoite surface (Marshall *et al.*, 1997). It was characterized after the MSP-1 (Holder and Freeman, 1984) and MSP-2 (Smythe *et al.*, 1988). The MSP-4 is a 40 kDa polypeptide. The expression of MSP-4 is varied on parasite life cycle. The highest rate of MSP-4 expression occurs during trophozoite and shizont stages parasite, detected by immunoblots with human antisera. Immunofluorescent assays showed that the protein is located on the merozoite surface (Marshall *et al.*, 1997). The protein integrates to the merozoite surface with glycosylphosphatidylinositol anchored domain at the carboxyl terminus. The polypeptide sequence of MSP-4 contains the hydrophobic signal domain at the N-terminus region and a single epidermal growth factor (EGF) - like domain at C-terminus region. Previous studies of other merozoite surface proteins, such as MSP-1, revealed that the EGF-like domain is immunogenic (Chappel and Holder, 1993). Although, the disruption of EGF – like domain of MSP-4 affects to its antigenicity property and disrupts the disulfide bonds in other part of polypeptide molecule, the mechanisms of those phenomena are still unclear (Wang et al., 1999).

Approximately 10 kbp region around the *msp*-4 gene on the *Plasmodium falciparum* chromosome 2 have been characterized. The *msp*-4 sequence contained 144 bp of intron which was AT-rich region. Moreover, the upstream untranslated region of msp-4 gene is also AT-rich (Marshall *et al.*, 1998). The putative structure of the *msp*-4 gene is shown in the following picture (Figure 7.1).

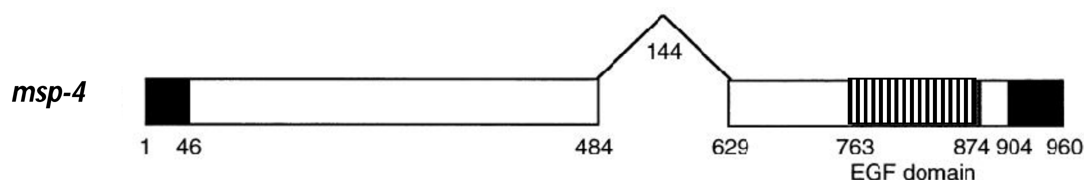

Figure 7.1 The picture shows putative structure of the *msp*-4 gene sequence. The black boxes represent fragments which encode the hydrophobic and GPI-anchored tail residues. The stripped box indicates EGF-like domain inside the *msp*-4 gene (Marshall *et al.*, 1998)

Most published data of merozoite surface protein-4 (*msp*-4) gene sequences are restricted to the immunogenic regions. The diversity status of the whole gene had never been truly determined.

7.2 Results

7.2.1 The gene amplification and sequencing

The primers, including msp4_primerF, msp4_primerR, Int_msp4F and Int_msp4R, were designed. The nested PCR technique was used for gene amplification process. The first round PCR using the msp4_primerF and msp4_primerR primers should give a 1,795 bp size amplicon but the product can not be detected by agarose gel electrophoresis. The second round PCR with the inner primers, Int_msp4F and Int_msp4R, should give smaller PCR product size (1,405 bp). The nested-PCR product detection using gel electrophoresis and, then, UV transillumination reveals a single band product. The gene was amplified in 4-5 reaction tubes per sample and pooled into a 1.5 ml microfuge tube. The products were purified and sequenced as described in Chapter III. The sequencing results from all 17 isolates showed the unreadable

electropherogram. Both forward and reverse strand sequences show signal interruption (Figure 7.2). No consensus sequence can be deduced from these sequencing.

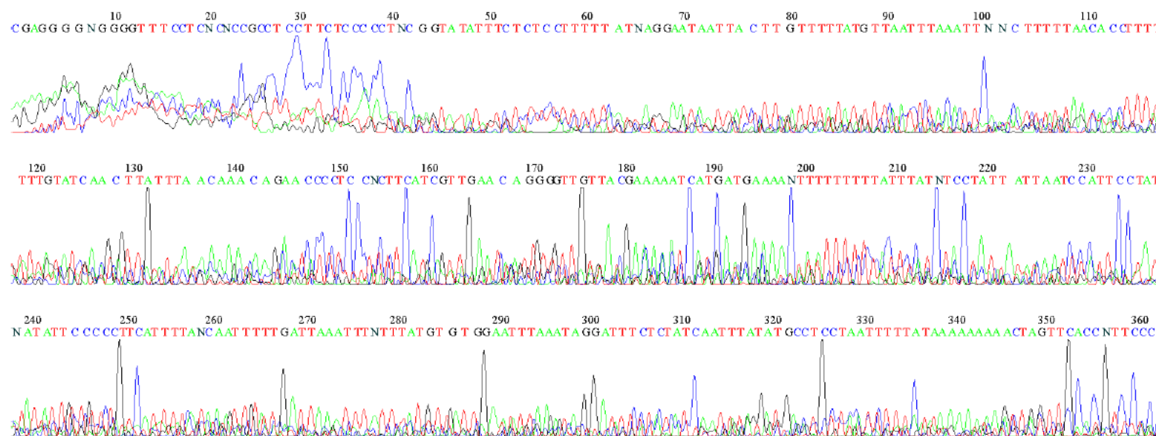

Figure 7.2 This picture reveals signal interruption of the UB85 isolate forward strand sequence. The reverse strand sequence (picture not shown) also displays signal interruption.

New set of primers, including newmsp4F, newmsp4R, new_intM4F and new_intM4R, were designed for nested-PCR technique. The first round PCR using newmsp4F and newmsp4R primers produced a PCR product as predicted, 1,276 bp in size. Consequently, the product was serial diluted to 1:2, 1:4 and 1:8 dilutions. The second round PCR using new_intM4F and new_intM4R primers gave single band product from the 1:4 and 1:8 dilution samples (Figure 7.3).

