

HUMAN GENETIC SUSCEPTIBILITY TO CLINICAL MALARIA



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
(BIOCHEMISTRY)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2008**

COPYRIGHT OF MAHIDOL UNIVERSITY

Thesis

Entitled

HUMAN GENETIC SUSCEPTIBILITY TO CLINICAL MALARIA

Mr. Bhee Witoonpanich

Mr. Bhee Witoonpanich

Candidate

Assoc. Prof. Sumalee Tungpradabkul

Assoc. Prof. Sumalee Tungpradabkul,

Ph.D. (Molecular Biology)

Major-Advisor

Assist. Prof. Anavaj Sakuntabhai

Assist. Prof. Anavaj Sakuntabhai,

M.D., D.Phil. (Clinical Medicine)

Co-Advisor

Assoc. Prof. Pratap Singhasivanon

Assoc. Prof. Pratap Singhasivanon,

M.B.B.S., D.T.M&H, M.P.H.,

Dr.P.H. (Epidemiology)

Co-Advisor

Prof. Banchong Mahaisavariya

Prof. Banchong Mahaisavariya, M.D.

Dean

Faculty of Graduate Studies

Assoc. Prof. Tuangporn Suthiphongchai

Assoc. Prof. Tuangporn Suthiphongchai

Ph.D. (Biochemistry)

Chair

Doctor of Philosophy Programme

in Biochemistry

Faculty of Science

Thesis

Entitled

HUMAN GENETIC SUSCEPTIBILITY TO CLINICAL MALARIA

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Doctor of Philosophy (Biochemistry)

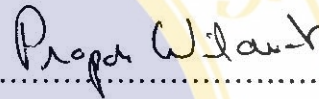
on

March 6, 2008



Mr. Bhee Witoonpanich

Candidate



Prof. Prapon Wilairat,

Ph.D. (Biochemistry)

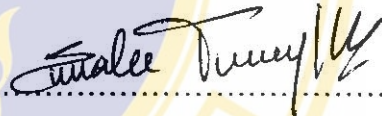
Chair



Lect. Prapat Suriyaphol,

Ph.D. (Biochemistry)

Member



Assoc. Prof. Sumalee Tungpradabkul,

Ph.D. (Molecular Biology)

Member



Prof. Jerapan Krungkrai,

Ph.D. (Biochemistry)

Member



Assist. Prof. Anavaj Sakuntabhai,

M.D., D.Phil. (Clinical Medicine)

Member



Prof. Banchong Mahaisavariya, M.D.

Dean

Faculty of Graduate Studies

Mahidol University



Prof. Skorn Mongkolsuk,

Ph.D. (Biological Science)

Dean

Faculty of Science

Mahidol University

ACKNOWLEDGEMENT

The success of this thesis can be attributed to the great assistance and support from my major advisor, Assoc. Prof. Sumalee Tungpradabkul and my co-advisor, Assist. Prof. Anavaj Sakuntabhai. I also would like to thank the following people;

The late Prof. Sornchai Looareesuwanfor and Assoc. Prof. Pratap Singhasivanon for their initiating ideas, expertise, insight and the project resources

Assist. Prof. Tanyachai Sura and his staffs for their help in DNA extraction.

Dr. Prapat Suriyaphol for providing me with many computer skills and advices.

Dr. Chayanon Peerapittayamongkol for his work in DNA collection and hemoglobin genotyping.

Dr. Chairat Turbpaiboon for his kindness, invaluable advices, and for his work in verifying and correction of the pedigree relationships.

Dr. Waraphon Phimraphi for her patience in providing me with the project information, and for her work in analysis and model fitting of the phenotypes.

Dr. Richard Paul for his work in analysis and model fitting of the phenotypes.

Dr. Simon Heath and the staffs at Centre National de Genotypage, Ivry, France, for their work in statistical analysis, microsatellite genotyping and the genome scan.

Dr. Denise Brocklebank for her advices and her work in pedigree partitioning.

The staffs of the Rajanagarindra Tropical Diseases International Center (RTIC) for their kindness and help in data and sample collection.

The staffs of the old GMIA unit for their help regarding the laboratory aspects.

I am grateful to the Medical Scholar Program (MSP), Mahidol University, for their financial support for studying in this Ph.D. program

I am grateful to the all the staffs of the French embassy, Bangkok, and the French government for their kindness, their help, and their financial support both for studying the language and for my stay in France.

I am greatly indebted to all the lab members from R303 and all the staffs of Department of Biochemistry, Mahidol University, for their consistent support.

Finally, I would like to thank my family for their love and support.

Bhee Witoonpanich

HUMAN GENETIC SUSCEPTIBILITY TO CLINICAL MALARIA

BHEE WITOONPANICH 4437256 SCBC/D

Ph.D. (BIOCHEMISTRY)

THESIS ADVISORS: SUMALEE TUNGPRADABKUL, Ph.D., ANAVAJ
SAKUNTABHAI, M.D., D.PHIL., PRATAP SINGHASIVANON, M.B.B.S.,
D.T.M&H, M.P.H., Dr.P.H.

ABSTRACT

Malaria, caused by protozoan parasites of the genus *Plasmodium*, is one of the most common infectious diseases and an enormous public health problem. Malaria episodes may vary considerably in their severity and clinical manifestations. So far, most genetic studies have been conducted in African countries where malaria caused by *P. falciparum*, the most lethal species, is endemic. Here, this study was based on the data of a longitudinal cohort study during 1994 to 2005 in the Karen population of 3,481 individuals living in Suan Phung district, Ratchaburi province, Thailand. This district is located near the Thai-Myanmar border and is an endemic area for malaria caused mostly by *P. falciparum* and *P. vivax*. Genetic heritability of the number of clinical malaria episodes and malaria parasite density phenotypes was assessed using the pedigree-based variance components model. Autosome-wide linkage analyses using variance components (VC) and Haseman-Elston (HE) approaches were then conducted on the phenotypes of which the heritability was significantly estimated. Several regions were identified with evidence for linkage but the most prominent one was on chromosome 3q27 in HE analyses of *P. vivax* asexual parasite density phenotypes with maximum LOD score of 5.89 (p -value $< 1 \times 10^{-5}$, locus-specific heritability of 60%; 95% confidence interval, 37.5%-82.5%). In this autosome-wide scan approach, the study provided for the first time systematic evidence of the genetic control of *P. vivax* parasite density as well as the selective pressure that this parasite species had exerted on the human genome.

KEY WORDS: GENETIC SUSCEPTIBILITY/ CLINICAL MALARIA/ GENOME-
WIDE SCAN/ *PLASMODIUM VIVAX*/ KAREN POPULATION

121 pp.

ลักษณะทางพันธุกรรมในคนที่มีความไวในการเกิดโรคมาลาเรีย (HUMAN GENETIC SUSCEPTIBILITY TO CLINICAL MALARIA)

พีร์ วิทยุพนิชย์ 4437256 SCBC/D

ปร.ด. (ชีวเคมี)

คณะกรรมการควบคุมวิทยานิพนธ์: สุมาลี ตั้งประดับกุล, Ph.D., อนวัช ศกุนตาทัย, M.D., D.PHIL.,
ประตাপ สิงหิวนนท์, M.B.B.S., D.T.M&H, M.P.H., Dr.P.H.

บทคัดย่อ

มาลาเรีย เป็น โรคติดเชื้อที่พบมากที่สุดโรคหนึ่งซึ่งเกิดจากปรสิตในตระกูล *Plasmodium* และก่อให้เกิดปัญหาด้านสุขภาพในประชากร ความรุนแรงของโรคและอาการแสดงออกอาจแตกต่างกันได้อย่างมากในการเป็นมาลาเรียแต่ละครั้ง จนถึงขณะนี้ การศึกษาทางด้านพันธุกรรมในคนที่เกี่ยวข้องกับการเกิดมาลาเรียนั้น มักจะทำการศึกษาในประเทศในแถบทวีปแอฟริกา ซึ่งมีการระบาดของมาลาเรียที่เกิดจากเชื้อ *P. falciparum* แต่ไม่มี *P. vivax* งานวิจัยนี้ ได้อ้างอิงข้อมูลจากการศึกษาการระบาดของมาลาเรียในกลุ่มประชากรกะเหรี่ยง 3,481 คนในอำเภอสวนผึ้ง จังหวัดราชบุรี ระหว่างปี พ.ศ. 2537 ถึง 2548 อำเภอสวนผึ้งอยู่ติดกับชายแดนไทยพม่าและการระบาดของมาลาเรียส่วนใหญ่เกิดจากเชื้อทั้ง *P. falciparum* และ *P. vivax* ในการศึกษาี้ ผู้วิจัยได้ทำการประเมินการถ่ายทอดทางพันธุกรรมของลักษณะของมาลาเรียได้แก่ การเป็นมาลาเรีย และปริมาณเชื้อในเลือดระหว่างการเป็นโดยใช้ pedigree-based variance components model จากนั้นเลือกเอาลักษณะที่พบว่าการถ่ายทอดทางพันธุกรรมมาทำ autosome-wide linkage analysis โดยวิธี variance components และ Haseman-Elston จากผลการวิจัย พบว่า หลายบริเวณบน autosome มีการถ่ายทอดไปกับลักษณะของมาลาเรีย แต่ที่ชัดเจนที่สุดคือ บริเวณโครโมโซม 3q27 โดยวิธี Haseman-Elston โดยบริเวณนี้มีการถ่ายทอดไปกับลักษณะปริมาณเชื้อ *P. vivax* แบบไม่มีเพศในเลือด ด้วยค่า LOD สูงสุดคือ 5.89 (ค่า p -value $< 1 \times 10^{-5}$, การถ่ายทอดทางพันธุกรรมเฉพาะบริเวณ 60% โดยที่ 95% confidence interval อยู่ระหว่าง 37.5%-82.5%) ในงานวิจัยนี้ เราได้แสดงหลักฐานให้เห็นเป็นครั้งแรกว่า ลักษณะทางพันธุกรรมมีผลต่อการควบคุมปริมาณเชื้อ *P. vivax* ในเลือด และเชื้อ *P. vivax* มีผลต่อการคัดเลือกทางพันธุกรรมในคน

121 หน้า

CONTENTS

	Page
ACKNOWLEDMENT	iii
ABSTRACT	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
CHAPTER	
1 INTRODUCTION	1
2 OBJECTIVES	3
3 LITERATURE REVIEW	4
1. Malaria	4
1.1 Malarial vectors and malaria life cycle	4
1.2 Malaria morbidity and mortality	7
1.3 Malaria diagnosis using microscopic examination	10
2. Genetics	11
2.1 Genetic material	11
2.2 The Hardy-Weinberg principle	17
2.3 Mendelian principles	19
2.4 Genetic linkage	19
2.5 Quantitative genetics	21
3. Familial Aggregation	22
3.1 Epidemiologic concepts of a disease	22
3.2 Familial factors as risk factors for a disease	23
3.3 Analysis of heritability	23
4. Linkage Analysis as a Tool for Disease Gene Mapping	25
4.1 Model-based or parametric linkage analysis	26
4.2 Model-free or non-parametric linkage analysis	27
5. Software Review	29

CONTENTS (continued)

	Page
6. Malaria and Human Genetic Studies	33
6.1 The red cell and malaria	33
6.2 Cytoadherence and malaria pathogenesis	37
6.3 Malaria and immune system	40
6.4 Malaria and family studies	44
4 MATERIALS AND METHODS	47
1. Subjects and Phenotypes Preparation	47
1.1 Study site	47
1.2 Family data	49
1.3 Pedigree partitioning	51
1.4 Data collection	52
1.5 Disease phenotypes	54
1.6 Data analyses for malaria and disease phenotypes	55
2. Genotype Preparation	58
2.1 DNA extraction, quantification and amplification	58
2.2 Microsatellite genotyping	65
3. Statistical Analysis	66
3.1 Estimation of IBD matrices	66
3.2 Estimation of heritability	66
3.3 Linkage analysis	67
5 RESULTS	70
1. Malaria Distribution During 8 Years of Study	70
1.1 Passive case detection	70
1.2 Active case detection	71
2. Pedigree Partitioning	75
3. Data Analyses for Malaria and Disease Phenotypes	75
4. Heritability to Clinical Malaria	81
5. Linkage Analysis	84
6 DISCUSSION	93

CONTENTS (continued)

	Page
7 CONCLUSION	99
REFERENCES	100
APPENDIX	118
BIOGRAPHY	121



LIST OF TABLES

Tables	Page
1. Genotype frequencies in Hardy-Weinberg equilibrium under random mating	18
2. Linkage analysis software	32
3. Common erythrocyte variants that affect resistance to malaria	36
4. Host molecules that mediated cytoadherence by <i>P. falciparum</i> -infected erythrocyte and that have been reported to show association with resistance or susceptibility to malaria	39
5. Immune genes reported to be associated with different malarial phenotypes	43
6. Genetic association with malaria in family studies	45
7. Disease phenotypes, and their methods of transformation	57
8. Number of malarial infection among 9 age-groups	73
9. Summary of samples used in statistical and genetic analyses	78
10. Significance of environmental variables in the statistical analyses performed for each phenotype	79
11. Percentage of variation explained by environmental factors in the statistical analyses	80
12. Estimation of heritabilities for number of clinical attacks, parasite density, proportion of clinical presentations, and non-malaria clinical presentations	82
13. The summary of the chromosome positions, methods of analyses, and the phenotypes of which the LOD score is higher than 1.5	92
14. Contribution (%) of genetic (heritability, h^2) and house (c^2) effects to variability in malaria and non-malaria clinical and biological phenotypes	119

LIST OF FIGURES

Figures	Page
1. Global distribution of dominant or potentially important malaria vectors	5
2. Diagram depicting the life cycle of human malaria	6
3. Malaria cases and elevation in Southeast Asia during 1998 to 2001	9
4. Human genome structure	15
5. The structure of the gene contains a series of exons (1 - 7) that are expressed in the mRNA, separated by a series of intervening sequences or introns (A - G) that have been "spliced out"	16
6. The Rajanagarindra Tropical Diseases International Center (RTIC)	48
7. Map of Tanaosri subdistrict showing the location of 7 hamlets which were selected to be the study site	50
8. Number of clinical malaria attacks by year during 1998-2005	72
9. Number of clinical malaria (all species), <i>P. falciparum</i> and <i>P. vivax</i> attacks in each age-group	74
10. Proportions of variation explained by genetic heritability and environmental factors found to have a significant effect on the phenotype measured	83
11. Linkage analysis results of number and proportion of clinical visits	86
12. Linkage analysis results of maximum, mean, and all <i>P. falciparum</i> parasite density	88
13. Linkage analysis results of maximum, mean, and all <i>P. vivax</i> parasite density	89
14. Linkage analysis results on chromosome 3 by Merlin regress with 3 <i>P. vivax</i> parasite density phenotypes	91
15. The distributions of residuals of the clinical phenotypes	118
16. The distributions of residuals of the parasite biological phenotypes	118

LIST OF ABBREVIATIONS

Abbreviation or symbol	Term
α	alpha
ANOVA	analysis of variance
β	beta
bp	base pair(s)
$^{\circ}\text{C}$	degree Celcius
cDNA	complementary DNA
cM	centiMorgan
conc	concentration
CR1	erythrocyte-complement receptor 1
DARC	Duffy antigen/chemokine receptor
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dsDNA	double strand deoxyribonucleic acid
ϵ	epsilon
EDTA	ethylenediamine tetraacetic acid
<i>e.g.</i>	example gratia
<i>et al.</i>	et alli
EtBr	ethidium bromide
FISH	Fluorescence <i>in situ</i> hybridization
γ	gamma
G6PD	glucose-6-phosphate dehydrogenase
h^2	heritability
Hb	hemoglobin
HbC	hemoglobin C
HbE	hemoglobin E
HbS	hemoglobin S

LIST OF ABBREVIATIONS (continued)

Abbreviation or symbol	Term
HLA	humeral leukocyte antigens
hr.	hour(s)
HS	hypersensitive site(s)
H ₂ O	water
IBD	identity-by-descent
IBS	identity-by-state
ICAM-1	intercellular adhesion molecule-1
IQ	intelligent quotient
IVS	intervening sequence(s)
km ²	square kilometre(s)
LCR	locus control region
LD	linkage disequilibrium
LOD	logarithm of odds
M	molar
mg	milligram(s)
µg	microgram(s)
MgCl	magnesium chloride
MHC	major histocompatibility complex
ml	millilitre(s)
µl	microlitre(s)
ng	nanogram(s)
No.	number
NO	nitric oxide
PCR	polymerase chain reaction
PCAM1	platelet-endothelial cell adhesion molecule
PfEMP-1	<i>P. falciparum</i> erythrocyte-membrane protein-1
pg	picogram(s)

LIST OF ABBREVIATIONS (continued)

Abbreviation or symbol	Term
<i>P. falciparum</i>	<i>Plasmodium falciparum</i>
<i>P. vivax</i>	<i>Plasmodium vivax</i>
<i>P. malariae</i>	<i>Plasmodium malariae</i>
<i>P. ovale</i>	<i>Plasmodium ovale</i>
<i>P. knowlesi</i>	<i>Plasmodium knowlesi</i>
RBC	red blood cell or erythrocyte(s)
ref.	reference(s)
RFLP	Restriction fragment length polymorphism
SDS	Sodium dodecyl sulfate
sec	second(s)
SNP	single nucleotide polymorphism
SOLAR	sequential oligogenic linkage analysis routines
spp.	specie(s)
TDT	transmission disequilibrium test
TE	Tris- ethylenediamine tetraacetic acid
TNF	tumour necrosis factor
V	voltage
Ve	environmental variance
Vg	polygenic variance
VNTR	variable number of tandem repeat
WBC	white blood cell or leucocyte(s)
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

Malaria remains the most important human parasitic disease worldwide causing over 350–500 million clinical cases per year and between 1.1 to 2.7 million deaths (1). Most of the treatments available have limited efficacy and side effects as well as the emergences of drug-resistant strains of the parasite. Vaccines are being developed with varying degree of success.

The common symptoms of clinical form of malaria are fever, chills, headaches, and diaphoresis. However, these signs and symptoms are generally nonspecific. The diagnosis of uncomplicated malaria is normally based on a history of fever and a positive blood slide. The severe form of malaria includes cerebral malaria with symptoms such as severe anemia and renal failure (1).

There were several convincing evidences that human genetic factors play an important role in determining the outcome of malarial infection. It was first introduced by Haldane's hypothesis that malaria had exerted considerable selective pressure on human genome and there was the likely evolutionary driving force behind common erythrocyte variants especially in tropical countries (2). An example of this is the high gene frequency of HbS in Africa that confers protection against severe malaria in heterozygotes but causes fatal sickle cell disease in homozygotes. Other hemoglobinopathies such as glucose-6-phosphate dehydrogenase (G6PD) deficiency, Melanesian Ovalocytosis and the thalassaemias (reviewed in (3, 4)) interfere with parasite survival or ability to re-invade red blood cells but have not led to the development of new treatments because each mutation carries serious consequences for the host when homozygous (5). There are several other studies that support the genetic role in determining malarial infection. Levels of parasitemia appear to be linked to 5q31-q33 in Cameroon and Burkina Faso (6, 7). Susceptibility to malarial fever appears to be linked strongly to the major histocompatibility complex (MHC) in

a Gambian twin study (8), and *TNF-376A* and *TNF-238A* are associated with cerebral malaria and severe malarial anemia respectively in The Gambia (9). In Gambian children, HLA-B53 has protective effect against severe malaria (10, 11), and in Gabon, protection from severe malarial anemia has been associated with a single nucleotide polymorphism of the inducible nitric oxide synthase promoter region (*NOS2*) (12, 13).

Despite these findings, very few systematic analyses have been reported to address human genetics more comprehensively especially on clinical malaria. To our knowledge, this research is the first autosome-wide linkage analysis on clinical malaria in Asian population. It is based on the data of local residents of Suan Phung district, Ratchaburi province, located near the Thai-Myanmar border which is an endemic area for malaria in Thailand. Almost 3,481 participants have been followed up for a number of clinical malaria attacks, types of malaria (*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovalae* and mixed infection), parasite densities and clinical responses to treatment during 1998-2005. Family structures have been constructed and the total pedigrees comprise 2,428 individuals. Phenotypes were transformed by controlling confounding factors such as age, sex, and hamlet using multivariate regression analysis. The phenotypes' residuals were then estimated for each individual using Genstat statistical package version 7.1. Genetic contributions to the phenotypes were performed by estimation of heritability using SOLAR program version 2.1.4. The identity by descent (IBD) matrices estimation was done by Loki version 2.4.5. Variance component based and Haseman-Elston's regression based linkage analyses were then performed by SOLAR version 2.1.4 and Merlin version 1.1.1 respectively to find region(s) of the genome contributed to the differences of the phenotype between each individual.