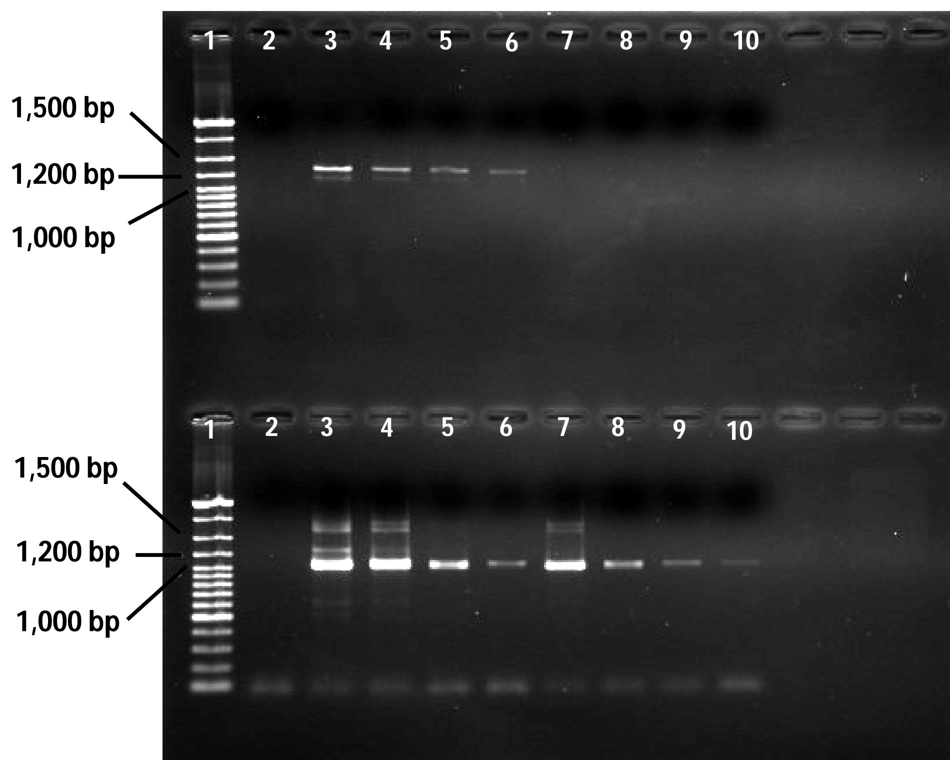

Figure 7.3 This picture shows the first round (upper row) and second round (lower row) product of the *msp*-4 gene. Lane 1 and 2 of the upper and lower rows is 100 bp ladder markers (Bioexcellent). and negative control, respectively. In the uper row, lane 3 is the first round PCR product of RN19 isolate and the following lane 4-6 are the diluted product of RN19 at 1:2, 1:4 and 1:8 dilution. Lane 7 is the first round PCR product of RN26 isolate and lanes 8-10 are the diluted product of RN19 at 1:2, 1:4 and 1:8 dilution. In the lower row, the second round PCR products of RN19 and RN26 show in lane3-6 and lane 7-10 (using non diluted, 1:2, 1:4 and 1:8 dilution DNA template) respectively. The thermal cycles of these PCR consist of initiation denaturation step at 95°C for 5 min and the following 25 cycles of 95°C for 40 sec, 60°C for 40 sec and 72°C for 2 min (first round PCR) and 30 cycles of 95°C for 40 sec, 60°C for 40 sec and 72°C for 1.30 min (second round PCR).

7.3 Discussions

The nested PCR product with the first set primers, including msp4_primerF, msp4_primerR, Int_msp4F and Int_msp4R, showed the single band product but the sequencing with these first set primers fail to give the readable electropherogram. The problem may cause by either primer specificity or *msp*-4 gene structure. The msp-4 gene locates on the chromosome 2 of *P. falciparum* and it is located next to the *msp*-5 gene. The intragenic region, which distinguished the *msp*-4 and *msp*-5 genes, displays AT-rich nature. The upstream region of the *msp*-4 gene has also high AT content. The first set primers were designed from those high AT sequences and may cause non specific priming.

New pairs of primers are designed. The results from agarose gel electrophoresis (Figure 7.3) show multiple bands in the first round PCR product of RN19, while no PCR product in the first round PCR can be detected from the RN26 sample. Surprisingly, both samples give multiple band products in nested-PCR reaction. Using diluted sample from the first round-PCR can greatly reduce non-specific product. This suggested that the carry-over from the first PCR reaction, at high concentration, may cause multiple band products and, so, the dilution helps reduce the problem. Although no sequence can be concluded from the experiment, no size variation can be revealed by agarose gel electrophoresis.

CHAPTER VIII

THE MEROZOITE SURFACE PROTEIN-5(MSP-5)

8.1 Introduction

The MSP-5 is an integral membrane protein which attaches on the merozoite surface with glycosylphosphatidylinsitol (GPI) - anchored tail (Marshall *et al.*, 1998). The MSPs antigens which have GPI-anchored domain, included MSP-1, MSP-2, and MSP-4. The MSP-5 is a 40 kDa protein which has antigenic property. The epidermal growth facter (EGF) - like domain, locates on the C-terminus, of this MSP-5 is recognized by infected human immunity. Other MSP proteins, which contain the EGF-like residue, included $MSP1_{19}$ fragment of the MSP-1 (Holder and Freeman, 1984) and MSP-4 (Marshall *et al.*, 1997). The function of EGF-like domain has not yet been illucidated but the blockage of these residues by antibodies can affect the parasite growth *in vitro* (Blackman and Holder, 1992, Chappel *et al.*, 1993 and Wang *et al.*, 1999). On the other hand, the MSP-2 and MSP-3, lacking EGF-like domain in their polypeptides, can also be recognized by antibodies from immuned or infected human. The repetitive region on the MSP-2 is the target of immunity (Smythe *et al.*, 1990).

The merozoite surface protein-5 (*msp-5*) gene, encoded for MSP-5, is located on the chromosome 2 of the *P. falciparum*. Other *msp* genes including *msp*-2 and *msp*-4 genes, encoded for MSP-2 and MSP-4 respectively, is also located on the same chromosome. These genes arrange on the chromosome as *msp*-4, *msp*-5 and *msp*-2 genes, respectively, in head to tail direction. The intragenic distance between *msp*-4 and *msp*-5 gene is 1021 bp (Marshall *et al.*, 1998). Both flanking parts of *msp*-5 gene, 5′ and 3′ untranslate regions (UTR), and 134 bp of an

intron contains AT-rich sequences. . The 3′ sequence of *msp*-5 gene encodes for the cysteine rich EGF-like domain and GPI-anchored protein residue (Figure 8.1). Sequence comparison between MSP-4 and MSP-5 shows the similarity of both peptide structures in which a hydrophobic domain, at N-terminus, 515 bp, is separted from the EGF-like domain and GPI-anchored tail residues, 303 bp, by an intron. This similarity suggested that both genes may originate from duplicated genes which are diverged from each other (Marshall *et al.*, 1998).

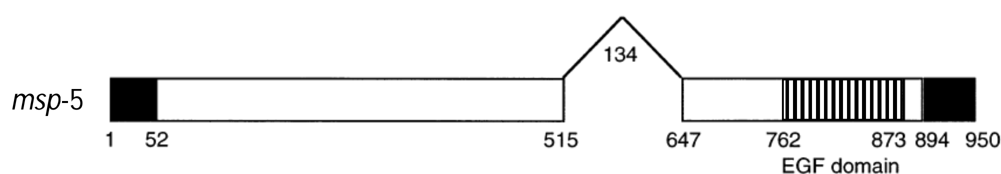

Figure 8.1 The general features of the *msp*-5 gene. Both black boxes at 5′ and 3′ terminus represent a hydrophobic and GPI-anchored tail encoded sequences, respectively. The stripped box indicates EGF-like domain (Marshall et al., 1998).

Because its importance for vaccine development, the EGF-like domain encoded sequence of *msp*-5 gene is well studied in *P. falciparum* samples from few parasite clones, but the data of entire *msp*-5 sequences which were collected in field isolate has not yet been reported. Our study will be the first report which reveals the diversity of the msp-5 gene of the *P. falciparum* populations of Thailand.

8.2 Results

8.2.1 The gene amplification

In this study, the primers for *msp*-5 gene amplification were designed as described in chapter III. The agarose gel electrophoresis and UV transilluminator were used to detect the PCR product. The first round PCR, using Ext_msp5F and Ext_msp5R primers, gave a single band at predicted size, at 1,856 bp. The second round amplification, using msp5_primerF and msp5_primerR primers, gave a single product at predicted size, 1,308 bp. Limit size variations among these second round PCR products were observed. (Figure 8.2).

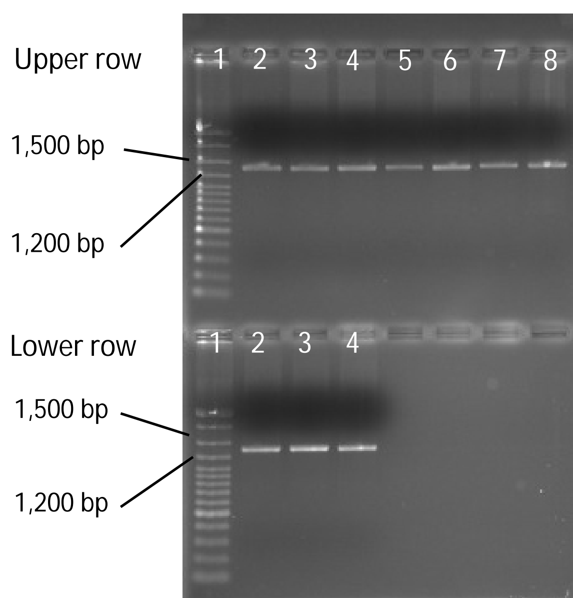

Figure 8.2 The picture shows the second round PCR product of 9 Kanchanaburi isolates and 1 Tak isolate (served as positive control). The PCR products located between 1,200 and 1,500 bp of 100 bp ladder marker has been detected. Lane 1 of both upper and lower rows are 100 bp ladder markers (Bioexcellent). Lane 2-8 in upper row show the PCR

products of T9/94, K195, K74, K215, K66, K391 and K58, respectively. Lower row shows the PCR products of K165 (lane 2), K205 (lane 3) and K397 (lane 3).

The 4-5 reaction tubes of each isolate were pooled; then, each sample was purified and sequenced (Macrogen Inc., Korea). The sequencing reactions failed to give any readable sequence. The electropherograms of those sequences revealed interfered peaks. One of signal interference was shown below (Figure 8.3).

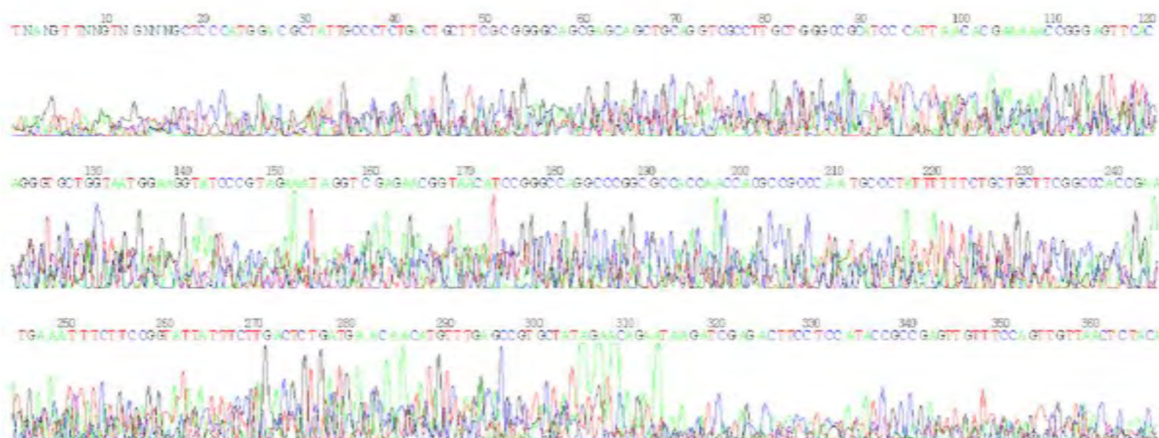

Figure 8.3 The picture shows an electropherogram with signal interference from sequencing reaction of K195 isolate.

As the result, the 3 new primers were designed and the semi-nested PCR technique was used to amplify the gene. First round PCR using 2 new primers, including new_outM5F and new_msp5R, gave a major band product after agarose gel electrophoresis detection, as predicted size (1,036 bp) located between 1,000 and 1,100 bp marker of 100 bp ladder marker

(Bioexcellent) (Figure 8.4). The gene was amplified again with second round PCR, using new_msp5F and new_msp5R primers. After UV detection, the result reveals multiple bands amplicon (Figure 8.3). The annealing temperature was increased to enhance reaction specificity. Although the final annealing temperature of both first round and second round PCR were increased to 61°C, multiple PCR products were still detected.

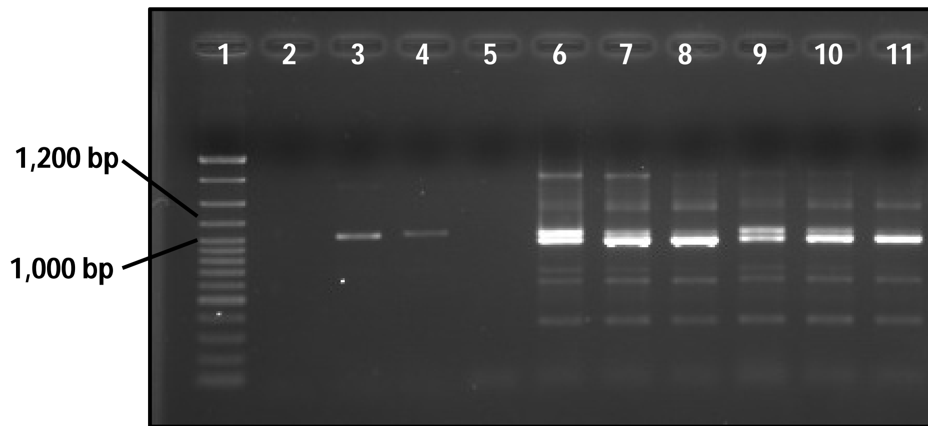

Figure 8.4 The picture shows the PCR product from RN19 and RN26 isolates. The lane 1 is 100 bp ladder markers (Bioexcellent). Negative control of first round and second round PCR are shown in land 2 and lane 5 respectively. The first round products of RN19 are in lane 3 and the RN26 product is in lane 4. The second round PCR products of RN19 and RN26 are shown in lane 6-8 (non diluted, 1:2, 1:5 dilution) and lane 9-11 (non diluted, 1:2, 1:5 dilution) respectively.

8.3 Discussions

There were limit size variation which revealed by agarose gel electrophoresis. Because the binding site of the second round PCR (Figure 8.2) and first round PCR (Figure 8.4) primers were located at the untranslated regions (UTR) of the gene. As the results, the insertion and deletion which caused size different may occur inside or outside of the *msp*-5 gene. However, if the insertion or deletion is inside the gene region these will be the first report of size variation in the *msp*-5 gene.

The noise in electrophilogram may cause by several problems. Although, the single banded products were sequenced, no readable sequence was obtained. The AT-rich nature of the *msp*-5 gene, including the intron and 5′ and 3′ untranslated regions (UTR) of the gene may be an obstacle in sequencing reaction. The new primers were designed to improve the signal interruption problem.

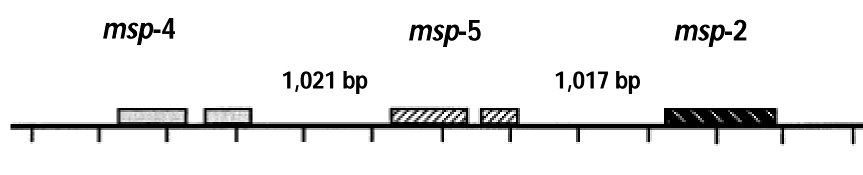

Figure 8.5 The picture shows the putative structure of the partial chromosome 2 of *P. falciparum* which contains *msp*-2, *msp*-4 and *msp*-5 genes. The 2 intragenic regions between 3 genes loci are shown in this picture. The 5′ and 3′ untranslated regions of msp-4 and msp-5 have high AT content. (Marshall *et al.*, 1998)

The problem of PCR with new 3 primers, including new_outM5F, new_msp5F and new_msp5R, is multiple sizes of PCR product. The increasing of annealing temperature to gain

more specificity did not improve the problem. It seems that these multiple bands may not result from non-specific binding of the primers. The larger band products, which show in figure 8.4, may cause by the remaining of the first round PCR primers and original DNA template. Using dilute first round PCR product seems to reduce the large nonspecific PCR product (Figure 8.4). Furthermore, the *msp-5* gene contains an AT-rich intron at its center. From our calculation, 144 bp of intron consists of approximately 90% of nucleotide A and T. The AT-rich region of intron may cause to the imbalance of dNTPs in the reaction and lead to the slippage of the DNA synthesis. As a result, some PCR products were shorter than expected size. More complicated methods may be used to obtain the single size of PCR product. For example, the purification of first round PCR product should be done to remove the unwanted primer of the first round PCR before using in the following step. The dATPs and dTTPs concentrations should be adjusted for both PCR and nested PCR reaction. The higher dATPs and dTTPs concentration may reduce the problem due to imbalance usage of dNTP during gene amplification.