CHAPTER 2

OBJECTIVES

The family-based clinical malarial phenotype analyses were performed on Karen population to determine the heritability or the genetic contribution to malaria risk. The quantitative traits used, after adjusting for environmental factors like age, sex, hamlet, dates and their interactions, are such as number of positive *P. falciparum* or *P. vivax* visits (attacks) and parasite density. The autosome-wide linkage analyses were then performed to locate the position(s) of the quantitative trait locus(loci) contributed to variation of the traits in the population.

CHAPTER 3

LITERATURE REVIEW

1. Malaria

Malaria is caused by protozoan infection of red blood cells with *Plasmodium* spp. Five species of protozoan parasite of the *Plasmodium* genus - *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and potentially *P. knowlesi* (14, 15)- cause malaria in humans. Malaria caused by *P. falciparum* is most common and most lethal (1).

1.1 Malarial vectors and malaria life cycle

Out of about 3,500 known species of mosquitoes worldwide, only around 30 to 40 species of Anopheles mosquitoes are capable of transmitting the infection (Figure 1). Anopheles mosquito is required for the transmission of malaria from one human host to another taking up gametocytes in the human blood meal. Appropriate development of the parasite in the mid-gut and salivary gland of the mosquito is then necessary to complete the sporogonic cycle (Figure 2).



Figure 1. Global distribution of dominant or potentially important malaria vectors
(adapted from www.medicalecology.org/.../print_malaria.htm)

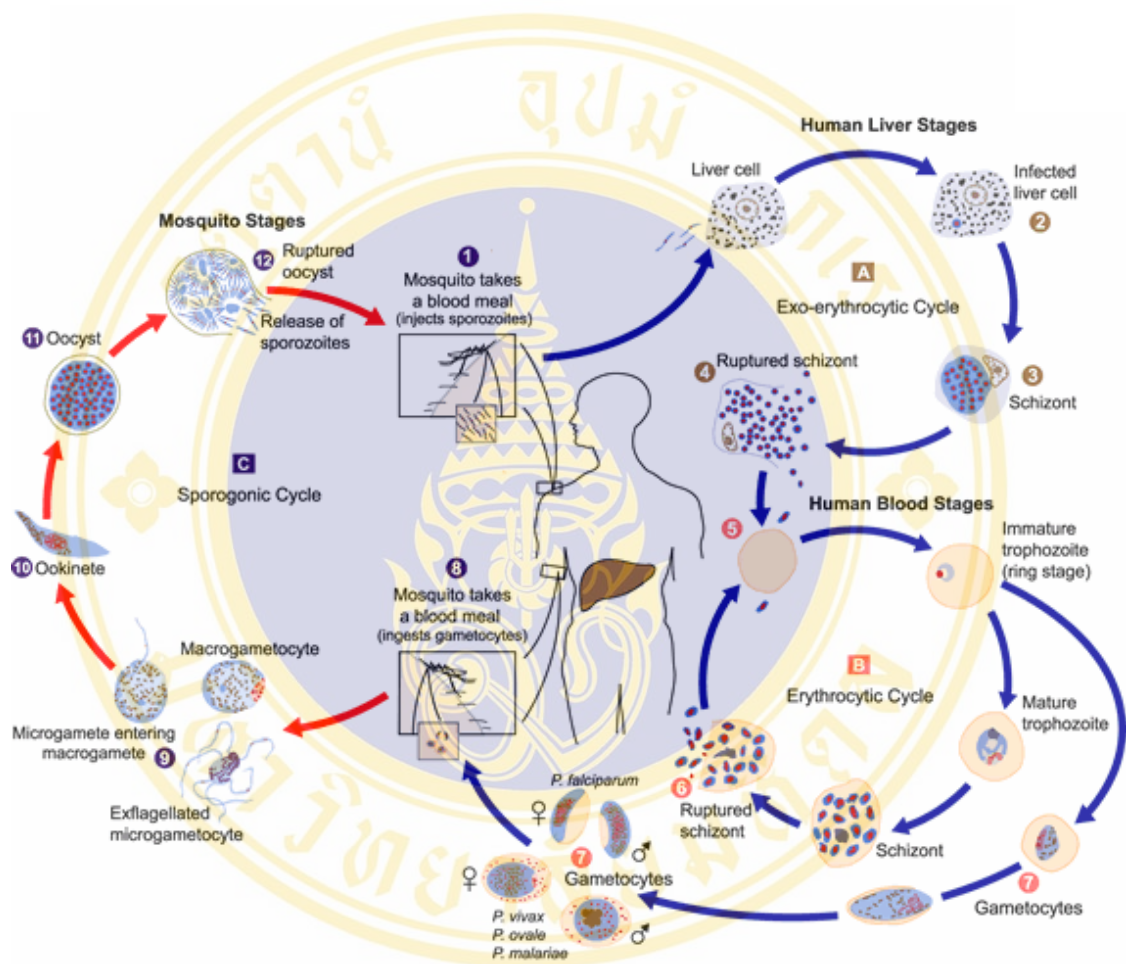


Figure 2. Diagram depicting the life cycle of human malaria (source:http://www.uni-tuebingen.de/modeling/Mod_Malaria_Cycle_en.html)

1.2 Malaria morbidity and mortality

The common symptoms of clinical form of malaria are fever, headache, back pain, chills, sweats, myalgia, nausea, vomiting, diarrhea, and cough. The clinical manifestations of the severe form which will occur if *Plasmodium falciparum* infection is left untreated include coma, severe anemia, renal failure, respiratory distress syndrome, hypoglycemia, shock, spontaneous hemorrhage, and convulsions (16).

Most malaria clinical cases occur in Africa and out of an estimated 1.1–2.7 million deaths each year, over 90% are children under 5 years of age, mainly from cerebral malaria and anemia (1). The immunological status of a person is one of the factors that play an important role in determining the effect of malaria on human health and in the intensity of disease transmission. In endemic areas of malaria, children under 5-6 years old are those most at risk of acquiring severe malaria (17) whereas in older children and adults who have partial immunity, the parasite density is observed to decrease with increasing age (18). In non-endemic areas, severe malaria affects adults and children as well as non-immune travelers and migrants.

In Southeast Asia, the number of malaria cases had increased during 1998 to 2001 (Figure 3). In Thailand, for the past four decades, the number of malaria cases throughout the country has been markedly reduced from the morbidity of about 268 per 1,000 populations in 1947 to 1.9 per 1000 population in 1994. However, the mortality rate of 1.7 per 100,000 in 1994 and 1.6 per 100,000 populations in 1997 has shown that the disease is still an important cause of death, although reported malaria mortality has continuously decreased since 1991 (19). Most malaria cases are found at the border of Thailand especially Thai-Myanmar and Thai-Cambodia border but the endemic areas are in the hilly or forested areas only. The presence of efficient vectors and the frequency of population movement makes malaria transmission in forested areas intense (20).

In 2004, WHO reported that there were 37,335 malaria cases in Thailand and 325 cases died. There were all four *Plasmodium* spp. infections but the major ones were *P. falciparum* (50.9%) and *P. vivax*. The primary malaria vectors were *Anopheles minimus*, *An. dirus* and *An. maculatus*. Malaria chemoprophylaxis was not recommended in Thailand because of high level of *P. falciparum* resistance to

antimalarial drugs including chloroquine, mefloquine, sulfadoxine pyrimethamine combination, and the serious, diminished sensitivity to quinine. But, the treatment regimens using mefloquine or its combination with artemisinin derivatives remained effective.



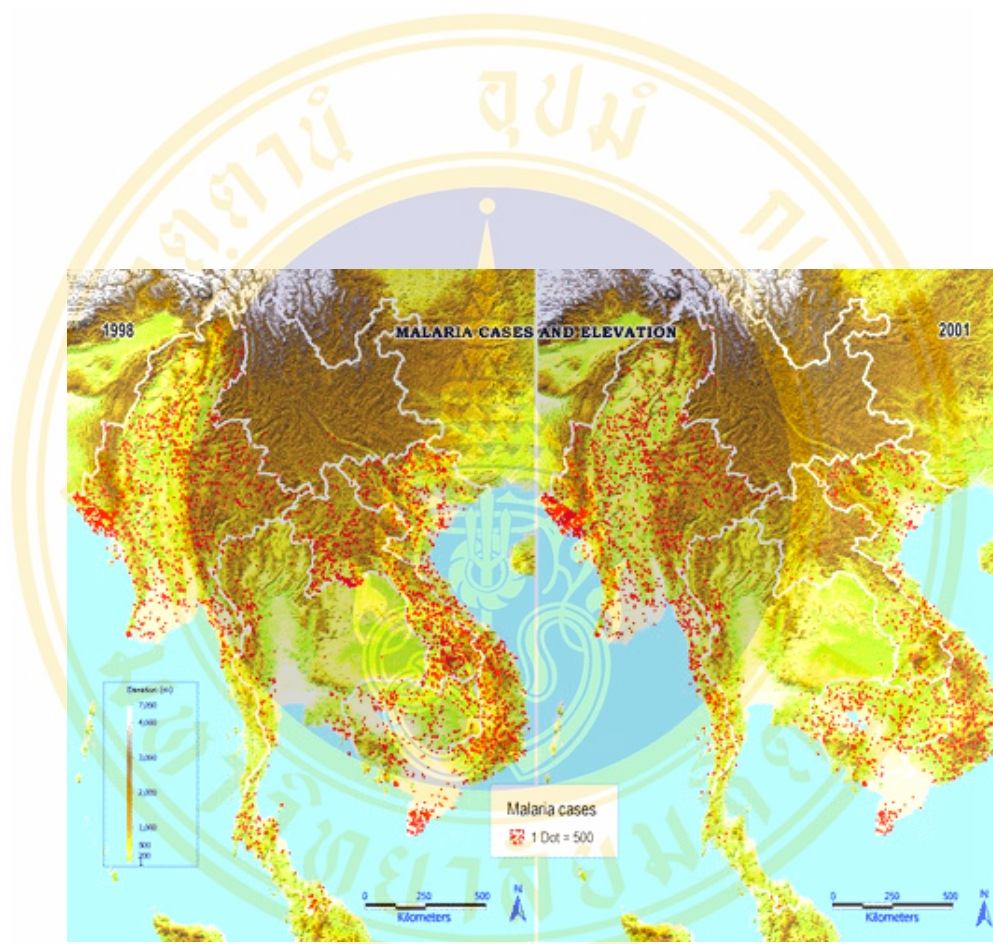


Figure 3. Malaria cases and elevation in Southeast Asia during 1998 to 2001 (source: www.thaitravelclinic.com/faq.htm)

1.3 Malaria diagnosis using microscopic examination

1.3.1 Examining thick smear

Since the erythrocytes (RBCs) are lysed and the parasites are more concentrated, the thick smear is useful for parasites screening and for detecting mixed infections. Malaria parasites are made a tentative species determination on the thick smear and then the thin smear is examined to determine the species present. Most often, thin smear is the appropriate sample for species identification. For determination of "No Parasites Found" (NPF): for malaria diagnosis, WHO recommends that at least 100 fields, each containing approximately 20 leucocytes (WBCs), be screened before calling a negative thick smear. Assuming an average WBC count of 8,000 per microliter of blood, this gives a threshold of sensitivity of 4 parasites per microliter of blood. In nonimmune patients, symptomatic malaria can occur at lower parasite densities, and screening more fields (e.g., 200, 300, or even the whole smear) might be warranted, depending on the clinical context and the availability of laboratory personnel and time.

1.3.2 Examining thin smear

Thin smear is useful for species identification of parasites already detected on thick smears; screening for parasites if adequate thick smears are not available, and a rapid screen while a thick smear is still drying.

1.3.3 Quantifying parasites

In some malaria cases, quantification of parasites yields clinically useful information. If this information is needed by a physician, malarial parasites can be quantified against blood elements such as RBCs or WBCs. To quantify malarial parasites against RBCs, the parasitized RBCs are counted among 500-2,000 RBCs on thin smear and express the results as % parasitemia.

$$\% \text{ parasitemia} = (\text{parasitized RBCs} / \text{total RBCs}) \times 100$$

If the parasitemia is high (e.g., > 10%), 500 RBCs are examined. If it is low (e.g., <1%), 2,000 RBCs (or more) are examined. Asexual blood stage parasites and gametocytes are counted separately.

To quantify malarial parasites against WBCs, on the thick smear, the parasites are counted against WBCs, 500 parasites or 1,000 WBCs, whichever comes first. The result is expressed as parasites per microliter of blood, assuming 8,000 WBCs per microliter of blood.

$$\text{Parasites/ } \mu\text{l blood} = (\text{parasites/WBCs}) \times \text{WBC count per } \mu\text{l (or 8,000)}$$

Result in % parasitized RBCs and parasites per microliter of blood can be interconverted if the WBC and RBC counts are known or otherwise by assuming 8,000 WBCs and 4×10^6 RBCs per microliter of blood.

2. Genetics

2.1 Genetic material

To understand the role of genetic factors in the occurrence of disease in human populations, it is important to have a basic knowledge of the structure and function of the genetic material as well as the principles underlying its transmission in families and populations.

Although cell biologists in the late 1800s suspected that the colored material in the cell nucleus (termed chromatin) carried the genetic material of the cell, it was not until the beginning of the twentieth century that the chromosomes (colored bodies) were actually described. The chromatin encapsulated in the nucleus is a complex of polymers of deoxyribonucleic acid (DNA) and basic proteins called histones.

2.1.1 Chromosomes

Human nuclear DNA is organized into 22 pairs of autosomal chromosomes and 2 sex-specific chromosomes (called X and Y) that become visible microscopically only during cell division. Since humans are diploid, there are two copies of each autosome in the cell nucleus; each parent donates one of each pair. Each chromosome is composed of two arms separated by centromere; by convention the shorter arm is denoted “p” and the longer arm is denoted “q”. In most cell divisions, the duplicated nuclear material divides into daughter cells by a process called mitosis, the only cell

division process for all somatic cells. In reproductive or germinal cells, a different process called meiosis occurs.

2.1.2 Genes

While nuclear DNA is organized into these separate and distinct chromosomes, the basic unit of hereditary information is the gene or locus, which codes for some gene products (such as an enzyme or a structural protein). Eukaryotic genes are composed of lengths of DNA in which the coding sequences (exons) are interrupted by intervening sequences (IVS or introns, Figure 5). Genes can still be thought of as discrete units located at specific sites along the chromosome, that is the terms gene and locus can be used interchangeably.

At a given locus, many different forms of gene representing individual mutations may exist, and these are termed alleles. Each person carries two copies of every autosomal gene, and these alleles may or may not be different. Heterozygous individuals have two different alleles, while homozygous individuals have two copies of the same alleles. Females also carry two alleles at all loci on the X chromosome, but males are hemizygous for every locus on both the X and Y chromosome. Many genes have one or several alleles represented at high frequencies in populations, and such loci are said to be polymorphic. Typically, a polymorphism exists when the less frequent allele has a frequency of more than 1% in a population (21), or to be strict, a polymorphism is the existence of two different alleles at significant frequencies in the population (22).

2.1.3 Genetic markers

A genetic markers is a polymorphic DNA or protein sequence deriving from a single chromosomal location (22). Genetic markers are not causally related to any disease but can be extremely useful for genetic linkage or cosegregation which can be demonstrated between the marker locus and a locus controlling a disease.

In the early 1980s, DNA polymorphism provided, for the first time, a set of markers that were sufficiently numerous and spaced across the entire genome. DNA markers have an additional advantage that they can all be typed by the same technique. Moreover, their chromosomal location can be determined using FISH or

hybrid mapping, allowing DNA-based genetic maps to be cross-referenced to physical map. The development of DNA markers allow human gene mapping to start in earnest.

2.1.3.1 Restriction fragment length polymorphisms (RFLPs)

The first generations of DNA markers were restriction fragment length polymorphisms (RFLPs). RFLPs were initially typed by preparing Southern blots from restriction digests of the test DNA, and hybridizing with radio-labeled probes. This technology required plenty of time, money and DNA. Nowadays this is less of a problem because RFLPs can usually be typed by polymerase chain reaction (PCR). A sequence including the variable restriction site is amplified; the product is incubated with the appropriate restriction enzyme and then run out on a gel to see if it has been cut. A more fundamental limitation is their limited informativeness. RFLPs have only two alleles: the site is present or it is absent. The maximum heterozygosity is 0.5. Disease mapping using RFLPs is frustrating because there are too often that a key meiosis in a family turns out to be uninformative.

2.1.3.2 Minisatellites

Minisatellite VNTR (variable number tandem repeat) markers were a great improvement. The VNTRs have many alleles and high heterozygosity. Most meioses are informative. However, the technical problems of Southern blotting and radioactive probes were still an obstacle to easy mapping, and VNTRs are not evenly spread across the genome.

2.1.3.3 Microsatellites

The advent of PCR finally made mapping relatively quick and easy. Minisatellites are too long to amplify well, and so the standard tools for PCR linkage analysis are microsatellites. These are mostly (CA)_n repeats. Tri- and tetranucleotide repeats are gradually replacing dinucleotide repeats as the markers of choice because they give cleaner results – dinucleotide repeat sequences are peculiarly prone to replication slippage during PCR amplification. Each allele gives a little ladder of ‘stutter band’ on a gel, making it hard to read. Much effort has been devoted to

producing compatible sets of microsatellite markers than can be amplified together in a multiplex PCR reaction and give non-overlapping allele sizes, so that they can be run in the same gel lane. With fluorescent labeling in several colors, it is possible to score perhaps ten markers on a sample in a single lane of an automated gel.

2.1.3.4 Single nucleotide polymorphisms (SNPs)

After 10 years of developing more and more polymorphic markers, it may seem perverse that the newest generations of markers are 2-allele single nucleotide polymorphisms. They include the classic RFLPs, but also polymorphisms that do not happen to create or abolish a restriction site. The advantage of SNPs is that they can be scored on solid-state arrays without recourse to gel electrophoresis (23). The gain in throughput is more than offsets in the lower informativeness of SNPs. Typically, the test DNA is PCR amplified in a very large multiplex and hybridized to an array comprising a series of anchored oligonucleotide primers, each terminating with a polymorphic nucleotide. A single primer extension step is carried out on the array, using a mixture of four fluorescently-labeled dideoxynucleotides. Label adds to primers that perfectly match the test DNA, but not to those with 3' mismatch. Reading the cells of the array for presence or absence of fluorescence allows the types for every SNP on the array to be read off. Although the technology is still being put together, it is hoped that the array of a few thousand primers can be used to genotype markers spaced closely across the whole genome in single chip hybridization (22).