CHAPTER IV

CONCLUSION

The merozoite surface proteins, including MSP-1_{19} and MSP-2, are considered as an important vaccine candidate protein. According to the MSP-1_{19} antigenic region, it contains epidermal growth factor (EGF)-like sequence which serves as a protein epitope for immune response (Holder and Freeman, 1984), while, the repetitive sequence of MSP-2 displays the target of human immunity (Smythe *et al.*, 1990). The sequence variation of EGF-like domain of MSP-1_{19} in Thailand has been restricted for the last 10 years. There are only 5 haplotypes of this antigenic residue which have been found in Thailand during 1999-2009 (Sakihama *et al.*, 1999 and Sakihama *et al.*, 2006). The frequency of 5 haplotypes in our 5 studied areas may be changed but the major haplotype and other alleles still remain in the population. On the other hand, the MSP-1_{19} haplotypes and their frequencies are different in the other parts of the world. Because the malaria vaccine against *Plasmodium falciparum* using MSP-1_{19} epitope should be designed to cover all haplotypes in an intervention area, the vaccine should be produced for a specific area based on its haplotypes data.

The repetitive sequence of MSP-2, especially in the 3D7 allele, revealed high variation. In Thailand, 50 parasite isolates displayed 35 allele variants. Similarly, other studies of worldwide MSP-2 repeated sequence also indicated a large number of MSP-2 variants (Ferreira et *al.*, 2007). The different of these sequences were produced by insertion and deletion in the repetitive region. Moreover, the hybrid between 2 alleles of MSP-2 repetitive region had been found in Africa, South America and Myanmar (Ferreira *et al.*, 2007 and Kang *et al.*, 2010). The highly polymorphic status of *msp*-2 gene suggested that it seemed to be difficult to design the vaccine using MSP-2 repetitive sequence as vaccine development epitope. Although, the

previous experiment of human immune response to this MSP-2 epitope had been promising, a few *in vitro* recombinant MSP-2 proteins were studied. More information of MSP-2 diversity will be needed to illuminate the problem of sequence variation in vaccine development.

The attempt to study other MSPs in this study, MSP-3, MSP-4 and MSP-5, were not succeeded. Because of low specificity of our designed primers, the sequencing has given signal interruption electropherogram which may cause by non-specific PCR product. This problem may be improved by increaseing the specificity of primers or more complicated molecular techniques.

Finally, the results from this work were not only show the polymorphism status of both merozoite surface protein-1 and -2 in Thailand, the data also show that it may be possible to identify *P. falciparum* clone by using the sequence data. At present, there is no available technique to identify a clone of malaria parasite. The parasites were named according to their collected sites and their original culture. It would be difficult to prove that a clone is contaminated. Because *msp* genes contained highly variable regions, each clone of the parasite may have the unique sequence of these genes which may be used for identification purpose.

REFERENCES

Blackman, M.J., Heidrich, H.G., Donachie, S., McBride, J.S. and Holder, A.A. 1990. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. Journal of Experimental Medicine 172: 397-382.

Blackman, M.J., Ling, I.T., Nicholls, S.C. and Holder, A.A. 1991. Preteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. Molecular and Biochemical Parasitology 49: 29-34.

Blackman, M.J. and Holder, A.A. 1992. Secondary processing of the *Plasmodium falciparum* merozoite surface protein-1 (MSP1) by a calcium dependent membrane-bound serine protease: shedding of MSP133 as a noncovalently associated complex with the other fragments of the MSP1. Molecular and Biochemical Parasitology 50:307-315.

Chang, S.P., Gibson, H.L., Lee-Ng, C.T., Barr, P.J. and Hui, A. 1992. Carboxyl-terminal fragment of *Plasmodium falciparum* gp195 expressed by a recombinant baculovirus induces antibodies that completely inhibit parasite growth. Journal of Immunology 149: 548-555.

Chappel, J.A. and Holder, A.A. 1993. Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion in vitro recognize the first growth factor-like domain of merozoite surface protein-1. Molecular and Biochemical Parasitology 60: 303-312.

Cowman, A.F., Baldi, D.L., Healer, J., Mills, K.E., O'Donnell, R.A., Reed, M.B., Triglia, T., Wickham, M.E. and Crabb, B.S. 2000. Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. Federation of European Biochemical Societies 476: 84-88.

Cowman, A.F., Baldi, D.L., Duraisingh, M., Healer, J., Mills, K.E., O'Donnell, R.A., Thompson, J., Triglia, T., Wickham, M.E. and Crabb, B.S. 2002. Functional analysis of *Plasmodium falciparum* merozoite antigens: implications for erythrocyte invasion and vaccine development. Philosophical Transactions of the Royal Society 357: 25-33.

Ferreira, M.U., Kaneko, O., Kimura, M., Liu, Q., Kawamoto, F. and Tanabe, K. 1998. Allelic diversity at the merozoite surface protein-1 (MSP-1) locus in natural *Plasmodium falciparum* populations: a brief overview. Memorias do Institudo Oswaldo Cruz, Rio de janeiro 93: 631-638.

Ferreira, M.U., Ribeiro, W.L., Tonon, A.P., Kawamoto, F. and Rich, S.M. 2003. Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*. Gene 304: 65-75.

Ferreira, M.U., and Danial, L.H. 2007. *Plasmodium falciparum*: Worldwide sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-2 (MSP-2). Experimental Parasitology 115: 32-40.

Flück, C., Smith, T., Beck, H.P., Irion, A., Beluela, I., Alpers, M.P., Anders, R., Saul, A., Genton, B. and Felger, I. 2004. Strain-specific humoral response to a polymorphic malaria vaccine. Infection and Immunity 72: 6300-6305.

Happi, C.T., Gbotosho, G.O., Sowunmi, A., Falade, C.O., Akinboye,D.O., Gerena, L., Kyle, D.E., Milhous, W., Wirth, D.F. and Oduola, M.J. 2004. Molecular analysis of *Plasmodium falciparum* recrudescent malaria infections in children treated with chloroquine in Nigeria. American Journal of Tropical Medicine and Hygiene 70: 20-26.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R.F., Brooker, S., Smith, D.L., Moyeed, R.A. and Snow, R.W. 2009. A World Malaria Map: *Plasmodium falciparum* Endemicity in 2007. The Public Library of Science Medicine 6: 286-302.

Holder, A.A. and Freeman, R.R. 1984. The three major antigens on the surface of *Plasmodium falciparum* merozoites are derived from a single high molecular weight precursor. The Journal of Experimental Medicine 160: 624-629.

Huber, W., Felger, I., Matile, H., Lipps, H.J., Steiger, S. and Beck, H.P. 1997. Limited sequence polymorphism in the *Plasmodium falciparum* merozoite surface protein 3. Molecular and Biochemical Parasitilogy 87: 231-234.

Jungwutiwes, S., Putapornthip, C. and Huges, A.L. 2010. Bottleneck effects on vaccine-candidate antigen diversity of malaria parasites in Thailand. Vaccine 28: 3112-3117.

Kaneko, O., Kimura, M., Kawamoto, F., Ferreira, M.U. and Tanabe, K. 1997. *Plasmodium falciparum*: allelic variation in the merozoite surface protein 1 gene in wild isolates from southern Vietnam. Experimental Parasitology 86: 45-67.

Kang, J.M., Moon, S.U., Kim, J.Y., Sho, S.H., Lin, K., Sohn, W.M., Kim, T.S. and Na, B.K. 2010. Genetic polymorphism of merozoite surface protein-1 and merozoite surface protein-2 in *Plasmodium falciparum* field isolates from Myanmar. Malaria Journal 9: 131-138.

Lalitha, P.V., Malrihta, P., Chattopadhyay, R. and Chauhan, V.S. 1999. *Plasmodium falciparum*: variations in the C-terminal cysteine-rich region of the merozoite surface protein-1 in field samples among Indian isolates. Experimental Parasitology 92: 12-18.

Lawrence, G., Cheng, Q.Q., Reed, C., Taylor, D., Stowers, A., Cloonan, N., Rzepczyk, C., Smillie, A., Anderson, K., Pombo, D., Allworth, A., Eisen, D., Anders, R. and Saul, A. 2000. Effect of vaccination with 3 recombinant asexual-stage malaria antigens on initial growth rates of *Plasmodium falciparum* in non-immune volunteers. Vaccine 18: 1925-1931.