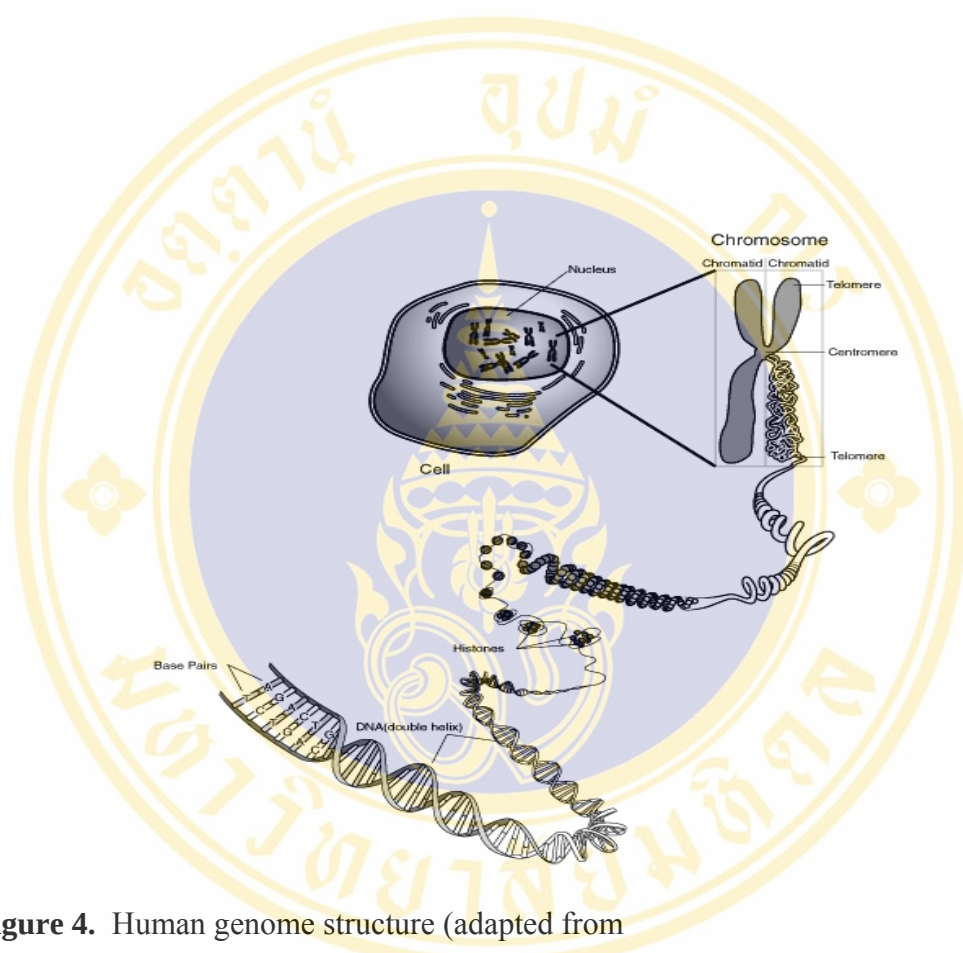


Figure 4. Human genome structure (adapted from <http://employees.csbsju.edu/hjakubowski/classes/ch331/bind/olbindtranscription.html>)

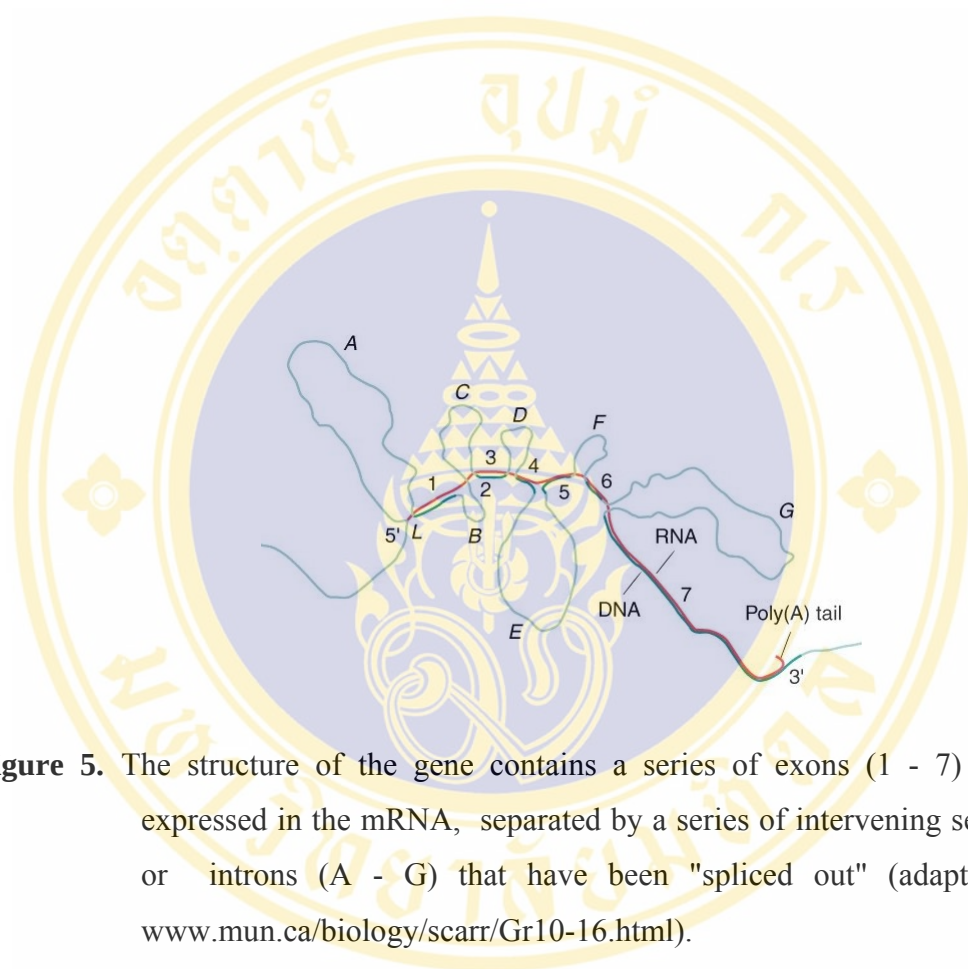


Figure 5. The structure of the gene contains a series of exons (1 - 7) that are expressed in the mRNA, separated by a series of intervening sequences or introns (A - G) that have been "spliced out" (adapted from www.mun.ca/biology/scarr/Gr10-16.html).

2.2 The Hardy-Weinberg principle

The milestone of population genetics is summarized in Hardy-Weinberg principle which was independently derived by G.H. Hardy and W. Weinberg in 1908. This principle offered a simple explanation for how genetic diversity for Mendelian traits can be maintained in the population. The Hardy-Weinberg law also defined the simple relationship between the frequency of genes in a population and the frequency of genotypes (i.e., individuals), and is therefore useful for estimating allele frequencies from the prevalence of Mendelian disease (24).

Since at any one autosomal locus, an individual carries two alleles (one paternal and one maternal), the population's pool of all alleles at a single autosomal locus can also be represented by the familial Punnett square. For example, if A and a are alleles at a single locus, and if the relative frequency of A is denoted p, with the relative frequency of a denoted q (where $p+q = 1$), then the proportion of genotypes of a random-mating population are established, working out the expected frequencies of genotypes among offspring in the next generation and showing that an equilibrium exists. Specifically, the frequencies of the three genotypes (AA, Aa, and aa) will be the same in all subsequent generations, that is, p^2 , $2pq$, q^2 . Thus, these genotype frequencies represent in Hardy-Weinberg equilibrium. The derivation is somewhat different for X-linked genes but the logic is identical. Similarly, Hardy-Weinberg equilibrium can be worked out for multiple alleles and multiple loci (21).

Table 1. Genotype frequencies in Hardy-Weinberg equilibrium under random mating

	gametes with A	gametes with a	
gametes with A	p^2	pq	$p^2 + pq = p(p+q) = p$
gametes with a	pq	q^2	$q^2 + pq = q(p+q) = q$
	p	q	1.0



2.3 Mendelian principles

Recent advance in human genetics have altered the concept of structure of genes. Furthermore, evidence from other species raises the possibility that transposition of genetic material, gene conversions, amplification and imprinting may underlie some diseases previously thought to follow simple *Mendelian inheritance*. Nevertheless, the principles of Mendelian transmission, both in families and in populations, still hold, and these serve as the milestone of genetic analysis.

The first cohesive of explanation of the inheritance of discrete traits was developed by Gregor Mendel in 1865. From his breeding experiments with pea plants, Mendel developed the concept that pair of unobserved “factors” was responsible for physical traits in individuals, and single member of these pairs were transmitted unaltered from parent to offspring. Mendel’s results are summarized into two basic principles: (1) the principle of segregation of alleles, and (2) the principle of the independent assortment.

Mendelian principles have been successfully applied to the study of transmission of single-gene traits and disease in families. Family data generally are collected as pedigrees, although sibships or other fixed structure groups of relatives can also be used to test specific hypotheses about genetic control. The term pedigree come from the French “pied de grue” which literally means crane’s foot. Like the branch footprint of the bird, the pedigree represents the branching process by which genes are transmitted from one generation to the next (21).

2.4 Genetic linkage

Genetic linkage represents departures from Mendel’s second law, the independent assortment of genes, and is the reflection of the cosegregation of alleles at loci physically located near one another on the same chromosome. While gametes produced by two independently segregation genes contain the four combinations of alleles in equal proportion, gametes arising from two linked loci will have a predominance of alleles paired in the same combinations seen in their parents. Due to crossing over or recombination during meiosis, however, linked genes are not always transmitted together. The amount of recombination depends on the physical distance between linked genes along the chromosome. The longer of the distance is the more

opportunity for meiotic crossing over, resulting in observable genetic recombination. Conversely, the shorter the physical distance, the less chance the recombination will occur and the tighter the linkage.

In the whole population, genetic linkage creates only minor imbalances in the possible combinations of genotypes, and overall there may be no discrepancy in the expected proportion of the various combinations of alleles at two linked loci (i.e., certain haplotypes or combinations of alleles at linked loci). At equilibrium, the frequencies of pairs of alleles at different loci (i.e., haplotypes) are simply the product of their respective allele frequencies. When there is an excess/deficiency of certain haplotypes, however, the loci are said to be in linkage disequilibrium. Linkage disequilibrium will decay with time, but this is a function of the recombination between loci. The rate of decay can be quite slow for tightly linked loci, and may span scores of generations. Observed disequilibrium in a population may merely reflect a transition point on the way to true multilocus equilibrium, or it may be due to recent merging of two or more subpopulations with different genetic frequencies. Random fluctuation in haplotype frequencies due to small population sizes, or possibly selection for certain multilocus combinations, should also be considered (21).

Although extreme linkage disequilibrium between a marker locus and a disease locus can result in a statistical association between a marker allele and the disease when population samples are examined, it is important not to confuse the effects of linkage with a more general association between a marker allele and a disease. Association can occur between an allele at a marker locus and a disease due to a number of complex biological interactions, including any in which the marker allele plays a direct or indirect role in the pathogenesis of disease. Linkage, however, is a property of the relative position of the loci, not their alleles, and the observed cosegregation within a given family can involve any allele at the marker locus.

The demonstration of genetic linkage in humans and the localization of genes on specific chromosomes can be accomplished by both statistical and experimental methods. Statistical methods involve linkage analysis of two or more genetic traits (i.e., two markers or a marker and a disease locus), using pedigree data. On the other hand, experimental methods to map a locus to a chromosome include rodent-human

somatic cell hybridization techniques, in situ hybridization, chromosome sorting techniques, as well as other techniques.

2.5 Quantitative genetics

The polygenic model of inheritance bridged the gap between Mendelian genetics, which traditionally involved the study of discrete traits (most frequent diseases) and biometric or quantitative genetics, which traditionally involved the study of continuous traits. Quantitative genetics developed along with and served as a catalyst for much of statistics. Many of those who developed the early principles of statistics were motivated by questions of familial resemblance and human genetics. Several common statistical techniques (correlation, regression, analysis of variance) were first developed to measure genetic factors underlying continuously varying traits. These traits showed obvious familial aggregation but were quite intractable under the principles of simple Mendelian genetics.

In domestic organisms such as plants, it is possible to modify Mendel's original experimental design to study quantitative traits. In human populations, it is more difficult to use the techniques of statistical genetics to extrapolate the genetic mechanisms, and the earliest work in quantitative genetics was aimed at simply measuring the degree of familial resemblance. Later workers showed that such regression to the mean could easily result if the observed phenotype were influenced by both genetic and environmental factors acting independently. Since parents with high phenotypic values can transmit only the genetic components to their offspring, their offspring would be expected to be closer to the population mean. If the genetic component to phenotype is the result of many independent genes each contributing in an additive fashion (i.e., polygenes), it will serve as an estimator of the proportion of phenotypic variance which is attributed to genetic differences, a quantity known as heritability (the ratio of genetic variance to the total variance). In 1918, R.A. Fisher showed that if there were no dominance for a large number of independently segregating loci, it is possible to express the observed correlation between types of relatives as a linear function of the phenotypic variation directly attributable to genetic differences and that attributable to nongenetic differences. Effects of these independent polygenes are summarized by genetic components of variance (21).

3. Familial Aggregation

The study of familial aggregation is a central theme genetic epidemiology in which major methodological and statistical advances have been recently made. Family studies encompass three approximately sequential step of scientific investigation (25). First researchers attempt to determine whether or not there is evidence of familial aggregation in a disease or trait. If familial aggregation is found, the second step is to discriminate among environmental, cultural, and/or genetic factors that may contribute to this clustering. Third, when evidence of a role for genetic factor is found, analyses designed to test for specific genetic mechanisms are carried out. Some basic epidemiologic principles and approach can be applied (1) to detect the presence of familial aggregation and (2) to evaluate whether familial aggregation can be attributed to genetic or environment factors. Note that unraveling the cause of familial aggregation does not constitute an end in itself in genetic epidemiology, but it is an important step in defining the role of genetic factors contributing to disease occurrence in the population at large. Epidemiologic principles and approaches are crucial not only for detecting the presence of familial aggregation, but also for application of specific genetic models.

3.1 Epidemiologic concepts of a disease

Familial aggregation studies provide a unique framework for evaluating epidemiologic concepts of a disease, for assessing the spectrum of expression of a disease, and for identifying its heterogeneity. As in all epidemiologic studies, the case definition is an important prerequisite. Traditionally, studies of familial aggregation begin by identifying a group of individuals with a specific disease (index cases) and determining whether relatives have an excess frequency of the same disease when compare to an appropriate reference population. While most often the phenotype of interest is a disease, it can also be a physiologic trait that has continuous distribution (26-28).

The study of familial members of index cases can provide an opportunity for understanding the familial nature of a disease or its precursors and risk factors. These risk factors are frequently quantitative measures of traits associated with a chronic disease (29). Familial aggregation studies can also be used to point out common

etiologic and pathogenetic mechanisms between different diseases (30, 31). Conversely, familial studies can reveal underlying heterogeneity within a disease (32).

3.2 Familial factors as risk factors for a disease

Although epidemiologists have long been interested in familial factors as risk factors for a disease (33), most epidemiologic studies do not collect or analyze family history data (34). In case-control studies, a positive family history for a disease is often treated as a risk factor. Positive family history is usually defined as the presence of the disease in one or more first degree relatives (parent, siblings and offsprings) for both cases and controls. The analytic strategy is to then test whether or not a positive family history confers any excess risk for the disease. However, relying exclusively on causal family histories obtained from the interviews of cases and controls has serious limitations and can lead to errors of interpretation (35).

Family history depends on many factors such as the number of relatives, their biological relationship to the index case, their age distribution, and the disease frequency in the population.

3.3 Analysis of heritability

After performing familial aggregation for a trait or a disease and showing that this aggregation can not be explained completely by environmental risk factors, the next logical step is to ask how much of the familial aggregation can be attributed to genetic causes. Heritability is typically used to answer this question because it is defined as the proportion of variation directly attributable to genetic differences among individuals relative to the total variation in a population (which includes variation due to both genetic and nongenetic factors).

The concept of heritability was developed as a part of quantitative genetics which focused exclusively on continuous traits (such as height, IQ, serum cholesterol). The statistical techniques used in quantitative genetics were largely derived from analysis of variance (ANOVA). However, the concept of heritability can also be applied to qualitative traits by changing the question to ask how much of the variation in the risks of a disease can be attributed to genetic difference among individuals (21).

3.3.1 The concept of heritability

In quantitative genetics, phenotypes are measured on a continuous scale, and the observed distribution in a sample is frequently unimodal giving little or no evidence for any finite number of distinct genotype. The goal of quantitative genetics is the same as traditional human genetics, which is to identify possible genetic mechanisms controlling the distribution of phenotypes in both family and populations.

Because the phenotypes in quantitative genetics are continuous, summary statistics (mean, variances, and correlations) must be used to describe a particular sample or population. Genetic analysis frequently focuses only on correlations or covariances among relatives and attempts to partition these observed correlations (or covariances) into components attributable to shared genes and shared environments. From these components, heritability can then be calculated.

3.3.2 Definitions of heritability

Heritability is defined as a ratio of variances, specifically the ratio of the genetic variance (i.e., those attributable to genotypic or allelic differences to the total phenotypic variances in the population). Two types of heritability should be distinguished. First, there is heritability in the broad sense, which is the ratio of variance attributed to all gene differences among individuals to the total phenotypic variance. Second, there is heritability in the narrow sense, which is the ratio of variance contributed by additive effects of alleles at one or more loci to the total phenotypic variance. Both of these measures of heritability are ratio of a genetic component of variance to the total variance, and thus they must both lie between 0 and 1. The lower limit of zero heritability occurs when all the observed phenotypic variation is attributable to nongenetic factors, while the upper limit of 100% heritability exists when there is no phenotypic variation not due to genetic differences.

When studying human populations, therefore, it is always important to consider the impact of small differences in these sources of variation that may well vary across populations. In general, greater diversity of environmental factors present in a population will lower the estimated heritability, while populations with more homogeneous environments will give higher estimates of heritability, even though the biological mechanism underlying the trait may be identical. A population that is

relatively genetically homogeneous will produce a lower estimate of heritability than will a genetically heterogeneous population (21).

3.3.3 Variance components model

Most of the quantitative genetics involves more elaborate statistical models for estimating heritability. These techniques require additional assumptions about the underlying biological mechanisms, but they do permit partitioning phenotypic variation into both genetic and nongenetic components, and they do allow greater predictions about the structure of population under various conditions.

The primary statistical tool for quantitative genetics is analysis of variance, where the observed phenotypic variance is partitioned into components reflecting differences in unobserved genetic factors and environmental factors, both of which must be estimated from covariances and correlations among relatives.

4. Linkage Analysis as a Tool for Disease Gene Mapping

Similar phenotypes may often reflect the effects of multiple genes, environmental factors and their interactions and thus obscure the inheritance pattern of common diseases. To reduce the expected genetic heterogeneity and environmental noise, and to improve the power of linkage methods for common diseases, several investigators have taken approaches such as;

- 1) Suitable selection of the sample. For example, related individuals with the same form of a disease's severity are more likely to share the same genetic determinant.
- 2) Focusing the linkage studies on a genetically homogeneous population (36). The shared environment and life style of genetically isolated populations can be expected to enhance the signal-to-noise ratio of linkage studies and reduced disease-influencing variants comparing to outbred populations is very likely.
- 3) Focusing on quantitative traits that are risk factors for the disease. The disease itself may not be etiologically as homogeneous as the underlying quantitative phenotypes that are indicators of disease development. For example, central obesity, serum lipid levels, insulin level and blood

pressure are etiologic factors underlying the pathogenesis of cardiovascular diseases and each of them is under a varying degree of genetic control. Therefore, a state of the disease among affected individuals might be an insensitive indicator in which there is possibly high number of susceptibility loci controlling each underlying process. Thus, focusing on the intermediate quantitative phenotypes may be more representative of the variation observed among different individuals (37).

Quantitative trait locus (QTL) is the term used to call a locus underlying variation of a quantitative phenotype. Specific effects of QTL alleles on quantitative phenotypes are sometimes observable when they are at or near loci of known function (candidate genes). The effect on the phenotypes can be either direct or indirect through linkage disequilibrium.

When the genetic effect of one or more QTLs is unobservable, their locations and effects are estimated through linkage analysis. There are 2 approaches for linkage analysis; parametric or non-parametric.

4.1 Model-based or parametric linkage analysis

The principal method of analysis for disease gene mapping of Mendelian traits in pedigrees is the lod-score method (38, 39). The basis of the lod-score method is to discover how frequent two loci are separated by meiotic recombination or known as recombination fraction (θ). The smaller the θ , the closer the 2 loci locate in the genome. Recombination fractions define genetic distance; two loci that show 1% recombination are defined as being 1 centimorgan (cM) apart on a genetic map. Although the total chiasma counts between human male and female are different, the approximation of 1 cM = 1 Mb (megabase; physical distance) is very useful (22). Recombination fractions do not exceed 0.5 no matter how large the physical distance is. It will be equal to 0.5 if the 2 loci are far apart on the same chromosome, or on different chromosome, and in this case, the 2 loci will segregate independently. By using maps of markers with known location in the human genome, we can thus identify the location of the disease gene of interest.

In standard linkage analysis, the recombination fraction θ is estimated by likelihood-based statistical methods (40-42). This linkage analysis is also known as

model-based, or parametric, as it requires a precise genetic model, detailing the mode of inheritance, gene frequencies and penetrance of each genotype. For quantitative traits, this standard lod-score method requires the specification of genetic model, QTL parameters, including allele frequencies, mean effect of the genotypes and genetic variance in the population. The likelihood measures the probability of the observations based on assumptions about genetic model parameters, and on the value of the recombination fraction (37). The overall likelihood of the pedigree is calculated on the alternative hypothesis that the loci of interest are linked (recombination fraction = θ) or not linked (recombination fraction = 0.5). The ratio of these 2 likelihoods gives the odds of linkage, and the logarithm of the odds is defined as the lod score.

Once the likelihood is defined, the lod score is calculated for a range of θ values varying between 0 (complete linkage) and 50% (independence, or no linkage). The most likely value of the recombination fraction is the one that maximizes the lod score. The threshold for accepting linkage is the lod score greater than or equal to 3. To be precise, a lod score of 3.3 has been shown to provide a genome-wide significant linkage with a 5% chance of error (p -value = 0.05) accounting for the multiple tests performed while searching linkage to a disease locus throughout the whole genome (43).

Nevertheless, an overestimation of the recombination fraction will occur if the trait model is not particularly accurate. This is also the case for complex nonmendelian diseases or conditions; the genetic model cannot be specified because these diseases or conditions are much less tractable.

4.2 Model-free or non-parametric linkage analysis

The development of model-free or non-parametric method of linkage analysis is the consequence of the high parameterization and the difficulties in obtaining correct parameter estimates of lod-score methods. These methods are based upon identical by descents sharing among relatives, evaluated by observing the segregation pattern of polymorphic genetic markers in the pedigrees. It is also important to distinguish segments identical by descent (IBD) from those identical by state (IBS). Alleles IBD are demonstrably copies of the same ancestral (usually parental) allele, whereas IBS

alleles look the same, and may have the same DNA sequence, but they are not derived from a known common ancestor or their common origin cannot be unequivocally determined (22). In these methods, 2 relatives affected by the same disease are expected to share the same IBD alleles at the disease susceptibility loci and at all markers tightly linked to it. The proportion of the marker alleles that a group of affected relatives of a given degree will share IBD can be predicted when there is no linkage between the disease and the marker itself (44). Then a deviation from the expected IBD sharing towards increased observed IBD sharing is thus considered as an indication of linkage between the disease and the marker locus. For quantitative traits, the degree of IBD sharing among different relatives is correlated to trait similarity, either by regression or by variance component (VC) analysis.