Marshall, V.M., Silva, A., Foley, M., Cranmer, S., Wang, L., McColl, D.J., Kemp, D.J. and Coppel, R.S. 1997. A second Merozoite Surface Protein (MSP-4) of *Plasmodium falciparum* That Contains an Epidermal Growth Factor-Like Domain. Infection and Immunity 65: 4460-4467.

Marshall, V.M., Tieqiao, W. and Coppel, R.L. 1998. Close linkage of three merozoite surface protein genes on chromosome 2 of *Plasmodium falciparum*. Molecular and Biochemical Parasitology 94: 13-25.

McColl, D.J., Silva, A., Foley, M., Kun, J.F.J., Favaloro, J.M., Thompson, J.K., Marshall, V.M., Coppel, R.L., Kemp, D.J. and Anders, R.F. 1994. Molecular variation in a novel polymorphic antigen associated with *Plasmodium faliparum* merozoites. Molecular and Biochemical Parasitilogy 68: 53-67.

McColl, D.J. and Anders, R.F. 1997. Conservation of structural motifs and antigenic diversity in the *Plasmodium falciparum* merozoite surface protein-3 (MSP-3). Molecular and Biochemical Parasitilogy 90: 21-31.

Mehrizi, A.A., Zakeri, S., Salmanian, A.H., Sanati, M.H. and Djadid, N.D. 2008. *Plasmodium falciparum*: Sequence analysis of the gene encoding the C-terminus region of the merozoite surface protein-1, a potential malaria vaccine antigen, in Iranian clinical isolates. Experimental Parasitology 118(3): 378-385.

Mullis, K.B. 1990. The unsual origin of the polymerase chain reaction. Scientific American 262: 56-61, 64-65.

Oeuvray, C., Bouharoun-Tayoun, H., Gras-Masse, H., Bottius, E., Kaidoh, T., Aikawa, M., Filgueira, M.C., Tartar, A. and Druilhe, P. 1994. Merozoite surface protein-3: a malaria inducing antibodies that promote *Plasmodium falciparum* killing by cooperation with blood monocyte. Blood 84: 1594-1602.

Polley, S.D., Conway, D.J., Cavanagh, D.R., McBride, J.S., Lowe, B.S., Williums, T.N., Mwangi, T.W. and Marsh, K. 2006. High levels of serum antibodies to merozoite surface protein 2 of *Plasmodium falciparum* are associated with reduced risk of clinical malaria in coastal Kenya. Vaccine 24: 4233-4246.

Public Health, Ministry. Department of Disease Control.

Putaporntip, C., Jongwutiwes, S. and Hughes, A.L. 2008. Differential selective pressures on the merozoite surface protein 2 locus of *Plasmodium falciparum* in a low endemic area. Gene 427: 51-57.

Qari, S.H., Shi, Y.P., Goldman, I.F., Nahlen, B.L., Tibeyrnc, M. and Lal, A.A. 1998. Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen. Molecular and Biochemical Parasitology 92: 241-252.

Sachs, J. and Malaney, P. 2002. The economic and social burden of malaria. Nature 415: 680-685.

Sakihama, N., Kimura, M., Hirayama. K., Kanda, T., Na-Bangchang, K., Jongwutiwes, S., Conway, D. and Tanabe, K. 1999. Allelic recombination and linkage disequilibrium within Msp-1 of *Plasmodium falciparum*, The malignant malaria parasite. Gene 230: 47-54.

Sakihama, N., Ohmae, H., Bakotee, B., Kawabata, M., Yama, K. and Tanabe, K. 2006. Limited allelic diversity of *Plasmodium falciparum* merozoite surface protein 1 gene from populations in the Solomon Islands. American Journal of Tropical Medicine and Hygiene 74(1): 31-40.

Sakihama, N., Nakamura, M., Palanca Jr, A.A., Argubano, R.A., Realon, E.P., Larracas, A.L., Espina, R.L. and Tanabe, K. 2007. Allelic diversity in the merozoite surface protein 1 gene of *Plasmodium falciparum* on Palawan Island, the Philippines. Parasitology International 56(3): 185-194.

Seugorn, A., Siripoon, N., Kanchanakarn, N., Rungsihirunrat, K., Pumpaipool, T., Vichaikatka, S., Thaithong, S. and Harnyuttanakorn, P. 2007. Drug susceptibility of *Plasmodium falciparum* collected from different areas of Thailand during 2000-2001. Journal of Health Research 21(2): 119-124.

Smythe, J.A., Coppel, R.L., Brown, G.V., Ramasamy, R., Kemp, D.J. and Anders, R.F. 1988. Identification of two integral membrane proteins of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences USA 85: 5195-5199.

Smythe, J.A., Peterson, M.G., Coppel, R.L., Saul, A.J., Kemp, D.J. and Anders, R.F. 1990. Structural diversity in the 45-kilodalton merozoite surface antigen of *Plasmodium falciparum*. Molecular and Biochemical Parasitology 39: 227-234.

Snounou, G., Viriyakosol, S., Jarra, W., Thaithong, S. and Brown, K.N. 1993. Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of a high prevalence of mixed infections. Molecular and Biochemical Parasitology 58: 283-292.

Snounou, G., Xinping, Z., Siripoon, N., Jarra, W., Thaithong, S., Brown, K.N. and Viriyakosol, S. 1999. Biased distribution of msp1 and msp2 allelic variants in *Plasmodium falciparum* populations in Thailand. Transaction of the Royal Society of Tropical Medicine and Hygiene 93: 369-374.

Takala, S., Branch, O., Escalante, A.A., Kariuki, S., Wootton, J. and Lal, A.A. 2002. Evidence for intragenic recombination in *Plasmodium falciparum*: identification of a novel allele family in block 2 of merozoite surface protein-1: Asembo bay Area Cohort Project XIV. Molecular and Biochemical Parasitology 125: 163-171.

Tanabe, K., Mackay, M., Goman, M.and Scaife, J.G. 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. Journal of Molecular Biology 195: 273-287.

Treger, W. and Jensen, J.B. 1976. Human malaria parasites in continuous culture. Science 193(4254): 673-675.

Vijay Kumar, S., Ranjan, S., Saxena, V., Rajesh, V., Roy, S.K., Kochar, D., Ranjan, A. and Das, A. 2005. *Plasmodium falciparum*: Genetic diversity of C-terminus region of MSP-1 in isolates from Indian sub-continant. Experimental Parasitology 110(4): 384-388.

Wang, L., Black, C.G., Marshall, V.M., and Coppel, R.L. 1999. Structural and Antigenic Properties of Merozoite Surface Protein 4 of *Plasmodium falciparum*. Infection and Immunity 67: 2193-2200.

APPENDICES

A.1 The 61 sequences of MSP-1_{19} are shown below. The non-synonymous mutation positions are position 1 (E/Q), position 48 (K/T), position 57 (N/G), position 58 (G/R) and position 73 (L/F).

```
        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
            5          15         25         35         45         55         65         75
UB85    QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB27    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB14    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB7     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSRKK ITCECTKPDS YPL-------
UB22    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB28    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB84    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB82    QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB52    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB50    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB59    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
UB51    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD542   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSRKK ITCECTKPDS YPL-------
TD530   QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD529   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD504   QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD515   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD510   QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD508   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD554   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSRKK ITCECTKPDS YPL-------
TD531   QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD556   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
TD533   ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSGKK ITCECTKPDS YPL-------
RN130   QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN70    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN19    QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN26    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN28    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN31    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
```

```
         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
             5          15         25         35         45         55         65         75
RN133    QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPF-------
RN131    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSRKK ITCECTKPDS YPL-------
RN129    QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN122    ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN63     QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN66     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSGKK ITCECTKPDS YPL-------
RN68     QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
RN72     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K391     QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPF-------
K205     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K215     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K64      QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K66      ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K60      ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K58      ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K74      QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K165     QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPF-------
K185     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSGKK ITCECTKPDS YPL-------
K195     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSRKK ITCECTKPDS YPL-------
K389     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSGKK ITCECTKPDS YPL-------
K392     QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K386     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
K397     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH10     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADATCT EEDSGSSRKK ITCECTKPDS YPL-------
MH6      ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH7      ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH11     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH32     QNSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH18     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH24     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH28     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
MH20     ENSGCFRHLD EREECKCLLN YKQEGDKCVE NPNPTCNENN GGCDADAKCT EEDSGSNGKK ITCECTKPDS YPL-------
```