In VC linkage analysis, maximum likelihood method is used to estimate the components that model the phenotypic covariance among relatives. For a simple model in which a quantitative trait is influenced by one QTL and residual polygenic effects, the overall phenotypic variance of the trait is modeled as the sum of the phenotypic variances due to additive effects of the QTL, an additive polygenic effect, assumed to be due to a large number of unlocalized loci acting additively, and a random environmental deviation. Using the VC model, the null hypothesis can be tested whether the additive genetic variance due to the QTL equals zero (no linkage) by comparing the likelihood of this model with that of an alternative model in which the additive effects of the QTL is estimated. A likelihood ratio test evaluates the statistical evidence for linkage and the result is presented as a lod score.

The first regression method for quantitative trait mapping was developed by Haseman and Elston (45). It is based on the regression of the trait-squared difference in a sibpair on their estimated proportion of IBD sharing at a marker. The marker is considered linked to a trait when small differences in the trait are associated with high IBD sharing, and *vice versa*. Modifications of the original test were subsequently proposed to improve its power, in particular, the extension of the regression-based methods to the analysis of different types of relative pairs, as all these methods were limited to the analysis of sibpair data (46).

Compared with VC methods, regression methods are computationally less demanding and are more robust to violation of normality of the trait distribution.

However, when the distribution of the trait is approximately normal, VC linkage methods are more powerful. The VC approach is noteworthy for its generality and flexible modeling capability, as virtually any type of effect and interaction can be easily incorporated.

5. Software Review

Localization and characterization of quantitative trait loci (QTLs) and causal polymorphisms influencing complex phenotypes are major goals in statistical genetic analyses. The important steps in this process are Linkage analyses of QTLs and association studies to correlate phenotypes and genotypes. These investigations are supported by the development of a variety of computer software allowing fast and efficient analysis of quantitative traits.

“The most commonly used methods for quantitative trait linkage analyses in humans are the variance components and Haseman-Elston (H-E) approaches” (47). Parametric, model-based methods can also be used but these require the specification of allele frequencies at the trait locus and genotype-specific trait means.

In variance components approaches, the covariance among family members is modeled as a function of unspecified aggregate additive genetic effects, effects due to a hypothetical gene in the region being tested for linkage, and a residual component that is uncorrelated among individuals which is sometimes described as an environmental component (48-50). Maximum likelihood methods are used to estimate these components of variance in most variance component programs and numerous complexities can be added onto variance component linkage models such as dominance genetic effects, epistatic interactions, gene-environment interactions, shared environment correlations, spouse correlations, corrections for non-normality of the trait distribution, estimation of empirical p-values, estimation of linkage power for a given study and multivariate models. Different programs provide automated routines for different subsets of these variance component extensions. GENEHUNTER automates the inclusion of dominance components and has an extensive set of linkage and association tests and it is a standard for statistical genetic analysis (51). “A disadvantage of this position is that any new program will aspire to

improve on GENEHUNTER, so that for many of its functions there are now other programs with better performance” (52).

The Markov Chain Monte Carlo (MCMC)-based approach implemented in Loki also estimates the variance due to a QTL, but adds the number of QTLs influencing the trait and their allele frequencies (53). This model can be expanded to incorporate dominance effects, epistatic and gene–environment interactions, and other more complex models. Unlike most programs which provide a LOD (logarithm of odds) score as a measure of the evidence, Loki estimates the posterior probability for any given chromosome region of at least 1 trait locus being in that region.

The source of the multipoint identity by descent (IBD) matrices that are used to estimate QTL-specific variance in a linkage analysis has become one of the greatest differences between specific implementations of the variance component linkage method. The exact method by GENEHUNTER and MERLIN use a Lander-Green algorithm (41) for which computational time increases exponentially with the increasing numbers of non-founders in a pedigree. Pedigree with more than 25 individuals cannot generally be analyzed before breaking up into smaller units. Mendel uses either the Elston–Stewart or the Lander–Green algorithm, depending on pedigree size. SOLAR uses an extension of the Fulker–Cardon interval approach to estimating multipoint IBDs (49, 54). Pedigrees with unlimited complexity and an unlimited number of genotyped markers can be analyzed with this interval approach. However, this approximation performs well for markers that are individually quite informative (such as short tandem repeats) but not for markers that are less informative (such as single nucleotide polymorphisms). Loki uses MCMC methods to estimate IBDs which are also approximate, but are more precise than the Fulker–Cardon interval approach. The linear increase in computational time for MCMC IBD estimation with the increase of the number of markers and the pedigree size allows the use of pedigrees with arbitrary size and complexity and unlimited number of markers. The only drawback is that it is more computationally intensive than the interval approach, and may take weeks to compute for pedigrees of 100 individuals or more. The MCMC approximation method by SIMWALK2 also has good report of accuracy, although the program has a lot of tuning parameters (42). Several programs (SOLAR,

ACT, SEGPATH and SAGE) can also import IBD matrices from a variety of sources if they are in the right program-specific format.

“Because a model with no QTL effects, with all genetic effects in an unspecified aggregate genetic component, forms the basis of comparison for the likelihood ratio test of linkage, most variance components programs also provide an overall estimate of the trait heritability” (47).

In the H-E linkage approaches, the squared difference in siblings' trait values is modeled as a function of their IBD allele sharing at a particular chromosomal location (45). The general H-E method has several extensions including the 'revisited' H-E which uses the mean corrected cross-product of the siblings' trait values (55) but is less powerful than the original H-E in some cases (56), and a variety of 'weighted' H-E tests using functions of the original and revisited H-E (57, 58). The latest version of the H-E which is very similar to variance component approach models the full variance-covariance matrix within a family (59). The major difference is that the estimation of various components is done by regression rather than maximum likelihood. Like variance component approach, it can be easily extended to include other components such as epistatic interactions and gene-environment interactions. The original H-E linkage approach is implemented in SAGE and GENEHUNTER. The latest expansion of the H-E is also available in SAGE. The MERLIN REGRESS routine has the implementation of an H-E extension developed by Sham et al.(46). The squared trait sums and differences between relative pairs are modeled as a function of their sharing IBD allele.

Regression approaches are computationally more efficient and more robust to non-normality of the trait distribution, whereas maximum likelihood approaches are, in theory, more powerful (47). Table 2 gives the summary of linkage analysis software discussed here.

Table2. Linkage analysis software (47).

Name	Linkage methods	Platform	References
ACT	Variance components	UNIX	(48, 60)
GENEHUNTER	Variance components; Haseman-Elston	UNIX; Mac OS X	(61, 62)
Loki	MCMC based multipoint linkage analysis	UNIX; Linux; Mac OS X ; FreeBSD ; OpenBSD ; OSFI; Irix; AIX	(53)
Mendel	Variance components; parametric	UNIX; Linux; Windows; Mac OS X; OSF/Tru64	(63-65)
Merlin	Variance components; Haseman-Elston	Unix; Linux	(66)
Mx	Variance components	UNIX; Linux; AIX; Irix; VMS ; some executables for Mac OS X; GUI version for Windows; DOS	(67-69)
SAGE	Haseman-Elston; parametric	UNIX; Linux; Windows; Tru64	(70)
SEGPATH	Variance components	UNIX	(71, 72)
SOLAR	Variance components	UNIX; Linux; Solaris; Irix	(49)

6. Malaria and Human Genetic Studies

The genetic basis of resistance to malaria is complex with a lot of different genes and their interactions with environmental variables and parasite genetic factors involved.

Regions with high levels of malaria transmission are usually selected for the study of susceptibility and resistance to malaria. Different genetic factors may control the risks for developing parasitemia in an exposed person, for becoming ill with malaria fever in a parasitemic person, and for developing severe malaria in a person with malaria fever. Parasitemia and fever can be regarded as quantitative phenotypes that are ascertained by repeated measurement within the community, whereas severe malaria is a qualitative phenotype that is typically ascertained in hospital-based studies (73). Genetic factors are expected to be detected at each stage in the causal chain of disease progression.

6.1 The red cell and malaria

A key step in the molecular process of erythrocyte is the invasion by *P. vivax*. The Duffy antigen, encoded by *FY* gene, is a chemokine receptor (DARC) that is expressed in various cell types. The Duffy antigen is expressed in erythrocytes in most populations (but not in sub-Saharan Africa). The disruption in a promoter SNP that is a binding site for the erythroid transcription factor GATA-1 (74) completely prevents *P. vivax* from invading erythrocytes, and accounts for the remarkable absence of *P. vivax* in parts of Africa (75). There are evidences that strongly suggest recent positive selection of African Duffy-negative allele, denote “*FY*O* (76, 77). Further support for selective pressure at this locus comes from the observation of an entirely independent *FY* polymorphism emerging in Papua New Guinea. This polymorphism decreases Duffy-antigen expression and thus reduces *P. vivax* invasion efficiency (78). What puzzles malariologists is the inclination of positive selection at the *FY* locus, preventing *P. vivax* infection which is generally not lethal. It has even been proposed that *P. vivax* infection interferes with the infection of much more lethal parasite *P. falciparum* (79).

G6PD gene on chromosome X encodes the enzyme glucose-6-phosphate dehydrogenase (G6PD) which produces the reduced state of the electron donor

nicotinamide adenine dinucleotide phosphate, conferring an important form of defense against oxidative stress within erythrocytes. Many different variants of *G6PD* have been geographically associated with malaria (80), and analysis of haplotypic structure at the *G6PD* locus supports the hypothesis of recent positive selection (81, 82). Having the common African form of G6PD deficiency (G6PD A⁻), female heterozygotes and male hemizygotes have ~50% reduced risk of severe malaria in Gambian and Kenyan children (83). In Nigerian children, protection against severe malaria correlates with deficiency in G6PD enzyme activity (84). Its proposed mechanism of protection is the reduction of parasite replication in G6PD-deficient erythrocytes (85). However, the parasite seems to produce its own G6PD to counter this (86).

Invasion of erythrocyte by *P. falciparum* is compromised by genetic deficiency of glycophorin A and B; the erythrocyte-membrane sialoglycoproteins which determine various blood groups (87). Glycophorin C, a minor component of erythrocyte membrane, serves as a receptor for *P. falciparum*. A deletion of *GYPC* exon 3, resulting in the Gerbich-negative blood group, reduces the invasion by *P. falciparum*. It is commonly found in coastal area of Papua New Guinea (88).

Case/control studies designed according to the definition of the spectral nature of malaria disease (i.e. severe compare to mild) (10, 89) have demonstrated associations of host polymorphisms with risk of severe malaria. There are evidences that β globin polymorphisms namely HbS (a glutamic acid→valine substitution at codon 6), Hb C (a glutamic acid→lysine substitution at codon 6), and HbE (a glutamic acid→lysine substitution at codon 26) have been associated with protection against malaria. Individuals heterozygous for sickle cell trait (hemoglobin AS) do not have any clinical abnormality. They have ~10-fold increased protection from severe malaria but a lesser degree of protection against milder forms (10, 90-92). The mechanisms of protection of HbAS are still unknown but 2 plausible mechanisms are suppression of parasite growth in red cells (93) and enhanced splenic clearance of parasitized erythrocytes (94). The protective effect of HbAS against malaria increased from 20% to 56% between ages of 2 and 10 years in Kenyan children implicating that it enhances or acts in synergy with the acquired immune response (95). Protection against severe malaria have been observed in both heterozygotes and homozygotes of

HbC (96-98), but homozygotes appears to have substantially greater protective effect (99). The proposed mechanism of protection of HbC is the increase of immune clearance of infected erythrocytes (100). For HbE, erythrocytes from HbE-heterozygous individuals have been observed to resist the invasion by *P. falciparum*. There is no evidence though, that HbE protects against severe malaria (101).

Since Haldane's insight that the geographic distribution of β -thalassemia may be due to the heterozygous condition affording protection against malaria, several other host erythrocyte polymorphisms such as α -thalassemia have been identified to be geographically associated with malaria (3). Much higher allele frequency of α -thalassemia, and much lower incidence of malarial illness were observed in the Tharu people than were other ethnic groups living in the same region of Nepal (102). The frequency of α^+ -thalassemia was highly correlated with malaria endemicity in parts of Melanesia (103). Both heterozygous and homozygous α^+ -thalassemia were protected against severe malaria in Kenyan children (104), whereas only heterozygotes were protected in Ghanaian children (105). In Papua New Guinea, the risk of severe malaria as well as other childhood infections was found to be reduced in heterozygous α -thalassemic children implicating that the protection did not seem to be malaria-specific (106).

"Band 3 protein," an erythrocyte-membrane and anion exchanger has been implicated in malaria resistance. A form of ovalocytosis that is common in parts of Southeast Asia resulted from a 27 bp deletion in this gene appears to be protective against either malarial infection (107, 108) and cerebral malaria (109, 110). The summary of common erythrocyte variants that affect resistance to malaria is given in Table 3.

Table 3. Common erythrocyte variants that affect resistance to malaria

Protein	Reported genetic association with malaria	References
Duffy antigen	<i>FY*O</i> allele completely protects against <i>P.vivax</i> infection	(74-76)
G6PD	G6PD deficiency protects against severe malaria.	(80, 82, 83, 85, 111)
Glycophorin A, B, and C	<i>GYPA</i> -, <i>GYPB</i> - and <i>GYPC</i> -deficiency RBC are resistant to invasion of <i>P. falciparum</i> .	(87, 112-114)
Beta-globin	<i>HbS</i> and <i>HbC</i> alleles protects against severe malaria. <i>HbE</i> allele reduces parasite invasion.	(3, 90-92, 95, 96, 98, 99, 101)
Alpha-globin	Alpha-thalassemia protects against severe malaria but appears to enhance mild malaria in some environments.	(104-106, 115)
Band 3 protein	Deletion causes ovalocytosis but protects against malarial infection and cerebral malaria.	(107-110)

6.2 Cytoadherence and malaria pathogenesis

One of the problems for malaria vaccine development is a remarkable degree of antigenic variation of the parasite. This is caused by switching expression of many different copies of the genes which encode the parasite's ligands. Among these genes, a gene family called "var" encodes the major ligand *P. falciparum* erythrocyte-membrane protein-1 (*PfEMP-1*). By constantly switching between different forms of *PfEMP-1* that bind to different host receptors (and different parts of the same receptor), the parasite varies its pattern of sequestration in different organs (116). Severe malaria has been associated with polymorphisms of host receptors for cytoadherence by *P. falciparum* infected erythrocytes. One of these receptors is the CD36 antigen, encoded by *CD36*, which binds parasitized erythrocytes to dendritic cells. This event appears to make the dendritic cell lost its function of presenting parasite antigens. Several *CD36* polymorphisms have been reported in malaria endemic region (117, 118), but in terms of disease-association studies, the results are confusing. Homozygotes for a nonsense polymorphism was associated with susceptibility to cerebral malaria in Gambian and Kenyan case-control study (117), but it was found that in Kenya alone, the heterozygotes was associated with protection against severe malaria (119). In Thailand, it has been reported that protection against cerebral malaria has been associated with a dinucleotide repeat sequence in intron 3 (118).

Another endothelial-binding receptor, intercellular adhesion molecule-1 (ICAM1), encoded by *ICAM1*, binds strongly to some isolates of *P. falciparum*. Reduction in binding efficiency was observed with a polymorphism in the N-terminal domain, present at high frequencies in African populations (120). The low binding allele was found to be associated with reduce susceptibility to severe malaria in Gabon (121), but increased susceptibility in Kenya (120), and no significant effect in the Gambia (122).

A phenomenon known as rosetting, where a parasitized erythrocyte binds to other erythrocytes is an alternative for parasite sequestration. Binding to erythrocyte-complement receptor 1 encoded by *CR1* through *PfEMP-1* is a mechanism for rosetting (123). In Papua New Guinea, erythrocyte *CR1* deficiency has been found to be associated with both *CR1* gene polymorphisms and with α^+ thalassemia which is

also common (124). Resistance to severe malaria was found to be independently associated with both *CR1* gene polymorphisms and α^+ thalassemia. However, no evidence of association was found between *CR1* polymorphisms and disease severity in the Gambia (122, 125). In Thailand, an RFLP that is associated with reduced expression of *CR1* on erythrocytes was reported and it was found that individuals homozygous for this were susceptible to severe malaria (126). The summary of host molecules that affect cytoadherence and malaria pathogenesis is given in Table 4.



Table 4. Host molecules that mediated cytoadherence by *P. falciparum*-infected erythrocyte and that have been reported to show association with resistance or susceptibility to malaria.

Gene	Reported genetic association with malaria	References
<i>CD36</i>	<i>CD36</i> polymorphisms show variable associations with severe malaria in the Gambia, Kenya and Thailand.	(117-119)
<i>ICAM1</i>	<i>ICAM1</i> polymorphisms show variable associations with severe malaria in Kenya, Gabon and Gambia.	(120-122)
<i>CR1</i>	<i>CR1</i> polymorphisms show variable associations with severe malaria in the Gambia, Thailand and Papua New Guinea	(122, 124, 126)

6.3 Malaria and immune system

Erythrocytes do not express MHC molecules therefore limit the opportunities for antigen presentation and immune recognition in malaria. Nevertheless, the fact that liver cells express MHC class I molecules makes it possible for cytotoxic T cell (CTL) recognition and responses when the parasites replicate in the liver cell prior to invading erythrocytes. Both MHC class I heavy chain, an exceptionally polymorphic gene encoded by *HLA-B*, and $\beta 2$ microglobulin comprise the *HLA-B* antigen-presentation complex. The *HLA-B53* allele which is extremely common in West Africa was found to be associated with a significantly reduced risk of severe malaria in Gambian children (10). Given that *HLA-B* is expressed by liver cells but not by erythrocytes, this association implies that naturally acquired immunity may significantly target liver-stage parasites.

The HLA class II β chain encoded by *HLA-DRB1*, together with α chain, comprises the *HLA-DR* antigen-presenting complex. This complex which is crucial for antibody production is expressed in B lymphocytes, dendritic cells, and macrophages. In Gambian children, a polymorphism of *HLA-DRB*; *HLA DRB1*1302-DQB1*0501*, has been associated with resistance to severe malaria (10). Also in Gambian children, it is reported that the distribution of MHC class II haplotypes shows an overall association with the incidence of malaria-fever episodes (127).

Interferon- γ , a key immunological mediator encoded by *IFNG*, is believed to play both a protective and a pathological role in malaria. In Gambian children, it was found that SNPs in the region of *IFNG* gene were weakly associated with severe malaria but no clear cut effect (128). The ligand-binding α chain of the interferon- γ receptor encoded by *IFNGR1* was studied in the Gambia and it was found that, in Mandinka which is the major ethnic group, heterozygotes for the *IFNGR1-56* polymorphism were associated with protection against cerebral malaria (129).

Interferon α receptor 1, a type I membrane protein encoded by *IFNAR1*, is one of the two chains of a receptor for interferon α and β . Two *IFNAR1* SNPs were associated with protection against severe malaria in a Gambian case-control study, and a resistant haplotype was identified (130).

Interleukin-10, encoded by *IL10*, is a key anti-inflammatory cytokine which has many functions including the down-regulation of the expression of Th1 cytokines,

MHC class II, and costimulatory molecules on macrophages. *IL10* has shown protective effect against severe malaria in several studies. In Gambian children, a common haplotype of *IL10* SNPs was strongly associated with protection against severe malaria by case-control but not TDT analysis (131).

IL12B encodes a subunit of interleukin-12 or T cell stimulating factor; a cytokine involving in the differentiation of naive T cells into Th1 cells, which is important in resistance against pathogens. Homozygotes for an *IL12B*-promotor polymorphism were found to be associated with reduced NO production measuring in blood samples, and this genotype has been associated with a mortality in cerebral malaria in Tanzanian but not in Kenyan children (132).

Interleukin-4, encoded by *IL4*, is a cytokine that induces differentiation of naive helper T cells (Th0 cells) to Th2 cells. Upon activation by IL-4, Th2 cells subsequently produce additional IL-4. In the Fulani of Burkina Faso, West Africa, the *IL4*-524T allele was associated with elevated antibody levels against malaria antigens. This might be a possible factor of increased resistance to malaria and explained why the Fulani have fewer malaria attacks than do other neighboring ethnic groups (133).