A.2 The 33 sequences of MSP-2 3D7 family are shown below. Blocks identifications of our study are block 1 (positions 1-28), block 2 (positions 29-39), block 3 (positions 40-218), block 4 (positions 219-309) and block 5 (positions 310-374).

```
             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 5         15         25         35         45         55         65
UB52         ---------- YSNTFINNAY NMSIRRSMEE SNPST----G AGGSGSAGGS GSAGGSGSAG GSGSAGGSGS
K205         ---------- YSNTFINNAY NMSIRRSMEE SNPST----G AGGSGSAGGS GSAGGSGSAG GSGSAGGSGS
RN70         ---------- YSNTFINNAY NMSIRRSMTE SKTPTPTPTG AGAGAGASGS AGSGDGASGS ASGSAGASGS
RN130        ---------- YSNTFINNAY NMSIRRSMTE SKTPTPT--G AGAGAGASGS AGSGDGASGS ASGSAGASGS
MH24         ---------- YSNTFINNAY NMSIRRSMTE SKTPTPT--G AGAGAGASGS AGSGDGASGS ASGSAGASGS
TD533        ---------- YSNTFINNAY NMSIRRSMTE SKTPTPT--G AGAGAGASGS AGSGDGASGS ASGSAGASGS
MH18         ---------- YSNTFINNAY NMSIRRSMAV SNPPT----G AGAVAGSGAG AGAVAGSGAG AGAVAGSGAG
K391         ---------- YSNTFINNAY NMSIRRSMAV SNPPT----G AGAVAGSGAG AGAVAGSGAG AGAVAGSGAG
RN31         ---------- YSNTFINNAY NMSIRRSMAV SNPPT----G AVAGSGAGAG AVAGSGAGAG AVAGSGAGAG
K64          VTFNIKNESK YSNTFINNAY NMSIRRSMAV SNPPT----G AVAGSGAGAG AVAGSGAGAG AVAGSGAGAG
K165         ---------- YSNTFINNAY NMSIRRSMAV SNPPT----G AVAGSGAGAG AVAGSGAGAG AVAGSGAGAG
TD504        ---------- YSNTFINNAY NMSIRRSMAV SNPPT----G AGAVAGSGAG AVAGSGAGAV AGSGAGAGAG
TD556        ---------- YSNTFINNAY NMSIRRSMAV SNPPT----G AGAVAGSGAG AVAGSGAGAV AGSGAGAGAG
UB59         ---------- YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGAGAS GSAGASGSAG ASGSAGASGS
UB7          ---------- YSNTFINNAY NMSIRRSMEE SNPPT----G AGAVAGSGAG AVAGSGAGAV AGSGAGAGAG
RN133        ---------- YSNTFINNAY NMSIRRSMAE SKTPT----G AGASGRAGSG DGAVASAGSG DGAVASAGSG
RN19         ---------- YSNTFINNAY NMSIRRSMAV SKPPT----G AGAGASGRAG SGDGAVASAG SGDGAVASA~
RN26         ---------- YSNTFINNAY NMSIRRSMEE SNPPT----G AGAVAGSGAG AVAGSGAVAG SGAGAGAVAG
RN131        ---------- YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGGSAG GSAGGSAGGS AGGSAGGSAG
K74          ---------- YSNTFINNAY NMSIRRSMAE SKTPT----G AGGSGSAGGS GSAGGSGSAG GSGSAGGSGS
K386         ---------- YSNTFINNAY NMSIRRSMEE SNPPT----G AGAVAGSGAG AVAGSGAVAG SGAGAGAVAG
K60          VTFNIKNESK YSNTFINNAY NMSIRRSMEE SNPPT----G AGAVAGSGAG AVAGSGAGAV AGSGAGAGAG
TD530        ---------- YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGAGAS GSAGAGASGS AGASGSAGAS
TD554        ---------- YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGGSAG GSAGGSAGGS AGGSAGGSAG
TD531        VTFNIKNESK YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGGSAG GSAGGSAGGS AGGSAGGSAG
RN122        ---------- ---------- ---IRRSMTE SNPPT----G ASGSAGGSAG GSAGGSAGGS AGGSAGGSAG
MH20         ---------- YSNTFINNAY NMSIRRSMKE SKPPT----G ASGSAGSGSG AVASA----- ----------
K185         ---------- YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGAGAS GSAGAGASGS AGASGSAGAG
K58          ---------- YSNTFINNAY NMSIRRSMTE SNPPT----G ASGSAGAGAS GSAGAGASGS AGASGSAGAG
K392         VTFNIKNESK YSNTFINNAY NMSIRRSMAE SKTPT----G ASGSAGSGAV ASAGSGAVAS AGSGAVASA-
TD515        ---------- YSNTFINNAY NMSIRRSMAE SKPPT----G TGASGSAGSG AGASGSAGSG DGAVASA---
TD529        ---------- YSNTFINNAY NMSIRRSMTE SKPPT----G TGASGSAGSG AGASGSAGSG DGAVASA---
TD508        ---------- YSNTFINNAY NMSIRRSMTE SKPPT----G TGASGSAGSG AGASGSAGSG DGAVASA---
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    75         85         95        105        115        125        135
UB52            AGGSGSAGGS GSAGGSGSAG SGD------- ---------- ---------- ---------- ----------
K205            AGGSGSAGGS GSAGGSGSAG SGD------- ---------- ---------- ---------- ----------
RN70            ASGSAGASGS ASGSAGASGS ASGSAGASGS ASGSAG---- ---------- ---------- ----------
RN130           ASGSAGASGS ASGSAGASGS ASGSAGASGS ASGSAG---- ---------- ---------- ----------
MH24            ASGSAGASGS ASGSAGASGS ASGSAGASGS ASGSAG---- ---------- ---------- ----------
TD533           ASGSAGASGS ASGSAGASGS ASGSAGASGS ASGSAG---- ---------- ---------- ----------
MH18            AGAGAGAVAG SGAGAGAGAV AGSGAGAGAV ASA------- ---------- ---------- ----------
K391            AGAGAGAVAG SGAGAGAGAV AGSGAGAGAV ASA------- ---------- ---------- ----------
RN31            AGAGAVAGSG AGAGAVAGSG AGAVASA--- ---------- ---------- ---------- ----------
K64             AGAGAVAGSG AGAGAVAGSG AGAVASA--- ---------- ---------- ---------- ----------
K165            AGAGAVAGSG AGAGAVAGSG AGAVASA--- ---------- ---------- ---------- ----------
TD504           AGAVAGSGAG AGAVAVAGSG AGAVASA--- ---------- ---------- ---------- ----------
TD556           AGAVAGSGAG AGAVAVAGSG AGAVASA--- ---------- ---------- ---------- ----------
UB59            AGASGSAGAS GSAGASGSAG ASGSAGSGA- ---------- ---------- ---------- ----------
UB7             AVAGSGAGAG AGAVAGSGAG AGAVAGSGAG ASA------- ---------- ---------- ----------
RN133           DGAVASA--- ---------- ---------- ---------- ---------- ---------- ----------
RN19            ---------- ---------- ---------- ---------- ---------- ---------- ----------
RN26            SGAGAGAGAV AGSGAGAGAV AGSGAGASA- ---------- ---------- ---------- ----------
RN131           GSAGGSAGSG D--------- ---------- ---------- ---------- ---------- ----------
K74             AGGSGSAGSG A--------- ---------- ---------- ---------- ---------- ----------
K386            SGAGAGAGAV AGSGAGAGAV AGSGAGASA- ---------- ---------- ---------- ----------
K60             AVAGSGAGAG AVAGSGAGAG AVAGSGAGAS A--------- ---------- ---------- ----------
TD530           GSAGASGSAG ASGSAGASGS AGASGSAGAS GSAGSGA--- ---------- ---------- ----------
TD554           GSAGGSAGGS AGGSAGGSAG GSAGGSAGGS AGGSAGGSAG GSAGSGD--- ---------- ----------
TD531           GSAGGSAGGS AGGSAGGSAG GSAGGSAGGS AGGSAGGSAG GSAGGSAGGS AGGSAGSGD- ----------
RN122           GSAGGSAGGS AGGSAGGSAG GSAGGSAGGS AGGSAGGSAG GSAGGSAGGS AGGSAGGSAG GSAGGSAGGS
MH20            ---------- ---------- ---------- ---------- ---------- ---------- ----------
K185            ASGSAGAGAS GSAGAGASGS AGAGASGSAG AGASGSAGAG ASGSAGAGAS GSAGSGA--- ----------
K58             ASGSAGAGAS GSAGAGASGS AGAGASGSAG AGASGSAGAS GSAGAGASGS AGSGA----- ----------
K392            ---------- ---------- ---------- ---------- ---------- ---------- ----------
TD515           ---------- ---------- ---------- ---------- ---------- ---------- ----------
TD529           ---------- ---------- ---------- ---------- ---------- ---------- ----------
TD508           ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     145        155        165        175        185        195        205
UB52            ---------- ---------- -----GNGAN PGADAERSPS TPA------- ---------- ----TTTTTT
K205            ---------- ---------- -----GNGAN PGADAERSPS TPA------- ---------- ----TTTTTT
RN70            ---------- ---------- ---------- ----AEGSPS TPA------- ---------- ----TTTTTT
RN130           ---------- ---------- ---------- ----AEGSPS TPA------- ---------- ----TTTTTT
MH24            ---------- ---------- ---------- ----AEGSPS TPA------- ---------- ----TTTTTT
TD533           ---------- ---------- ---------- ----AEGSPS TPA------- ---------- ----TTTTTT
MH18            ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
K391            ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
RN31            ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
K64             ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
K165            ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
TD504           ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
TD556           ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
UB59            ---------- ---------- ---------- ---DAERSPS TPA------- ---------- ----TTTTTT
UB7             ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
RN133           ---------- ---------- -----RNGAN PGADAKRSPS TPA------- ---------- ----TTTTTT
RN19            ---------- ---------- -----GNGAN PGADAEGSSS TPA------- ---------- ----TTTTTT
RN26            ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
RN131           ---------- ---------- -----GNGAN PGADAERSPS TPA------- ---------- ----TTTTTT
K74             ---------- ---------- -----GNGAN PGADAKRSPS TPA------- ---------- ----TTTTTT
K386            ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
K60             ---------- ---------- -----GNGA- ---DAKRSPS TPA------- ---------- ----TTTTTT
TD530           ---------- ---------- ---------- ---DAERSPS TPA------- ---------- ----TTTTTT
TD554           ---------- ---------- -----GNGAN PGADAERSPS TPA------- ---------- ----TTTTTT
TD531           ---------- ---------- -----GNGAN PGADAERSPS TPA------- ---------- ----TTTTTT
RN122           AGGSAGGSAG GSAGGSAGGS AGSGDGNGAN PGADAERSPS TPA------- ---------- ----TTTTTT
MH20            ---------- ---------- -----GNGAN PGADAERSPS TPATPATPAT PATPATPATP ATPATTTTTT
K185            ---------- ---------- ---------- ---DAERSPS TPA------- ---------- ----TTTTTT
K58             ---------- ---------- ---------- ---DAERSPS TPA------- ---------- ----TTTTTT
K392            ---------- ---------- -----GNGAN PGADAEGSSS TPA------- ---------- ----TTTTTT
TD515           ---------- ---------- -----RNGAN PGADAEGSSS TPA------- ---------- ----TTTTTT
TD529           ---------- ---------- -----RNGAN PGADAEGSSS TPA------- ---------- ----TTTTTT
TD508           ---------- ---------- -----RNGAN PGADAEGSSS TPA------- ---------- ----TTTTTT
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     215        225        235        245        255        265        275
UB52             TT------ND AEASTSTSSE NPNHNNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K205             TT------ND AEASTSTSSE NPNHNNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN70             TT------ND AEASTSTSSE NPNHNNAKTN PKGNGGVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN130            TT------ND AEASTSTSSE NPNHNNAKTN PKGNGGVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
MH24             TT------ND AEASTSTSSE NPNHNNAKTN PKGNGGVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD533            TT------ND AEASTSTSSE NPNHNNAKTN PKGNGGVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
MH18             TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K391             TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN31             TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K64              TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K165             TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD504            TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD556            TT------ND AEASTSTSSE NPNHKNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
UB59             TT------ND AEASTSTSSE NPNHNNAKTN PKGNGGVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
UB7              TT------ND AEASTSTSSE NPNHNNAETN PKGKGQVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN133            TT------ND AEASTSTSSE NSNHNNAETN PKGKGEVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN19             TT------ND AEASTSTSSE NPNHNNAETN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN26             TT------ND AEASTSTSSE NPNHNNAETN PKGKGQVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
RN131            TT------ND AEASTSTSSE NPNHNNAETN ---------- -QANKETQNN SNVQQDSQTK SNVPPTQDAD
K74              TT------ND AEASTSTSSE NPNHNNAETN PKGKGEVQEQ NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K386             TT------ND AEASTSTSSE NPNHNNAETN PKGKGQVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K60              TT------ND AEASTSTSSE NPNHNNAETN PKGKGQVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD530            TT------ND AEASTSTSSE NPNHNNAKTN PKGNGGVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD554            TT------ND AEASTSTSSE NPNHNNAETN ---------- -QANKETQNN SNVQQDSQTK SNVPPTQDAD
TD531            TT------ND AEASTSTSSE NPNHNNAETN ---------- -QANKETQNN SNVQQDSQTK SNVPPTQDAD
RN122            TT------ND AEASTSTSSE NPNHNNAETN ---------- -QANKETQNN SNVQQDSQTK SNVPPTQDAD
MH20             TTTTT---ND AEASTSTSSE NPNHNNAKTN PKGNGGVQEP NKANTETQNN SNVQQDSQTK SNVPPTQDAD
K185             TTTTT---ND AEASTSTSSE NPNHNNAKTN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K58              TTTTT---ND AEASTSTSSE NPNHNNAKTN PKGKGEVQKP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
K392             TTTTTTTTND AEASTSTSSE NPNHNNAKTN PKGNGGVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD515            TTTTTTTTND AEASTSTSSE NPNHNNAETN PKGNGKVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD529            TTTTTTTTND AEASTSTSSE NPNHNNAETN PKGNGKVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
TD508            TTTTTTTTND AEASTSTSSE NPNHNNAETN PKGNGKVQEP NQANKETQNN SNVQQDSQTK SNVPPTQDAD
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     285        295        305        315        325        335        345
UB52            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K205            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN70            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN130           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
MH24            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD533           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
MH18            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K391            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN31            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K64             TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K165            TKSPTAQPEQ AENSAPTAEQ TES------- ---------- ---------- ---------- ----------
TD504           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD556           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
UB59            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
UB7             TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN133           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN19            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN26            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN131           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K74             TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K386            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K60             TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD530           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD554           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD531           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
RN122           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
MH20            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K185            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K58             TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
K392            TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD515           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD529           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
TD508           TKSPTAQPEQ AENSAPTAEQ TESPELQSAP ENKGTGQHGH MHGSRNNHPQ NTSDSQKECT DGNKENCGAA
```

```
                ....|....| ....|....| ....|....|
                    355        365        375
UB52            TSLLNNSSNI AS-------- ----
K205            TSLLNNSSNI AS-------- ----
RN70            TSLLNNSSNI AS-------- ----
RN130           TSLLNNSSNI AS-------- ----
MH24            TSLLNNSSNI AS-------- ----
TD533           TSLLNNSSNI AS-------- ----
MH18            TSLLNNSSNI AS-------- ----
K391            TSLLNNSSNI AS-------- ----
RN31            TSLLNNSSNI AS-------- ----
K64             TSLLNNSSNI ASINKFVVLI SATL
K165            ---------- ---------- ----
TD504           PSLLSNSSNI AS-------- ----
TD556           PSLLSNSSNI AS-------- ----
UB59            TSLLNNSSNI AS-------- ----
UB7             PSLLSNSSNI AS-------- ----
RN133           TSLLNNSSNI AS-------- ----
RN19            TSLLNNSSNI AS-------- ----
RN26            TSLLNNSSNI AS-------- ----
RN131           TSLLNNSSNI AS-------- ----
K74             TSLLNNSSNI AS-------- ----
K386            TSLLNNSSNI AS-------- ----
K60             PSLLSNSSNI ASINKFVVLI SATL
TD530           TSLLNNSSNI AS-------- ----
TD554           TSLLNNSSNI AS-------- ----
TD531           TSLLNNSSNI ASINKFVVL- ----
RN122           TS-------- ---------- ----
MH20            TSLLNNSSNI AS-------- ----
K185            TSLLNNSSNI AS-------- ----
K58             TSLLNNSSNI AS-------- ----
K392            TSLLNNSSNI ASINKFVVLI SATL
TD515           TSLLSNSSNI AS-------- ----
TD529           TSLLNNSSNI AS-------- ----
TD508           TSLLNNSSNI AS-------- ----
```