Nitric oxide (NO) synthase, encoded by *NOS2A*, synthesizes nitric oxide (NO) from the terminal nitrogen atom of L-arginine in the presence of NADPH and dioxygen (O₂). NO is a free radical and a key biological messenger. It also has antiparasitic, potential immunosuppressive, and neurotransmissive properties and has been proposed as a factor in cerebral malaria. In a study in Gabon, the *NOS2A*-954C allele has been associated with resistance to severe malaria and reinfection (13, 134) although no association was found by the studies in the Gambia and Tanzania (135). The *NOS2A*-1173T allele is associated with protection from malarial illness in Tanzania and from severe malarial anemia in Kenya (136), but no protective effect against severe malaria in the Gambia. In Gambian children, associations were found between a *NOS2A* microsatellite polymorphism and susceptibility to fatal malaria (12), and a haplotype carrying the *NOS2A*-1659T allele and cerebral malaria (135).

The *TNF* gene encoded tumor necrosis factor is located in the MHC class III region. It is a multifunctional proinflammatory cytokine secreted predominantly by monocytes/macrophages and is critical for innate immunity against malarial parasites (137). In the Gambia and Burkina Faso, the populations were tested for linkage

between polymorphisms within the MHC region and mild malaria. Multipoint analysis have identified linkage between the MHC region and susceptibility to malaria fever episode, with a peak close to *TNF* (138, 139). Moreover, polymorphisms of *TNF* have been associated with severe malaria. A study in Gambia has found increased relative risk of having cerebral malaria in children homozygous for *TNF*-308A allele (140). Moreover, those who carried this allele were found to have increased risk for symptomatic reinfections with *P. falciparum* in Gabon (141), severe infectious diseases in general in Sri Lanka (142), and mortality and malaria morbidity in Kenyan infant (143). *TNF*-376A, a transcription factor OCT-1 binding genotype was associated with 4-fold increased susceptibility to cerebral malaria in large studies comparing cases and controls in West African and East African populations (144), whereas the *TNF*-238A allele has been associated with susceptibility to severe malarial anemia (9). Several different *TNF*-promotor SNPs were also suggested to be involved in the regulation of parasite density in a study in Burkina Faso (145).

FCGR2A encodes low-affinity IIa receptor for the Fc fragment of IgG which is present on monocytes, macrophages, neutrophils, natural killer (NK) cells, and T and B lymphocytes. The Fc fragment receptor plays an essential role in the protection against foreign antigens by removing antigen-antibody complexes from the circulation. A Arg131 allele of *FCGR2A* causes lost of binding to IgG2 and has been associated with protection against high levels of *P. falciparum* parasitemia in Kenya (146). Homozygotes for the His131 genotype were susceptible to cerebral malaria in the studies in Thailand and the Gambia (147, 148). The summary of immune genes that were reported to have association with malaria phenotypes is given in Table 5.

Table 5. Immune genes reported to be associated with different malarial phenotypes.

Gene	Reported associations with malaria	References
<i>HLA-B</i> and <i>HLA-DR</i>	<i>HLA-B53</i> and <i>HLA-DRB1</i> associated with severe malaria in the Gambia	(10, 11, 127)
<i>IFNR1</i> , <i>IFNG</i> and <i>IFNGR1</i>	Association with severe malaria in the Gambia	(128-130)
<i>IL10</i>	Haplotypic association with severe malaria in Gambia	(131)
<i>IL12B</i>	Association with severe malaria in Tanzania	(132)
<i>IL4</i>	Association with antimalarial Ab levels in Fulani people of Burkina Faso	(133)
<i>NOS2A</i>	Various associations with severe malaria in Gabon, Tanzania, Kenya and the Gambia	(12, 13, 134-136)
<i>TNF</i>	Various associations with severe malaria and re-infection risk in the Gambia, Burkina Faso, Kenya, Gabon and Sri Lanka	(138, 139, 141-143, 145)
<i>FCGR2A</i>	Association with severe malaria in the Gambia	(146-148)

6.4 Malaria and family studies

Detailed longitudinal analysis which includes studying a set of twins or a village is needed for an accurate family study of malaria. In this kind of study, quantitative traits such as the level of parasitemia and the frequency of malaria fever episodes can be evaluated. From the study of Gambian twins, genetic factors were partly responsible for susceptibility to malaria fever episodes (8), and the linkage region was mapped to major histocompatibility complex (MHC) region on chromosome 6 (139). In a longitudinal family analysis in Sri Lanka, the differences in susceptibility to clinical malaria episodes among some individuals appeared to be due to a genetic basis (149). For the level of parasitemia, it was first suggested that there was an involvement of a major gene controlling blood infection levels (150) but later indicated that complex genetic factors were involved (151, 152) and there was linkage to the MHC and the 5q31-33 region (6, 7, 138).

In a twin study in Liberia, genetic effect controlling individual variation in immunological responses was estimated. An evidence of heritability in antimalarial antibody responses was found but this did not appear to be due to HLA class II genes (153). In Papua New Guinea, a familial segregation analysis of immunological responses to malarial antigens gave complex overall picture but there was evidence suggesting that Mendelian effects might contribute to some specific antigen responses (154-156). The summary genetic association with malaria in family studies is given in Table 6.

Table 6. Genetic association with malaria in family studies.

Study area / genetic approach	Results	References
Liberia / twin study	Antibody levels after immunization through repeated infections have been compared to Pfl55/RESA in monozygotic twins with those in dizygotic twins, siblings, or unrelated controls. Antibody responses were found to be more concordant within monozygotic twin pairs than others living under similar environmental conditions. When the responses were assessed for possible associations with different HLA class II polymorphisms, no associations were found.	(153)
Cameroon/ longitudinal family studies	Segregation analysis of blood parasite densities in 42 Cameroonian families was performed. The results provided clear evidence for the presence of a recessive major gene controlling the degree of infection in human malaria, indicating that about 21% of the population was predisposed to high levels of infection.	(150)
Gambia / Twin study	A longitudinal survey of malaria morbidity in a cohort of 258 pairs of twin children was conducted in the Gambia to assess the extent to which genetic factors determine the host's susceptibility and clinical response to infection. The results showed that both members of an MZ pair were more likely to develop fever than were twins of a DZ pair, suggesting that genetic factors influenced the presentation of clinical disease. Sharing of MHC alleles was not increased among the 13 pairs of dizygous twins who were discordant for malaria. These findings confirm the importance of genetic factors to the risk of uncomplicated malaria.	(8, 139)
Cameroon / Sib-pairs linkage analysis	A segregation analysis of blood infection levels was performed here in 44 pedigrees and the results confirmed the existence of complex genetic factors controlling blood infection levels but were not consistent with the parent-offspring transmission of a single gene. No linkage was found with any of the four markers outside the 5q31-q33 region. A trend in favor of linkage was observed in the distal part of the 5q31-q33 region, especially with the marker D5S636 which was the marker that provided the highest evidence for linkage with SM1.	(6)

Table 6. Genetic association with malaria in family studies (cont.).

Study area / genetic approach	Results	References
Burkina Faso / sib-pair linkage analysis	Linkage analyses were performed on 153 sibs from 34 families between blood infection levels and chromosome 5q31-q33, which contains numerous candidate genes encoding immunological molecules. The results showed linkage of parasitemia to D5S393 and D5S658. The heritability indicated that its variation accounted for approximately 45% of the variance of blood infection levels. Multipoint analyses confirmed linkage, with a peak close to D5S658 suggesting a central role of the locus in the control of parasitemia.	(7)
Burkina Faso / family- based study	34 families were analyzed to test for linkage between polymorphisms within the MHC region and mild malaria .Two-point analysis indicated linkage of mild malaria to TNF. Using multipoint analysis, they also found evidence for linkage of mild malaria to the MHC region, with a peak close to TNF. These results support genes within the MHC region being involved in mild malaria. In particular, the genetic variation within TNF may influence susceptibility to mild malaria.	(138)
Papua New Guinea / Familial segregation analysis	Studied in 214 adults but found no significant associations between HLA class I or II alleles of these RESA and MSA-2 antigens in this population although there was evidence of familial aggregation of antibody responses to MSA-2 (3D7).	(154-156)
Sri Lanka / Longitudinal family analysis	Found that human genetics, housing, and predisposing systematic effects (e.g., sex, age, occupation, history of infections, village) each explained approximately 15% of the variation in the frequency of malarial infection. Genetic control of <i>P. falciparum</i> infections appeared to modulate the frequency and intensity of infections, whereas genetic control of <i>P. vivax</i> infections appeared to confer absolute susceptibility or refractoriness but not intensity of disease.	(149)
Mali / Family-based approach	This study was of 2,811 inhabitants of Bamako, Mali, clustered in 407 nuclear families. The probands were 136 children with severe malaria and 271 healthy children. The data suggested strong familial aggregation of CM and SMA that the odds of having CM were 2.49 times greater and the odds of having SMA were 4.92 times greater than the odds for a child whose sibling had no such history.	(157)

CHAPTER 4

MATERIALS AND METHODS

1. Subjects and Phenotypes Preparation

1.1 Study site

Ratchaburi is one of the several provinces that outline Thailand from its neighboring nation, Myanmar. The Tanaosri subdistrict, Suan Phung district, Ratchaburi province was selected as the study site. Suan Phung is a small district located near the Thai-Myanmar border surrounded by the long Tanaosri ranges on the western side. The Tanaosri subdistrict is located at the southern part of Suan Phung, about 163 km west of Bangkok. It has a population of 5,368 and majority of the people are Karen.

Tanaosri subdistrict has the total area of 126,343.75 rai (202.153 km²). Most of the areas are mountainous areas. The majority of both male and female labor forces are working in agricultural field while some females work at home as housewives. On the Thai-Myanmar border, malaria transmission estimates are low and unstable (1). The mosquito vectors in this area are *A. minimus*, *A. maculatus* and *A. dirus*, where *A. minimus* is the most common (data unpublished). The incidence of malaria is moderately seasonal, and the peak of malaria transmission occurs during April to June. Most of the people have been infected by either *P. falciparum* or *P. vivax* at some points and there were a few of bed net usage during this study period.

The Rajanagarindra Tropical Diseases International Center (RTIC) or known as Suan Phung field station, situated near Tanaosri subdistrict, is managed by the Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University.

The RTIC provides the people free diagnostic services of malaria and free medications such as antipyretics, antimalarials and antihelminthics. The RTIC gets more than 10,000 visits per year coming from the villages in the Tanaosri subdistrict and other neighboring areas in Suan Phung district.

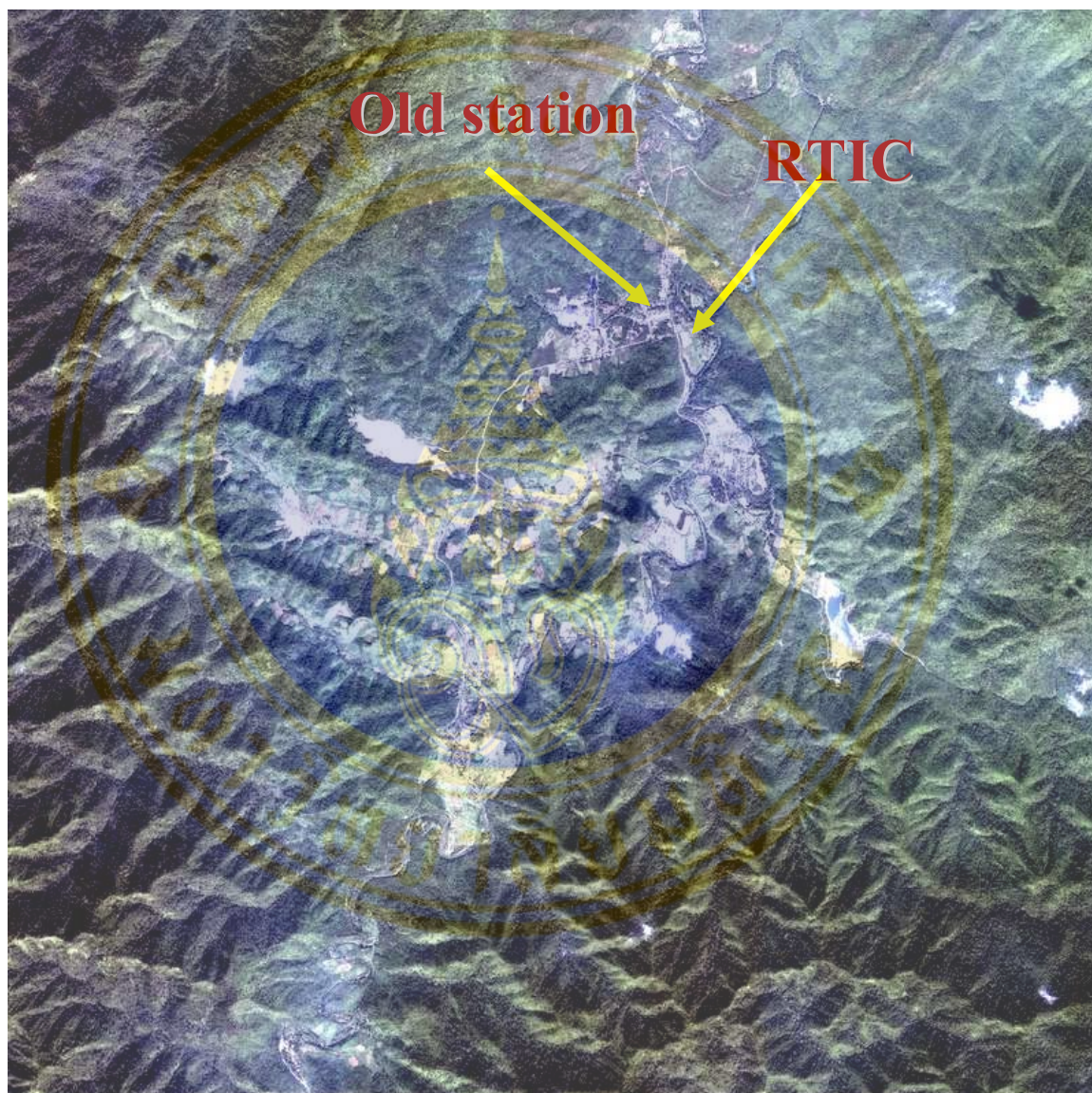


Figure 6. The Rajanagarindra Tropical Diseases International Center (RTIC).

1.2 Family data

The studied population comprised 3,481 villagers of all ages living in 7 hamlets in a total of 482 households in Tanaosri subdistrict which is an endemic area of malaria, and the peak of malaria transmission occurs during April to June. The project protocol and objectives were explained to the population and informed consent was individually obtained from all study participants or their parents or guardians. The ethical permission for the study was approved by the Ethical Committee of the Ministry of Public Health of Thailand.

To construct the family structures, 2,447 participants had been recruited using a questionnaire interviewing each individual or key representatives of the household to obtain both demographic information such as birth date, age, sex, day of fever, and genetic relationships between children, their parents, and sometimes their grandparents or non-relatives in the same household, and between households (2). The total pedigrees comprised of 2,428 individuals, including absent or dead relatives, forming 238 families that varied from 3 to 959 inhabitants per one family. The family trees had been constructed using Cyrillic pedigree drawing software. Multiple marriages are commons in every hamlet resulting in a large number of half siblings. DNA samples of 1,032 participants were extracted using Phenol-chloroform method from the blood of each participant who was five years old or older and from the buccal swab for the ones who were less than 5 years old. The family structures were confirmed by the ibscheck software (Centre Nationale de G notypage, France) using identity by state (IBS) allele sharing of each pair of pedigree members observed from the genotyping results of 400 microsatellite markers.

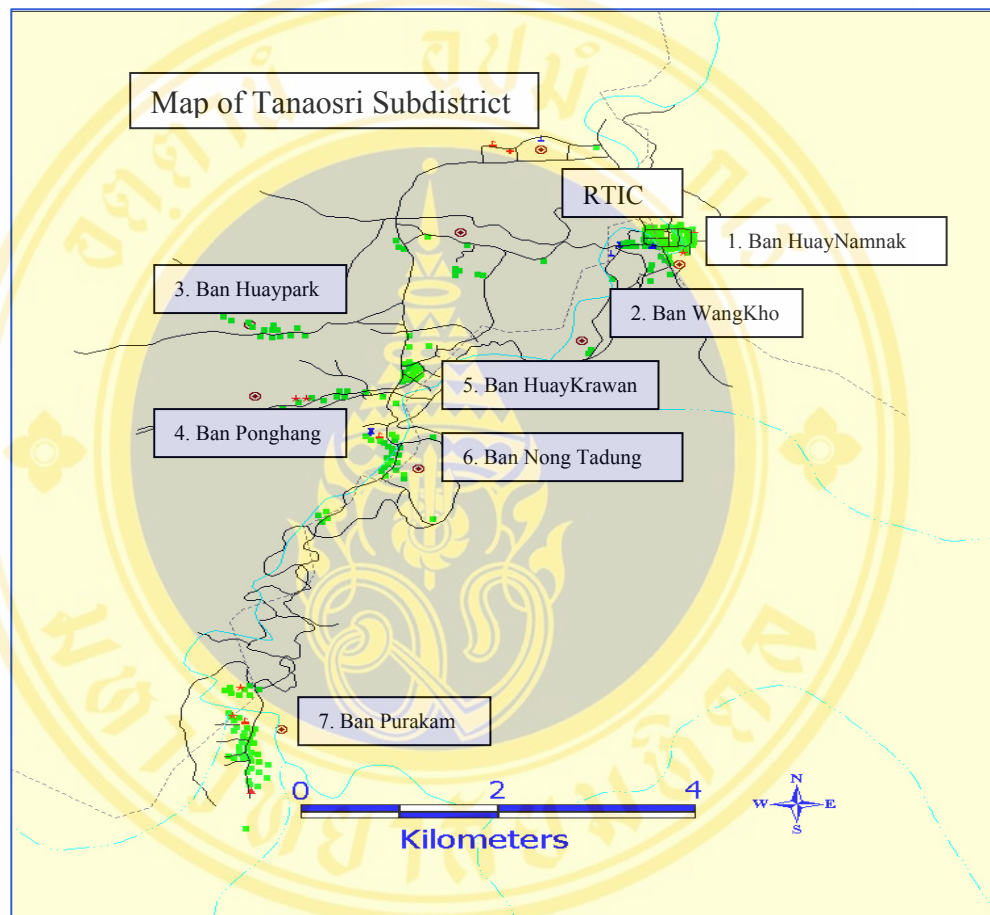


Figure 7. Map of Tanaosri subdistrict showing the location of 7 hamlets which were selected to be the study site.

1.3 Pedigree partitioning

There is a very large and complex family in Suan Peung consisting of 959 individuals. Pedigrees partitioning techniques were necessarily applied to reduce computational workload while running the linkage analysis programs. The aim was to partition the pedigree structures while minimizing information loss. The program developed by Brocklebank et al. (Brocklebank et al., in preparation) efficiently encodes and processes arbitrarily large and complex pedigree structures employing concepts from graph theory. Algorithms for pedigree management and pedigree partitioning while aiming to retain maximum information for analysis were implemented in their program.

The methods use a graphical representation for the pedigree structure and the problem of pedigree decomposition is formulated as a combinatorial optimization problem. The computation time for exhaustive enumeration of all possible pedigree configurations becomes rapidly prohibitive as pedigree size increases, so they implemented a stochastic method to search for an optimally informative partitioned pedigree structure. A Simulated Annealing (SA) algorithm is employed, which is a random combinatorial optimization procedure that iteratively updates a solution. The ability to conduct pedigree partitioning with fine control over the complexity of the generated structures, while maximizing the information retained for analysis is a major strength of this approach.

The constraint on pedigree structure is defined in terms of a maximum permitted complexity, where complexity is defined by “bits” which equals to $2n - f$, where n is the number of non-founders in a fully connected pedigree structure and f is the number of founders. This constraint allows direct usage of the subpedigrees identified in the popular programs GENEHUNTER, Allegro and Merlin. In addition, partitioning with larger constraints allows practical use of MCMC methods such as implemented in SIMWALK2 and Loki.

The algorithms evaluate pedigree structures by calculating information score for each structure based on expected kinship and fraternity coefficients for the complete pedigree, given by

$$f(X) = \text{sum}(\phi_{ij} + \psi_{ij}),$$

where ϕ_{ij} is the kinship coefficient between nodes i and j or the probability that a random allele from individual i is IBD with a random allele from j , ψ_{ij} is the fraternity coefficient or the probability relatives i and j share both alleles IBD, and the sum is over all pairs in which both nodes are informative. A node is considered informative or not informative based on available genotype and/or phenotype data. Simulations performed by Brocklebank *et al.* demonstrate that the score correlates well with the theoretical power for linkage.

1.4 Data collection

The clinical malaria cohort study had been conducted from June 1994 to May 2005. We monitored the incidence of malaria in all villagers. This community is made up of 4 closely related ethnic groups. Most of them are Karen (85%). The rest are Thai (14%), Mon and Burmese (1%).