A.3 The 17 sequences of MSP-2 FC27 family are shown below. Block identifications of this family are block 1 (positions 1-28), block 2 (positions 29-44), block 3 (positions 45-207), block 4 (positions 208-256) and block 5 (positions 257-321).

```
        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                5         15         25         35         45         55         65
RN68    ---------- YSNTFINNAY NMSIRRSMAN EGSNTNSVDA KAPNADTIAS GSQSSTNSAS TSTTNNGESQ
UB85    ---------- YSNTFINNAY NMSIRRSMAN EGSNTNRVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
UB27    ---------- YSNTFINNAY NMSIRRSMAN EGSNTTSVGA NAPNADTIAS GSQSSTNSAS TSTTNNGESQ
UB14    ---------- YSNTFINNAY NMSIRRSMAN EGSNTTSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
UB28    ---------- YSNTFINNAY NMSIRRSMAN EGSNTTSVGA NAPNADTIAS GSQSSTNSAS TSTTNNGESQ
UB82    VTFNIKNESK YSNTFINNAY NMSIRRSMAN EGSNTNRVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
UB50    ---------- YSNTFINNAY NMSIRRSMAN EGSNTTSVGA NAPNADTIAS GSQSSTNSAS TSTTNNGESQ
UB51    ---------- YSNTFINNAY NMSIRRSMAN EGSNTKSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
UB84    ---------- YSNTFINNAY NMSIRRSMAN EGSNTTSVGA NAPNADTIAS GSQSSTNSAS TSTTNNGESQ
MH32    ---------- YSNTFINNAY NMSIRRSMAN EGSNTKSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
MH10    ---------- YSNTFINNAY NMSIRRSMAN EGSNTTSVGA NAPNADTIAS GSQSSTNSAS TSTTNNGESQ
RN72    ---------- YSNTFINNAY NMSIRRSMAN EGSNTKSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
RN66    ---------- YSNTFINNAY NMSIRRSMAN EGSNTNRVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
RN129   ---------- YSNTFINNAY NMSIRRSMAN EGSNTNRVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
K66     VTFNIKNESK YSNTFINNAY NMSIRRSMAN EGSNTKSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
K397    ---------- YSNTFINNAY NMSIRRSMAN EGSNTKSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ
K195    ---------- YSNTFINNAY NMSIRRSMAN EGSNTKSVGA NAPKADTIAS GSQSSTNSAS TSTTNNGESQ

        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                75         85         95        105        115        125        135
RN68    TTTPTAADTI ASGSQRSTNS ASTSTTNNGE SQTTTPTAAD TIASGSQRST NSASTSTTNN GESQTTTPTA
UB85    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB27    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB14    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB28    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB82    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB50    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB51    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
UB84    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
MH32    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
MH10    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
RN72    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
RN66    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
RN129   TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
K66     TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
K397    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
K195    TTTPTA---- ---------- ---------- ---------- ---------- ---------- ----------
```

```
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               145        155        165        175        185        195        205
RN68      ADTPTATESN SPSPPITTT- ---------- ---------- ---------- ---------- -------ESS
UB85      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
UB27      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
UB14      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
UB28      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
UB82      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
UB50      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
UB51      ADTPTATESN SRSPPITTTE SNSRSPPITT TESNSRSPPI TTTESNSRSP PITTT----- -------ESS
UB84      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTTESNSPSP PITTT----- -------ESS
MH32      ADTPTATESN SRSPPITTTE SNSRSPPITT TESNSRSPPI TTT------- ---------- -------ESS
MH10      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
RN72      ADTPTATESN SRSPPITTTE SNSRSPPITT TESNSRSPPI TTTESNSRSP PITTTESNSR SPPITTTESS
RN66      ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
RN129     ADTPTATKSN SPSPPITTTE SNSPSPPITT TESNSPSPPI TTT------- ---------- -------ESS
K66       ADTPTATESN SRSPPITTTE SNSRSPPITT TESNSRSPPI TTTESNSRSP PITTTESNSR SPPITTTESS
K397      ADTPTATESN SRSPPITTTE SNSRSPPITT TESNSRSPPI TTTESNSRSP PITTTESNSR SPPITTTESS
K195      ADTPTATESN SRSPPITTTE SNSRSPPITT TESNSRSPPI TTTESNSRSP PITTTESNSR SPPITTTESS
```

```
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               215        225        235        245        255        265        275
RN68      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB85      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB27      SSGNAPNKTD GKGEESKKKN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB14      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB28      SSGNAPNKTD GKGEESKKKN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB82      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB50      SSGNAPNKTD GKGEESKKKN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB51      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
UB84      SSGNAPNKTD GKGEESKKKN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
MH32      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
MH10      SSGNAPNKTD GKGEESKKKN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
RN72      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
RN66      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
RN129     SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
K66       SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
K397      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
K195      SSGNAPNKTD GKGEESEKQN ELNESTEEGP KAPQEPQTAE NENPAAPENK GTGQHGHMHG SRNNHPQNTS
```

```
         ....|....| ....|....| ....|....| ....|....| ....|....|
             285        295        305        315        325
RN68     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB85     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB27     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB14     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB28     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB82     DSQKECTDGN KENCGAAPSL LSNSSNIASI NKFVVLISAT L
UB50     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB51     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
UB84     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
MH32     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
MH10     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
RN72     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
RN66     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
RN129    DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
K66      DSQKECTDGN KENCGAATSL LNNSSNIASI NKFVVLISAT L
K397     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
K195     DSQKECTDGN KENCGAATSL LNNSSNIAS- ---------- -
```

BIOGRAPHY

Mister Phumin Simpalipan was born on August 8, 1984 in Bangkok, Thailand. He graduated in high school level from Prakanongpittayalai School, Bangkok in 2001. He received his Bachelors Degree in Biology from the Faculty of Science, Chulalongkorn University in 2005. Then, he continued his Master Degree of Science at Chulalongkorn University. During the Master Degree education, he received research support scholarship from the Research Program on Conservation and Utilization of Biodiversity and the Center of Excellence in Biodiversity, Faculty of Science, Chulalongkorn University (CEB_M_49_2009).