1.4.1 Laboratory investigation

Thick and thin blood films were methanol-fixed, and then stained by 3% Giemsa's stain. Blood films were examined under oil immersion objective at x1000 magnification by trained laboratory technicians and 200 thick film fields were examined before films were declared negative. Parasite species were identified on thin films and densities (per μl .) were calculated from thick film by establishing the ratio of the parasites to white blood cells (WBCs) after at least 200-500 WBCs had been counted. The parasites / WBCs ratio was then multiplied by 8,000 which is the average WBC count per μl . of blood to get the parasite density. Asexual and sexual forms were recorded separately.

1.4.2 Passive case detection

We defined clinical malaria episodes as a measured fever (axillary temperature $> 37.5^\circ\text{C}$) or fever related symptoms including headache, back pain, chills, myalgia, nausea, vomiting, diarrhea, and cough associated with a slide positive for blood-stage, asexual and/or sexual form of *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale* parasites at any density. Passive case detection was effectively performed during June 1998 and May 2005.

1.4.3 Active case detection

In order to determine the prevalence of malaria, malaria survey had been carried out every 45 days from June 1994 to May 1995 and June 2003 to May 2004 in all hamlets, where 2,663 subjects were selected. Thick and thin blood smears were collected on the same slides from capillary blood of all individuals during the first week of each survey.

To calculate the phenotypic residuals, the survey period was assigned to person-years for each inhabitant. Every villager who had been checked for malaria in each year was supposed to stay in our study site for the whole year. To determine the correct number of clinical episodes, individuals presenting on consecutive occasions with blood-stage malaria parasite of the same species within 30 days after treatment of the initial infection were excluded from the analysis. Individuals presenting on consecutive occasions with non-malaria fever (or aforementioned symptoms) within 7 days following first presentation were likewise excluded from analyses. In addition, initial parasite negative visits were excluded if followed by a parasite-positive visit within the following two days. All positive malaria cases were treated with appropriate antimalarial drugs according to the recommendation of the Malaria Division, Ministry of Public Health and there were no self-treatment in our study area. Early treatment failure (ETF) and Late treatment failure (LTF) were defined as presentation with the same malaria species within 7 days and 7-30 days post-treatment respectively.

Incidence rates for *P. falciparum* and *P. vivax* malaria were calculated using these case definitions. Annual incidence rates (cases per 1,000 person-years) for clinical malaria caused by *P. falciparum* and *P. vivax* were derived from the following formula;

$$\text{Incidence rate} = \left(\frac{\text{No. of incidence cases}}{(\text{No. of individuals} \times 365) - (\text{No. of incidence cases} \times 28)} \right) \times 1000$$

P. vivax has a relapse mechanism that results in the reappearance of parasitemia. These parasites arise from the pre-erythrocytic hepatic-stage hypnozoite (3) which

develops at characteristic predetermined intervals following primary infection. It is well known that *P. vivax* exhibits two primary types of relapse pattern, apparently depending on the geographical origin of the parasites. The typical Chesson strains, referred to as the tropical-zone type, from the South Pacific and Southeast Asian regions (New Guinea-South Pacific), are characterized by early primary attack followed by a short latent period before the appearance of frequent relapses, whereas strains that originate in temperate zones (e.g., St. Elizabeth [United States]) exhibit early primary attack followed by a long latent period of 6 to 14 months succeeded by a series of relapses at short intervals (4, 5).

1.4.4 Malaria Control Program Treatment Guideline (WHO, 2003)

1. Treatment of uncomplicated *P. falciparum* cases

First line drug: Mefloquine 750 mg. and primaquine 30 mg. single dose

Second line drug: Quinine + tetracycline for 7 days or quinine alone for 7 days and primaquine 30 mg. on the last day.

Third line drug: Artesunate or artemether 700 mg. in divided dose over 5 days and primaquine 30 mg. on the last day.

2. Treatment of *P. vivax* and *P. ovale* cases

Chloroquine 1500 mg. base over 3 days

Primaquine 15 mg. daily for 14 days

3. Treatment of *P. malariae* cases

Same as *P. vivax* but no primaquine

4. Treatment of mixed infections

If mixed with *P. falciparum*, treat as *P. falciparum* malaria.

1.5 Disease phenotypes

From clinical case analyses (passive case detection) during 1998-2005, the following disease phenotypes were considered.

1.5.1 Clinical phenotypes

1. The number of clinical visits that were positive for *P. falciparum* (Number of visits Pf+) or *P. vivax* (Number of visits Pv+) or which were negative (Number of non-malaria fever visits).

2. The proportion of visits that were positive for *P. falciparum* (Proportion of visits Pf+) or *P. vivax* (Proportion of visits Pv+), that is the proportion of the number of visits positive for the parasite to all visits, and the proportion of positive visits with gametocyte for *P. falciparum* (Proportion of visits Pf gam+) or *P. vivax* (Proportion of visits Pv gam+), that is the proportion of the number of positive visits with gametocytes to all positive visits.

Using both number and proportion is an attempt to take into consideration differing tendency to visit the clinic

1.5.2 Parasite biological phenotypes

1. *P. falciparum* or *P. vivax* asexual parasite density, excluding zero parasite density values (i.e. exclude uninfected individuals) (Pf all parasite densities & Pv all parasite densities). Individual maximum asexual parasite densities (Pf max parasite density & Pv max parasite density). Calculated arithmetic mean asexual parasite densities, excluding zero values (Pf mean parasite density & Pv mean parasite density).

2. *P. falciparum* or *P. vivax* gametocyte parasite density (Pf all gametocyte densities & Pv all gametocyte densities). Individual maximum gametocyte parasite densities (Pf max gametocyte density & Pv max gametocyte density). Calculated arithmetic mean gametocyte parasite densities, excluding zero values (Pf mean gametocyte density & Pv mean gametocyte density).

Maximum density was chosen to assess whether there were individual human differences in tolerance to parasite burden prior to onset of clinical reaction. Mean density was chosen to consider the variation in the parasite genotype on human reaction and thus the human-parasite genotype interaction.

1.6 Data analyses for malaria and disease phenotypes

Statistical analyses and model fitting were conducted using the statistical package Genstat 7.1. All individuals in the study protocol were included in the analyses, irrespective of whether their family structure was known. Factors influencing the number of clinical visits and parasite density of either *P. falciparum* or *P. vivax* were analysed by fitting a GLM with a Poisson error structure (loglinear regression). While evaluating the proportion of visits that were positive for either species or proportion of infections positive for gametocytes, a binomial error structure (logistic regression) was performed. For phenotypes with repeated measures for the same individual, a GLMM was fitted with individual person as a factor in the random model. When single values per person were analysed a GLM was fitted. Factors include year (1998-2005), month, hamlet (7), and gender, with age factored into eight groups (<1, 1-4, 5-9, 10-14, 15-24, 25-39, 40-60 and 60+ years of age). Initially, a full model including all interactions possible was fitted and then the model was refined to include only statistically significant parameters and interactions. Date, a dimension, was only considered as a main effect and not as an interaction with the other physical variables (age, gender, hamlet). The seasonality of malaria transmission in this region varies annually and thus month was considered only via its interaction with year. Because the data were over-dispersed, a dispersion parameter was estimated. F-statistics in the GLM and Wald statistics, which approximate to a χ^2 distribution in the GLMM, were established. The residual variance not explained by these “environmental” factors was generated. Because a non-normal error distribution was used, Pearson rather than standardized normal residuals were generated. When multiple values were recorded for the same person, the mean residual value per person was then calculated thus generating one residual value per person per phenotype. Only phenotypes from individuals for whom family structure was available were then analysed for heritability. The effect of bednet use was analysed using the reduced data set (N=726), incorporating bednet use as an extra explanatory factor and analysed as above. Table 7 gives a summary of the disease phenotypes and their method of transformation.

Table 7. Disease phenotypes, and their methods of transformation

Disease phenotypes	Model type
1. Number of clinical visits	
Number of visits Pf+	GLMM loglinear regression (poisson error structure)
Number of visits Pv+	
Number of non-malaria fever visits	
2. Proportion of clinical visits	
Proportion of visits Pf+	GLMM logistic regression (binomial error structure)
Proportion of visits Pv+	
Proportion of visits Pf gam+	
Proportion of visits Pv gam+	
3. Asexual parasite density	
Pf max parasite density	loglinear regression
Pf mean parasite density	
Pv max parasite density	
Pv mean parasite density	
Pf all parasite densities	GLMM loglinear regression (poisson error structure)
Pv all parasite densities	
3. Gametocyte density	
Pf max gametocyte density	loglinear regression
Pf mean gametocyte density	
Pv max gametocyte density	
Pv mean gametocyte density	
Pf all gametocyte densities	GLMM loglinear regression (poisson error structure)
Pv all gametocyte densities	

2. Genotype Preparation

2.1 DNA extraction, quantification and amplification

We used Phenol-Chloroform as a standard method for DNA extraction from 3 ml. of EDTA whole blood of persons of 5 years old or older, from buccal swab of children younger than 5 years of age and from 2-3 tubes of heparinized blood in capillary tubes of those whom we can not take venous blood. After the extraction, the DNA was diluted by TE-4 (1X) buffer. The GenomiPhi™ DNA Amplification kit (Amersham) and Picogreen® Fluorimetry were used for DNA amplification and quantification respectively. The DNA was then kept at -20 °C.

2.1.1 Phenol-Chloroform DNA extraction

2.1.1.1 Reagents

Ammonium chloride (NH_4Cl)

Chloroform

Ethanol, absolute

HCl

Isoamyl alcohol

Isopropanol (2-Propanol)

Phenol

Potassium carbonate (KHCO_3)

Proteinase K

Sodium acetate

Sodium chloride

Sodium EDTA (Na_2EDTA)

Sodium dodecyl sulfate (SDS 10%)

Tris EDTA buffer, pH 8.

2.1.1.2 Preparation

Lysis Buffer

NH_4Cl 8.29 g [155 mM]

KHCO_3 1 g [10 mM]

Na_2EDTA 0.034 g or 200 μl EDTA 0.5 M [0.1mM]

Fill to 1000 ml with distilled water

Adjust to pH 7.4 with 1 M HCl or NaOH for each use

Chloroform/Isoamyl alcohol 24:1

Chloroform 24 ml

Isoamylalcohol 1 ml

SE-Buffer

NaCl 4.39 g [75 mM]

Na₂EDTA 8.41 g or 50 ml EDTA 0.5 M [25 mM]

Fill to 1000 ml with distilled water

Adjust to pH 8.0 with 1 M NaOH for each use

Sodium acetate

3 M Sodium acetate 246 g/L

Adjust to pH 5.2 with CH₃COOH

Proteinase K (10mg/ml)

Dissolve 100 mg Proteinase K in 10 ml TE for 30 min at room temperature

Aliquot and store at -20°C

Tris EDTA (TE-4) buffer (1X)

10 ml of 1M Tris-HCL

200 µl EDTA

Fill to 1000 ml with distilled water

2.1.1.3 Procedure

1. Add 10 ml lysis buffer into 3 ml. whole blood (or 2-3 capillary tubes of heparinized blood or buccal swab), shake gently, incubate for 30 min on ice, and centrifuge at 1200 r.p.m. for 10 min at 4°C.

2. Remove supernatant (blood waste), add 10 ml. lysis buffer, resuspend the pellet, and centrifuge for 10 min at 4°C (1200 r.p.m.).

3. Remove supernatant (blood waste), add 5 ml. SE-buffer, resuspend the pellet, and centrifuge for 10 min at 4°C (1200 r.p.m.).

4. Remove supernatant (blood waste). Add 5 ml. SE-buffer and resuspend the pellet, add 40 µl. proteinase K (10mg/ml) and 250 µl. 20% SDS shake gently, and incubate overnight at 37°C in a water bath.

5. Add 5 ml. SE-buffer and 10 ml. phenol shake by hand for 10 min, and centrifuge at 3000 r.p.m. for 5 min at 10°C.

6. Transfer the supernatant into a new tube, add 10 ml. phenol/chloroform/isoamyl alcohol (25:24:1), shake by hand for 10 min, and centrifuge at 3000 r.p.m. for 5 min at 10°C.

7. Again, transfer the supernatant into a new tube, add 10 ml. chloroform / isoamylalcohol (24:1), shake by hand for 10 min, and centrifuge at 3000 r.p.m. for 5 min at 10°C.

8. Transfer the supernatant into a new tube, add 300 µl. of 3 M sodium acetate (pH 5.2) and 10 ml isopropanol, shake gently until the DNA precipitate, use a plastic pipette, and capture the DNA.

9. Wash the DNA in 70% ethanol and dissolve the DNA in 0.5-1 ml TE-4 buffer overnight.

10. Measure the DNA concentration.

2.1.2 PicoGreen® dsDNA quantitation kit

PicoGreen® dsDNA quantitation reagent is an ultrasensitive fluorescent nucleic acid stain used in quantitating double-stranded DNA (dsDNA) in solution which enables quantitation of dsDNA as little as 25 pg/ml. with a standard spectrofluorometer and fluorescein excitation and emission wavelengths.

2.1.2.1 Reagents

PicoGreen® ds DNA quantitation reagent ref. P-7581 (conc. 1:200)

Tampon TE-3 (1X) (1 mM Tris-HCL pH 7.5, 0.1 mM EDTA)

Standard DNA (Ceph DNA 200 ng/µl.)

2.1.2.2 Procedure

1. Prepare an aqueous working solution of the PicoGreen reagent by making a 200-fold dilution of the stocking solution in TE-3. Protect the working solution from light by covering it with foil or placing it in the dark since the PicoGreen reagent is susceptible to photodegradation. For the best results, this solution should be used within a few hours of its preparation.

2. Prepare the 50 ng/50 μ l. and 100 ng/50 μ l. of standard DNA for standard curve by diluted 200 ng/ μ l. of Ceph DNA with TE-3.

3. Add 50 μ l. TE-3, 50 μ l. of 50 ng standard DNA and 100 ng standards DNA in each well of 96 well NUNC plate as blank and standard DNA for standard curve.

4. Add 1 μ l. of sample DNA with 49 μ l. TE-3 in each well of the rest of NUNC plate. Then add 50 μ l. PicoGreen reagent to all wells.

5. Cover the plate with aluminum foil, mix well on shaker and incubate for 5 minutes at room temperature.

6. After incubation, measure the sample fluorescence using a MFX Dynex fluorometer and standard fluorescein wavelengths (excitation ~ 485 nm, emission ~ 530 nm) in Ascent software.

7. Subtract the fluorescence value of the reagent blank from that of each of the samples. Use corrected data to generate a standard curve of fluorescence versus standard DNA concentration and determine the sample DNA concentration from the DNA Standard Curve.

8. After DNA quantitation, make 50 ng/ μ l. of DNA by diluting with 1X TE-4 and keep in Ependorf tube as a stocking DNA.

2.1.3 GenomiPhi™ DNA Amplification Kit (Amersham)

Each reaction of GenomiPhi™ DNA Amplification Kit must generate 4-6.5 μ g of DNA from 10 ng of purified human genomic DNA in 18 hours.

2.1.3.1 Reagents

QIAquick 96 PCR Purification Kit ref. 28181

QIAquick 96 PCR Purification Kit ref. 28183

Buffer PM

Buffer PE

Buffer EB

Absolute Ethanol

2.1.3.2 Kit contents

414896 GenomiPhi Sample Buffer (1x 0.9 ml), Lot 18294

414897 GenomiPhi Reaction buffer (1x 0.9 ml), Lot 18436

414898 GenomiPhi Enzyme Mix (1x 100 μ l.), Lot 18437

414895 GenomiPhi Control DNA, 10 μ g/ml (1x 20 μ l.), Lot 18323

2.1.3.3 DNA Amplification

- A. Heat denaturation of template DNA in sample buffer
 - Mix 5 μ l. of template DNA (conc. 50 ng/ μ l.) with 9 μ l. of sample buffer in Abgene plate and cover with ABI7000 adhesive sticker.
 - Heat to 95 °C for 3 min and then cool to 4 °C on ice.
- B. Preparation of amplification reaction and incubation
 - For each amplification reaction, prepare master mix by combining 9 μ l. of Reaction buffer with 1 μ l. of enzyme and mix on ice.
 - Add this master mix to the cool sample and mix by pipette.
 - Cover with the same adhesive sticker and incubate the sample at 30 °C for 16-18 hr.
- C. Post-amplification heat activation
 - Heat the sample to 65 °C for 10 min.
 - Cool to 4 °C

2.1.3.4 PCR and Electrophoresis for amplified DNA checking

- A. PCR reaction
 - master mix preparation
 - 10X Buffer 1.5 μ l.
 - dNTP 0.23 μ l.
 - MgCl 50 mM (1mM) 0.3 μ l.

Primer RST.119527 Forward 0.25 μ l.

Primer RST.119527 Reverse 0.25 μ l.

Taq polymerase 0.04 μ l.

H₂O 7.43 μ l.

- Add 1 μ l. of amplified product with 4 μ l. of H₂O and 10 μ l. of master mix in 96- well PCR plate, then use pipette for mixing.
- Add 50 μ l. of mineral oil in each well and put the plate on PCR machine, using RV65 program that the PCR condition is 65°C for 35 cycles.

B. Agarose gel

- 500 ml of 4% Agarose gel preparation
GibcoBRL 15 g
Nusieve 5 g
1X TBE 500 ml
EtBr 5 drops (1 drop/100 ml)
- Dissolve 15 g of GibcoBRL and 5 g of Nusieve with 500 ml TBE buffer by using stir rod and put the flask on warm plate.
- Boil the gel in microwave until the gel is clear, cool down until the gel gets warm then drop EtBr and mix gently.
- Pour warm gel into gel plate and put the comb on the gel then remove the comb since the gel is setting.

C. Gel Electrophoresis

- Put 3-5 μ l. of diluted PCR product to Falcon plate; add 5 μ l. loading dye and mix by pipetting.
- Load the mixture into each well of agarose gel
- Load 5 μ l. of Øx174 as a size marker (211 bp)
- Run electrophoresis at 300 V for 30-45 min. The PCR product bands can be seen under UV lighting.

2.1.3.5 Amplified DNA Purification

A. DNA sample

- After amplification, add 180 μ l. of distilled water to the 20 μ l. reaction mix.
- Use 3-5 μ l. diluted reaction mix for PCR to check that GenomiPhi amplification worked.

B. Prepare QIAvac 96 and Qiaquick 96-well plate

- Place waste tray inside QIAvac base
- Place QIAvac top plate over the base
- Seal unused wells of QIAquick by using thermowell Alu adhesive

C. Preparation and dispensation of DNA samples

- Add 3 volumes of buffer PM to 1 volume of the sample (200 μ l. + 600 μ l. buffer PM)
- After mixing, apply the samples to the wells of the QIAquick plate by using extended Biohit tips
- Switch on vacuum source.
- After all liquid has passed through the membrane, switch off the vacuum source.
- Repeat these steps if the total volume of the sample is more than 200 μ l. and the membrane must not be dried.
- Empty the waste tray in the specific trash for buffer PM. Clean the waste tray with water and dry because the component of QIAvac 96 are not resistant to concentrated alcohols so it is important to clean them after each step.

D. Washing

- Prepare buffer PE, buffer PE is supplied as a concentrate. Before using, add ethanol as indicated on the bottle and note on the bottle that the ethanol have been added.
- Wash wells of QIAquick plate by adding 900 μ l. of prepared buffer PE to each well and switch on vacuum source. Repeat this step twice.

- After buffer PE in all wells has passed through the membrane, apply vacuum for an additional 10 min to dry the membrane.
- Lift the top plate from the based (not the QIAquick 96 plate from the top plate), tap the top plate on absorbent papers until no drop comes out.
- Check that residual buffer PE is removed from the membrane.
- Empty the waste tray in the specific trash then clean with water and dry.

E. Elution

- For elution into provided collection microtubes: replace waste tray with the provided blue collection microtube rack containing 1.2 ml collection microtubes.
- To elute, add 100 μ l. of buffer EB (10 mM Tris-Cl, pH 8.5) to the center of each well of the QIAquick 96 plate (check that liquid covers all the membrane), incubate for 1 min and switch on vacuum source for at least 5 min or wait until all liquid has passed through the membrane. Switch off the vacuum source. Average eluted volume is 80 μ l. from 100 μ l. elution buffer volume.
- Rinse all components of the QIAvac 96 with water and dry.
- Dilute purified DNA with TE-4 buffer to 1 ml of total volume and separate into 2 parts, keep 500 μ l. at -20°C for stocking DNA and the other 500 μ l. for working DNA.

2.2 Microsatellite genotyping

The extracted DNA was randomly amplified using a primer extension pre-amplification method (6). Genotyping of microsatellites was carried out using an ABI366 system (Applied Biosystems, Foster City, USA). We typed 400 markers selected by Applied Biosystems with average interval of 10 cM.

3. Statistical Analysis

Genotyping results were checked for excess homozygotes or excess heterozygotes assuming Hardy-Weinberg equilibrium. Markers which showed excessive homo- or heterozygosity were re-examined. Mendelian inheritance within families was confirmed using PedCheck (7) and inconsistencies were resolved by re-examination of the raw data and re-genotyping where necessary. After Mendelian inheritance was resolved, unlikely genotypes were detected by MERLIN (8). Suggested microsatellite genotyping errors were corrected by re-examination of the raw data and re-genotyping where necessary.

3.1 Estimation of IBD matrices

The program Loki version 2.4.5 (<http://loki.homeunix.net/>) was used to estimate pairwise realized kinship coefficients which are IBD coefficients for all pairs of individuals in the pedigree at particular genomic positions. In fact, the values estimated are twice the kinship coefficients, as this is normally what is required. By using Markov Chain Monte Carlo method conditional on the linked markers, the program produces kinship coefficient matrices for the variance component analysis by SOLAR. This method can handle very large numbers of highly polymorphic markers as well as pedigree with high complexity. Loki has been used to provide IBD matrices for the analyses using external programs (9, 10).

3.2 Estimation of heritability

The determination of contribution of genetic factors to the phenotypes were performed by SOLAR (Sequential Oligogenic Linkage Analysis Routines) software package version 2.1.4 (<http://www.sfbr.org/sfbr/public/software/solar/>). The running files for SOLAR were prepared using MEGA2 program version 4.0 beta R1 (11). SOLAR performs a variance-component analysis of family data that model the variance of the phenotypes among family members as a function of genetic (polygenic) effects, measured covariates, and environmental effects (12). Let the quantitative phenotype, y , be written as a linear function of the n QTLs that influence it:

$$y = \mu + \sum_{i=1}^n \gamma_i + e,$$

where μ is the grand mean, γ_i is the effect of the i th QTL, and e represents a random environmental deviation. Assume γ_i and e are uncorrelated random variables with expectation 0 so that the variance of y is $\sigma_y^2 = \sigma_q^2 + \sigma_e^2$, where σ_q^2 is the variance of the additive effects of the QTL. Under this model, we can test the null hypothesis that the additive genetic variance due to the QTL equals zero (equivalent to the hypothesis of no linkage) by comparing the likelihood of this model with that of an alternative model in which σ_q^2 is estimated:

$$\begin{aligned} H_0: & \quad \sigma_y^2 = \sigma_e^2 \\ H_1: & \quad \sigma_y^2 = \sigma_q^2 + \sigma_e^2 \end{aligned}$$

Twice the difference in $\log_e$ likelihood of the two models provides a test statistic that is asymptotically distributed as a 1/2:1/2 mixture of a χ_1^2 variable and a point mass of zero (13). When the null hypothesis is rejected, the relative contribution of genetic factors to phenotypic variation is then estimated by the heritability (h^2) defined by the ratio of genetic variance component to the residual phenotypic variance (after adjusting for covariates); σ_q^2/σ_y^2 (12). As several phenotypes showed residual kurtosis of more than 0.8, *tdist* option which creates an extra parameter in the model to describe the distribution of the phenotype had been applied in all analyses.

3.3 Linkage analysis

Phenotypes in which the h^2 were significantly estimated were then selected for a quantitative trait locus autosomal linkage analysis by variance component and Haseman-Elston (H-E) methods. Variance component method was performed by SOLAR software package (version 2.1.4) and Merlin (version 1.1.2). H-E method was performed by Merlin-regress (version 1.1.2). LOD scores from both variance component and H-E method are reported. A linkage signal of LOD score more than 3 was considered significant. Candidate genes were defined as genes that were involved in red blood cell structure/ metabolism or immunological responses.

3.3.1 The variance component approach of linkage analysis

By using data from pedigree structures of arbitrary complexity, it is possible to make inferences regarding the localization and effect sizes of QTLs. For the simple additive model in which n QTLs and an unknown number of residual polygenes influence a trait, the covariance matrix for a pedigree can be decomposed into components that are due to the QTLs, genetic (polygenic), and environmental effects;

$$\mathbf{\Omega} = \sum_{i=1}^n \mathbf{\Pi}_i \sigma_{ai}^2 + 2\mathbf{\Phi} \sigma_g^2 + \mathbf{I} \sigma_e^2,$$

where $\mathbf{\Omega}$ is the covariance matrix for a pedigree, σ_{ai}^2 is the variance of the additive effects of the QTL, σ_g^2 is the residual genetic variance, σ_e^2 is the environmental variance, $\mathbf{\Pi}_i$ is the matrix whose elements (π_{ijl}) provide the predicted proportion of genes that individuals j and l share IBD at a QTL that is linked to a genetic marker locus, $\mathbf{\Phi}$ is the kinship matrix, and $\mathbf{I}$ is an identity matrix.

Using the variance-component model, we can test the null hypothesis that the additive genetic variance due to the i th QTL equals zero (no linkage) by comparing the likelihood of this restricted model with that of a model in which the variance due to the i th QTL is estimated. The difference between the two $\log_{10}$ likelihoods produces a LOD score that is the equivalent of the classical LOD score of linkage analysis. Twice the difference in loge likelihoods of these two models yields a test statistic that is asymptotically distributed as a 1/2:1/2 mixture of a χ_1^2 variable and a point mass at zero (13). When multiple QTLs are jointly considered, the resulting likelihood-ratio test statistic has a more complex asymptotic distribution that continues to be a mixture of χ^2 distributions.

For SOLAR, a household or shared environment effect can be added by an additional variance component with a coefficient matrix ($\mathbf{H}$) whose elements are 1 if the relative pair in question shares the environmental exposure or 0 otherwise. This simple incorporation of a shared household component leads to the following model for the phenotypic covariance matrix

$$\Omega = \sum_{i=1}^n \Pi_i \sigma_{ai}^2 + 2\Phi \sigma_g^2 + H \sigma_h^2 + I \sigma_e^2 .$$

Multipoint variance component linkage analyses were performed at 1 to 10 cM intervals according to the procedures outlined in each program.

3.3.2 The Haseman-Elston (H-E) approach of linkage analysis

The original H-E method of quantitative-trait linkage analysis for sib-pair data is based on regression of squared trait difference on the estimated proportion of alleles shared IBD at a marker locus (14). This method is simple and robust (15) but is less powerful than variance component method (16). Its modification developed by Sham and Purcell (17) was shown to lead to a test that had similar power to variance component approach. This method concerns regression of multipoint IBD sharing on trait squared sums and squared differences weighted by the inverse of their variances among all pairs of relatives. It has been implemented with the fast pedigree likelihood calculations provided by Merlin (8). The likelihood calculations required for obtaining IBD sharing probabilities are based on the Lander-Green algorithm and can handle pedigrees of around 20 individuals for multipoint analysis.

CHAPTER 5

RESULTS

1. Malaria Distribution During 8 Years of Study

During June 1998 to May 2005, 3,484 individuals of all ages were recruited into the study cohort and 4,303 cases were investigated from 21,041 observations. The total pedigrees comprised of 2,428 individuals were constructed from participants of 238 families living in 7 hamlets, including absent or dead relatives. DNA samples were obtained from 1,032 individuals of whom their family data were recorded.

We also collected 1,268 blood samples from the population and hematological studies including hemoglobin typing were conducted on these samples. From hematological data, at least 86 individuals were compatible with beta-thalassemia trait, 34 with alpha-thalassemia (HbH disease and alpha-thal1 trait), 90 with hereditary persistent fetal hemoglobin and 16 with HbE trait. Qualitative assay of G6PD was conducted on blood samples of 668 individuals, of which 106 (75 males) were deficient, and 61 (18 males) were partially deficient. Morphological studies revealed the presence of unique red cell shape abnormalities segregated in some families.

1.1 Passive case detection

Fifty three percent of participants did not have any clinical attack while 47% experienced at least one clinical attack during 8-year period. After discarding the recorded attacks that seemed to be due to recrudescences or relapses, of those people who had at least an attack, 66% had 1 to 2 attacks, 19% had 3 to 4 attacks, 13% had 5 to 10 attacks and the remaining 2% had more than 10 attacks, giving an average number of attacks of 1.24 per person within 8 years. The observed incidence rate was 2.4 cases per 1000 person-year for *P. falciparum* and 1.1 cases per 1000 person-year for *P. vivax*.

During 8 years, the peak of all species infections occurred in year 2000 (950 cases) as did the peak of *P. falciparum* infection (621 cases), while it was 1999 for *P. vivax* (319 cases). However, malaria cases declined after 2000 and the proportion of *P. falciparum* infection was, on average, twice as great as that of *P. vivax* in all years (Figure 8).

Of all the cases, malaria occurrence in males was almost twice as high as in females (63% and 36% respectively). Fifty four percent of malaria patients were children under 15 years of age and 44% were persons over 15 years of age (Table 8).

There was an age preference pattern for the two species; the number of *P. vivax* infections decreased in a linear fashion after the peak at the age of 1-4 years, whereas *P. falciparum* attacks did not reach the peak until the age of 5-9, after which it declined (Table 8, Figure 9).

We found all 4 parasite species during our study period; the two majorities were 2,825 cases of *P. falciparum* and 1,388 cases of *P. vivax* infection (66% and 32% respectively). Forty eight cases of mixed infections, 31 cases of *P. malariae* and 11 cases of *P. ovalae* were detected.

1.2 Active case detection

Average malaria prevalence rate from 45 days survey (during June 2003 – May 2004) among 2,663 individuals was 0.05 and the incidence rate was 0.09 cases per 1000 person-year for *P. falciparum* and 0.01 cases per 1000 person-year for *P. vivax*.

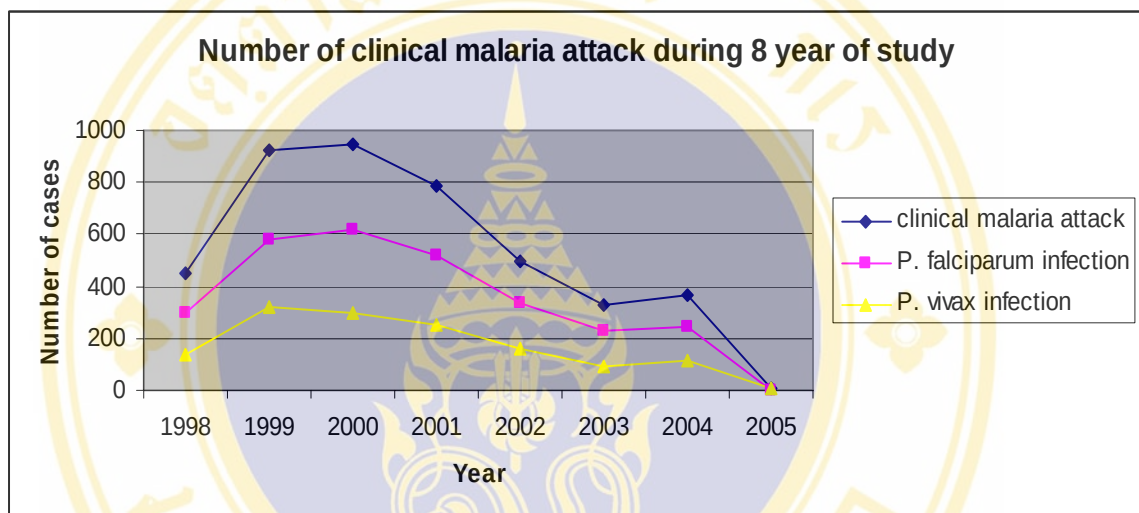


Figure 8. Number of clinical malaria attacks by year during 1998-2005

Table 8. Number of malarial infection among 9 age-groups

Age-group	Episode count	Clinical attack	Neg	Pf	Pv	Pm	Po	Mixed
1 (<1)	594	82	512	25	55	0	0	2
2 (1-4)	3074	684	2390	300	372	1	1	10
3 (5-9)	3378	915	2463	596	300	9	1	9
4 (10-14)	2428	633	1795	440	171	10	3	9
5 (15-24)	2591	512	2079	403	102	3	2	2
6 (25-39)	2830	401	2429	315	80	1	2	3
7 (40-59)	1629	244	1385	198	45	0	0	1
8 (>=60)	246	44	202	33	10	0	0	1
9(unknown)	4271	788	3483	515	253	7	2	11
All	21041	4303	16738	2825	1388	31	11	48

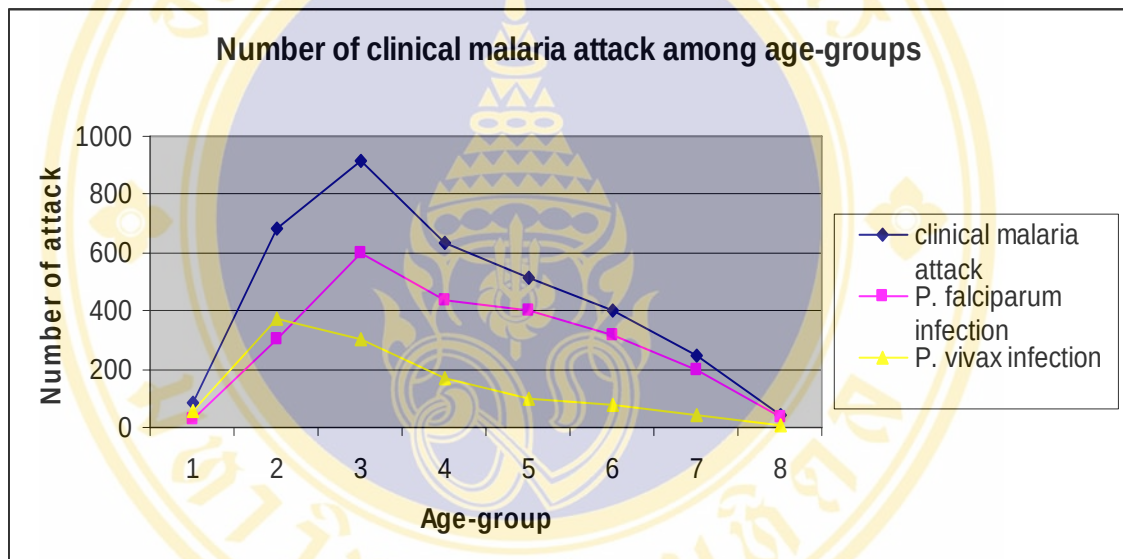


Figure 9. Number of clinical malaria (all species), *P. falciparum* and *P. vivax* attacks in each age-group

2. Pedigree Partitioning

Pedigrees partitioning was performed by the program developed by Brocklebank et al. (Brocklebank et al., in preparation) which efficiently encodes and processes arbitrarily large and complex pedigree structures while retaining maximum information for analysis.

For the original pedigrees, the total number of individuals was 2428, the information score was 2282.98 and the average complexity (bits) was 9.88 (0 – 1126). The pedigrees were then partitioned into maximum permitted complexity of 20 bits and 100 bits. The number of individuals for the pedigrees of maximum permitted complexity of 20 bits was 2485, the information score was 1846.31 (80.9% of original structure retained), and the average complexity was 6.39 (0 – 19). For the 100 bits one, the number of individuals was 2384, the information score was 2213.77 (96.9% of original structure retained), and the average complexity was 8.9 (0 – 99). The 20 bits one was used for linkage analyses by the program Merlin. For running Loki to produce kinship coefficient matrices and for linkage analyses by SOLAR which can take pedigree structure of higher complexity, the 100 bits permitted complexity pedigrees were used.

3. Data Analyses for Malaria and Disease Phenotypes

By using the statistical package Genstat 7.1, statistical analyses of the data and model fitting were conducted. Initially, a full model with all the factors including all interactions possible was fitted according to each phenotype. The factors include year (1998-2005), month (12), hamlet (7), house (477) gender, and age groups (8). Then the model was refined to include only statistically significant parameters and interactions. Table 9 gives the summary of samples used in the analyses.

For the number of clinical visits positive for *P. falciparum*, environmental factors explained the total variation by 12.0%. The significant factors were 2-way interactions (age.gender, age.hamlet, gender.hamlet) (<1.0%), age group (2.2%), gender (0.4%), hamlet (1.9%), and date (month and year, 6.5%). For the number of clinical visits positive for *P. vivax*, environmental factors explained 11.2% of the total variation, similar to those of *P. falciparum* positive visits. The significant factors were

2-way interactions (age.gender, age.hamlet, gender.hamlet) (1.9%), age group (4.8%), gender (0.2%), hamlet (0.5%), and date (3.8%). For the number of non-malaria clinical visits, 9.1% of the total variation could be explained by environmental factors which were 3-way and 2-way interactions (age.gender.hamlet, age.gender, age.hamlet, gender.hamlet) (<1.0%), age group (1.8%), gender (0.5%), hamlet (1.5%), and date (4.3%; Table 10, 11).

For the proportion of visits that were positive for *P. falciparum*, environmental factors explained 9.8% of the total variation with 3-way and 2-way interactions (age.gender.hamlet, age.gender, age.hamlet, gender.hamlet) (<1.0%), age group (1.8%), gender (0.5%), hamlet (1.5%), and date (5.0%) as significant factors. For the proportion of visits that were positive for *P. vivax*, environmental factors explained the total variation by 8.8%. The significant factors were 2-way interactions (age.gender, age.hamlet, gender.hamlet) (1.4%), age group (4%), gender (0.1%), hamlet (0.4%), and date (2.9%). The proportion of positive visits with gametocytes for either *P. falciparum* or *P. vivax* had just one significant environmental factor; age group for *P. falciparum* and date for *P. vivax*, which were responsible for 3.2% and 5.7% of the total variation respectively (Table 10, 11).

For *P. falciparum* maximum asexual parasite density, 34.4% of the total variation was explained by environmental factors with age group (18.9%), hamlet (1.1%), and date (14.4) as significant factors. The mean asexual *P. falciparum* parasitemia had only one significant factor; age group, which was responsible for 16.3% of the total variation. The *P. vivax* maximum asexual parasite density had 4 significant factors; age group (10.1%), hamlet (2.3%), date (13.1), and age.hamlet (8.5%), which were responsible for 34% of the total variation. For the *P. vivax* mean parasitemia, the environmental factors explained 18.7% of the total variation and the significant factors were age group (5.3%), hamlet (3.2%), and age.hamlet (10.2%). For *P. falciparum* all asexual parasite densities, 18.2% of the total variation could be explained by environmental factors which were age group (11%) and date (7.2%). The 3 significant factors for *P. vivax* all asexual parasite densities which contribute to 11.9% of the total variation were age group (3.3%), hamlet (2.5%), and date (6.1%). There was no significant factor for *P. falciparum* maximum and mean gametocytes densities, thus we used real raw data instead for these two phenotypes. For *P. vivax*

maximum gametocytes densities, only asexual parasite density was the significant factor explaining 4.4% of the total variation but for its mean gametocytes densities, there were 3 significant factors; hamlet, age.gender, and age.gender.hamlet, which were responsible for 3.9%, 8.8%, and 6.2% of the total variation respectively. For *P. falciparum* all gametocyte densities, the significant factors were age group and asexual parasite density which were responsible for 12.6% and 6.2% of the total variation respectively. For *P. vivax* all gametocyte densities, 8.4% of the total variation could be explained by age group (5.6%), hamlet (1.6%), and asexual parasite density (1.2%; Table 10, 11).

After the model for each phenotype was obtained, the residual variance not explained by these “environmental” factors was generated. Because a non-normal error distribution was used, Pearson rather than standardized normal residuals were generated.

Table 9. Summary of samples used in statistical and genetic analyses. Shown on the left hand side are the number of data points (observations and calculated mean/max values) analysed for each phenotype, the corresponding number of individuals implicated and hence residual values generated. On the right hand side are the number of these individuals for whom pedigree information was available and thus the number of pedigrees for each phenotype in the heritability analyses.

Phenotype	Statistical analyses		Genetic analyses	
	# Data points	# Individuals.	# Individuals.	# Pedigrees
Proportion of visits Pf+	19,162	2,545	2,018	334
Number of visits Pf+	19,162	2,545	2,018	334
Number of non-malaria fever visits	19,162	2,545	2,018	334
Proportion of visits Pv+	19,162	2,545	2,018	334
Number of visits Pv+	19,162	2,545	2,018	334
Pf mean parasite density	949	949	859	240
Pf max parasite density	949	949	859	240
Pf all parasite densities	1,796	949	859	240
Proportion of visits Pf gam+	1,796	949	859	240
Raw Pf mean gametocyte densities	84	80	73	47
Raw Pf maximum gametocyte densities	84	80	73	47
Pf all gametocyte densities	1,796	949	859	240
Pv mean parasite density	517	517	470	169
Pv max parasite density	517	517	470	169
Pv all parasite densities	978	517	470	169
Proportion of visits Pv gam+	978	517	470	169
Pv mean gametocyte densities	323	230	206	102
Pv max gametocyte densities	323	230	206	102
Pv all gametocyte densities	978	517	470	169

Table 10. Significance of environmental variables in the statistical analyses performed for each phenotype.

Phenotype	Variable								
	Age group	Gender	Hamlet	Date	Age. Gender	Age. Hamlet	Gender. Hamlet	Age. Gender. Hamlet	Asexual parasite density
Proportion of visits Pf+	<0.001	<0.001	<0.001	<0.001	0.012	<0.001	<0.001	0.041	ND
Number of visits Pf+	<0.001	<0.001	<0.001	<0.001	0.009	0.001	0.002	ns	ND
Number of non-malaria fever visits	<0.001	<0.001	<0.001	<0.001	0.004	<0.001	0.19 ^a	0.001	ND
Proportion of visits Pv+	<0.001	<0.001	<0.001	<0.001	0.003	<0.001	0.007	ns	ND
Number of visits Pv+	<0.001	<0.001	<0.001	<0.001	0.003	<0.001	0.005	ns	ND
Pf mean parasite density	<0.001	ns	ns	ND	ns	ns	ns	ns	ND
Pf max parasite density	<0.001	ns	0.035	<0.001	ns	ns	ns	ns	ND
Pf all parasite densities	<0.001	ns	ns	<0.001	ns	ns	ns	ns	ND
Proportion of visits Pf gam+	0.015	ns	ns	ns	ns	ns	ns	ns	ND
Raw Pf mean gametocyte density	ns	ns	ns	ND	ns	ns	ns	ns	ND
Raw Pf max gametocyte density	ns	ns	ns	ns	ns	ns	ns	ns	ns
Pf all gametocyte densities	0.002	ns	ns	ns	ns	ns	ns	ns	0.005
Pv mean parasite density	<0.001	ns	<0.001	ND	ns	<0.001	ns	ns	ND
Pv max parasite density	<0.001	ns	0.002	0.013	ns	0.002	ns	ns	ND
Pv all parasite densities	<0.001	ns	0.03	0.008	ns	ns	ns	ns	ND
Proportion of visits Pv gam+	ns	ns	ns	0.003	ns	ns	ns	ns	ND
Pv mean gametocyte density	0.06 ^a	0.46 ^a	0.005	ND	<0.001	0.062 ^a	0.61 ^a	0.025	ND
Pv max gametocyte density	ns	ns	ns	ns	ns	ns	ns	ns	0.001
Pv all gametocyte densities	0.011	ns	0.03	ns	ns	ns	ns	ns	0.001
^a retained as 3-way interaction significant									
ns – not significant (P>0.05) ND – variable not included in analyses									

Table 11. Percentage of variation explained by environmental factors in the statistical analyses

Phenotype	Variable									
	Age group	Gender	Hamlet	Date	Age. Gender	Age. Hamlet	Gender. Hamlet	Age. Gender. Hamlet	Asexual parasite density	
Proportion of visits Pf+	1.8	0.5	1.5	5.0	2&3-way Interactions <1%					-
Number of visits Pf+	2.2	0.4	1.9	6.5	2-way Interactions <1%					0
Number of non-malaria fever visits	1.8	0.5	1.5	4.3	2&3-way Interactions <1%					-
Proportion of visits Pv+	4.0	0.1	0.4	2.9	2-way Interactions 1.4%					0
Number of visits Pv+	4.8	0.2	0.5	3.8	2-way Interactions 1.9%					0
Pf mean parasite density	16.3	0	0	-	0	0	0	0	0	-
Pf max parasite density	18.9	0	1.1	14.4	0	0	0	0	0	-
Pf all parasite densities	11.0	0	0	7.2	0	0	0	0	0	-
Proportion of visits Pf gam+	3.2	0	0	0	0	0	0	0	0	-
Raw Pf mean gametocyte density	0	0	0	-	0	0	0	0	0	-
Raw Pf max gametocyte density	0	0	0	0	0	0	0	0	0	0
Pf all gametocyte densities	12.6	0	0	0	0	0	0	0	0	6.2
Pv mean parasite density	5.3	0	3.2	-	0	10.2	0	0	0	-
Pv max parasite density	10.1	0	2.3	13.1	0	8.5	0	0	0	-
Pv all parasite densities	3.3	0	2.5	6.1	0	0	0	0	0	-
Proportion of visits Pv gam+	0	0	0	5.7	0	0	0	0	0	-
Pv mean gametocyte density	0	0	3.9	-	8.8	0	0	0	6.2	-
Pv max gametocyte density	0	0	0	0	0	0	0	0	0	4.4
Pv all gametocyte densities	5.6	0	1.6	0	0	0	0	0	0	1.2

4. Heritability to Clinical Malaria

The heritability (h^2) evaluated by variance component method implemented in the program SOLAR is the relative contribution of genetic factors or actually additively acting host genes to phenotypic variation. Heritability defined by the ratio of genetic variance component to the residual phenotypic variance (after adjusting for covariates) was estimated for standardized normal deviance residual and Pearson's residual of each phenotype.

For the number of *P. falciparum* clinical attacks, the evaluated h^2 was 18.3% (p -value = 3.47×10^{-8}), lower than that of the number of *P. vivax* clinical attacks which was 19.3% (p -value = 9.77×10^{-13}). For the number of non-malaria clinical visits, 18.5% (p -value = 6.26×10^{-9}) of the total variation could be explained by additively acting host genes (Table 12).

When considering proportion of positive visits rather than the number of visits, similar values were obtained as genetic factor explained 18.2% (p -value = 3.48×10^{-8}) of the total variation of proportion of positive visits for *P. falciparum*, and 18.7% (p -value = 7.35×10^{-12}) for *P. vivax*. For the proportion of both *P. falciparum* and *P. vivax* infections with gametocytes, h^2 could not be significantly evaluated (Table 12).

For parasite density, additively acting host genes explained 27.8% (p -value = 2.6×10^{-6}) of the total variation of maximum asexual *P. falciparum* parasitemia, whereas for the mean asexual *P. falciparum* parasitemia, it was 19.2% (p -value = 0.0016). For *P. vivax* parasitemia, the h^2 were relatively higher than those of *P. falciparum* where it was 31.9% (p -value = 0.0017) for maximum *P. vivax* parasitemia, and 42% (p -value = 0.00037) for mean *P. vivax* parasitemia. For all asexual densities, while the h^2 of *P. falciparum* was only 20.3% (p -value = 0.0015), the one of *P. vivax* went up to 34.9% (p -value = 0.0019). Neither the h^2 of the residuals of maximum and mean gametocytes nor those of all gametocyte densities of *P. falciparum* and *P. vivax* could be significantly determined (Table 12).

Figure 10 gives the total variation in some of the clinical and parasite density phenotypes partitioned into their genetic and environmental components. Of particular interest are (i) the similarity in heritability on clinical *P. falciparum* phenotypes and non-malaria clinical episodes, and (ii) the very large genetic contribution to *P. vivax* asexual parasite density phenotypes (31.9-42%).

Table 12. Estimation of heritabilities for number of clinical attacks, parasite density, proportion of clinical presentations, and non-malaria clinical presentations.

Disease phenotypes	h^2 (%)	Std. Error (%)	p -value
Clinical visits			
Proportion of visits Pf+	18.2	4	3.48×10^{-8}
Proportion of visits Pf gam+	5.6	5.4	NS
Number of visits Pf+	18.3	4	3.47×10^{-8}
Number of non-malaria fever visits	18.5	3.9	6.26×10^{-9}
Proportion of visits Pv+	18.7	3.5	7.35×10^{-12}
Proportion of visits Pv gam+	10.7	10.9	NS
Number of visits Pv+	19.3	3.5	9.77×10^{-13}
Parasite biological traits			
Pf mean parasite density	19.2	7.7	0.0016
Pf max parasite density	27.8	7.4	2.6×10^{-6}
Pf all parasite densities	20.3	8	0.0015
Raw Pf mean gametocyte density	26.5	58.2	NS
Raw Pf max gametocyte density	51.8	51.5	NS
Pf all gametocyte densities	0.066	4.7	NS
Pv mean parasite density	42	13.5	0.0004
Pv max parasite density	31.9	12	0.0017
Pv all parasite densities	34.9	13.1	0.0019
Pv mean gametocyte density	0		NS
Pv max gametocyte density	31.5	25	NS
Pv all gametocyte densities	13.6	11	NS

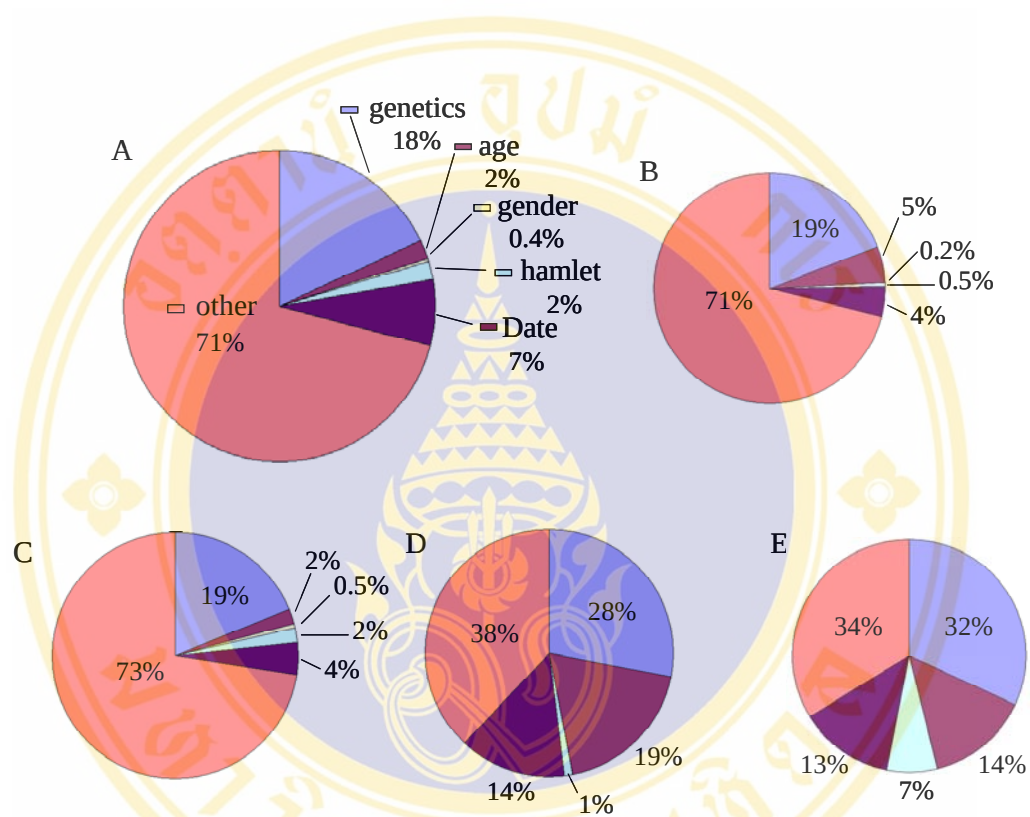


Figure 10. Proportions of variation explained by genetic heritability and environmental factors found to have a significant effect on the phenotype measured (Table 10, 11). (A) Number of clinical episodes of *P. falciparum*; (B) Number of clinical episodes of *P. vivax*; (C) Number of non-malaria clinical episodes; (D) Maximum *P. falciparum* parasite density; (E) Maximum *P. vivax* parasite density. Color coding for all five charts follows that in the top left.

5. Linkage Analysis

The variance component (VC) and Haseman-Elston (HE) multipoint linkage analyses were applied (Figure 11-14). The number of clinical visits, the proportion of clinical visits, asexual parasite density, and gametocyte density were analyzed as quantitative phenotypes. The most significant results were in *P. vivax* parasite density on chromosome 3q27.1-3q27.3. The linkage signal reached statistical significance in HE analyses with all *P. vivax* parasite density phenotypes. All signals were identified at 197 cM between markers D3S1565-D3S1262 (corresponding to 186-201 cM or 174.9-187.8 Mb of the genetic and physical chromosomal map). For Pv max parasite density, the maximum logarithm of odds (LOD) score was 4.57 (p -value $< 1 \times 10^{-5}$, locus-specific heritability of 42%; 95% confidence interval, 24.2%-59.8%), for Pv mean parasite density, the LOD score was 4.39 (p -value $< 1 \times 10^{-5}$, locus-specific heritability of 46.6%; 95% confidence interval, 26.2%-67%), and for Pv all parasite densities, the LOD score was 5.89 (p -value $< 1 \times 10^{-5}$, locus-specific heritability of 60%; 95% confidence interval, 37.5%-82.5%; Figure 13). No significant signal by VC from either Merlin or SOLAR was detected in this region. Figure 16 gives the linkage analyses results on chromosome 3 by Merlin regress with all 3 *P. vivax* parasite density phenotypes. The 1 LOD score support interval encompassed a 14 cM/ 6.4 Mb distance containing 62 annotated or hypothetical genes. Functional candidates include genes encoding dendritic cell-specific lysosome-associated membrane glycoprotein (LAMP3), kelch-like protein 6 (KLHL6), histidine-rich glycoprotein (HRG), and B-cell lymphoma 6 protein (BCL6).

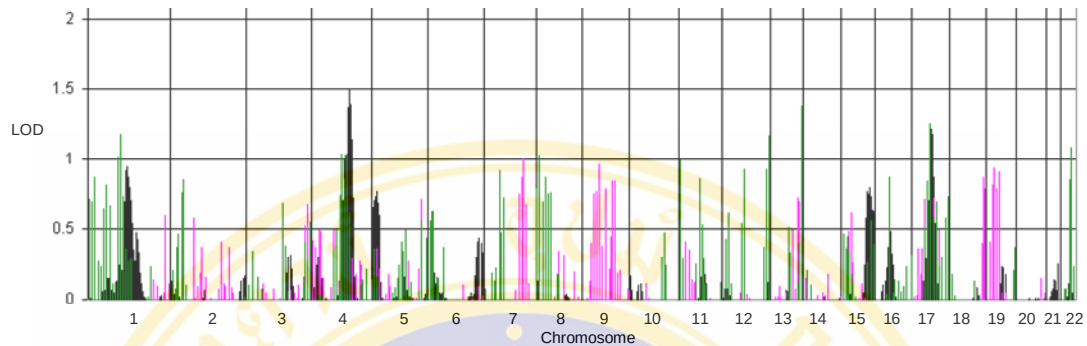
Furthermore, the signals on chromosome 1p36.23 and 1p36.22-1p36.21 provided evidence for linkage at regions between markers D1S214 - D1S450 (14 cM/6.98 Mb - 21 cM/9.61 Mb) and D1S2667-D1S2697 (25 cM/11.31 Mb - 37 cM/16.39 Mb) respectively with all 3 *P. vivax* parasite density phenotypes in HE analyses. In the first region, the signals were detected at 15 cM giving the LOD score of 2.75 (p -value = 0.0002, locus-specific heritability of 28.4%; 95% confidence interval, 12.7%-44.1%), 2.62 (p -value = 0.0003, locus-specific heritability of 31.9%; 95% confidence interval, 13.9%-49.9%), and 2.65 (p -value = 0.0002, locus-specific heritability of 30.3%; 95% confidence interval, 13.2%-47.4%) for maximum, mean and all *P. vivax* parasite density respectively. In the second region, the LOD score of

2.39 (p -value = 0.0004, locus-specific heritability of 21%; 95% confidence interval, 8.7%-33.3%), 2.35 (p -value = 0.0005, locus-specific heritability of 24.3%; 95% confidence interval, 9.8%-38.8%), and 2.53 (p -value = 0.0003, locus-specific heritability of 25.2%; 95% confidence interval, 10.7%-39.7%) for maximum, mean and all *P. vivax* parasite density respectively were detected at 28 cM. Almost no signal was detected in either of these regions in VC analyses (LOD score < 0.6; Figure 13). This finding on chromosome 1p36 is in the same region previously identified by Timmann et al. (1) where linkage signals were detected with *P. falciparum* parasite prevalence and malarial anemia.

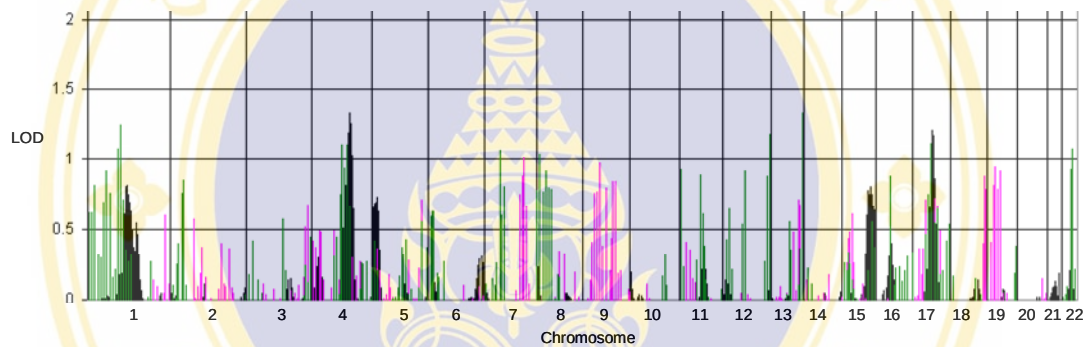
A further region with evidence for linkage was on chromosome 5p13.1 at marker D5S418 (59 cM/40.1 Mb) with all 3 *P. falciparum* parasite densities phenotypes in HE analyses, giving the LOD score of 2.25 (p -value = 0.0006, locus-specific heritability of 57.3%; 95% confidence interval, 22.4%-92.2%), 2.31 (p -value = 0.0006, locus-specific heritability of 48.2%; 95% confidence interval, 19.2%-77.2%) and 2.3 (p -value = 0.0006, locus-specific heritability of 49.4%; 95% confidence interval, 19.6%-79.2%) maximum, mean, and all *P. falciparum* parasite densities respectively (figure 12). No significant signal from VC methods was detected.

Finally, we found no evidence to confirm previously reported linkage of parasite density to chromosome 5q31-5q33 (2, 3), or malaria clinical episodes to chromosome 6p21.3 (4, 5) and 10p15.3-10p14 (1). Table 13 gives the summary of the chromosome positions, methods of analyses, and the phenotypes of which the LOD score is higher than 1.5.

A. Number of visits Pf+



B. Proportion of visits Pf+



C. Number of non-malaria fever visits

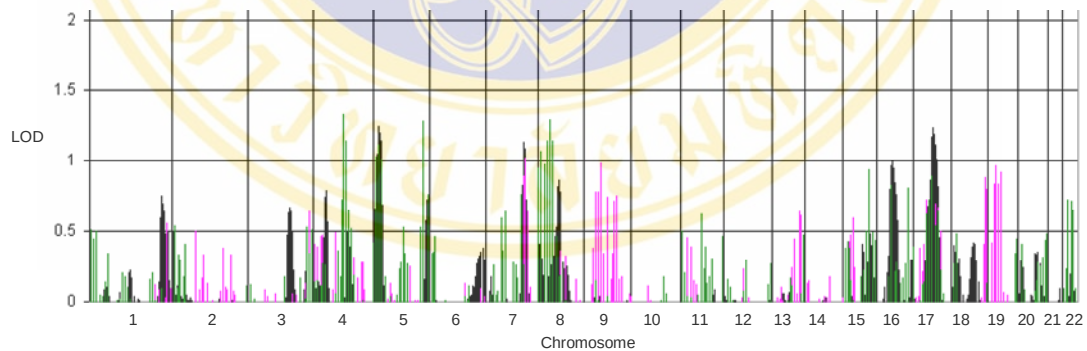
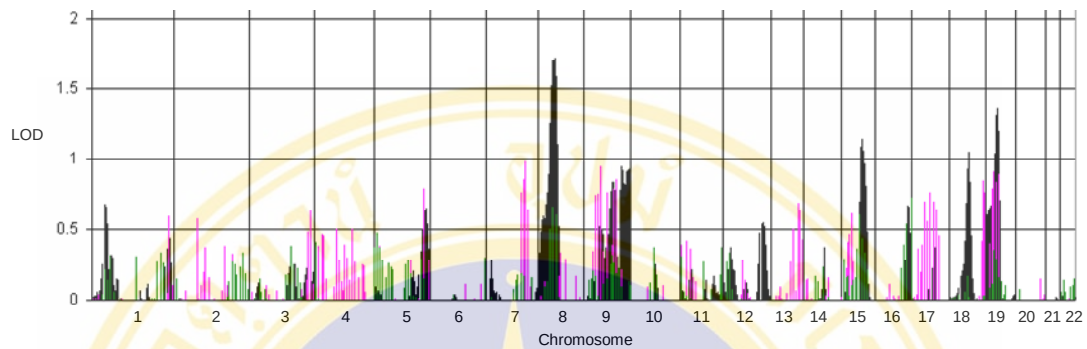


Figure 11. Linkage analysis results of number and proportion of clinical visits. (A) Number of visits Pf+. (B) Proportion of visits Pf+. (C) Number of non-malaria fever visits. (D) Number of visits Pv+. (E) Proportion of visits Pv+. Pink, green, and black columns represent the results from Merlin regression, Merlin variance component, and SOLAR variance component respectively.

D. Number of visits Pv+



E. Proportion of visits Pv+

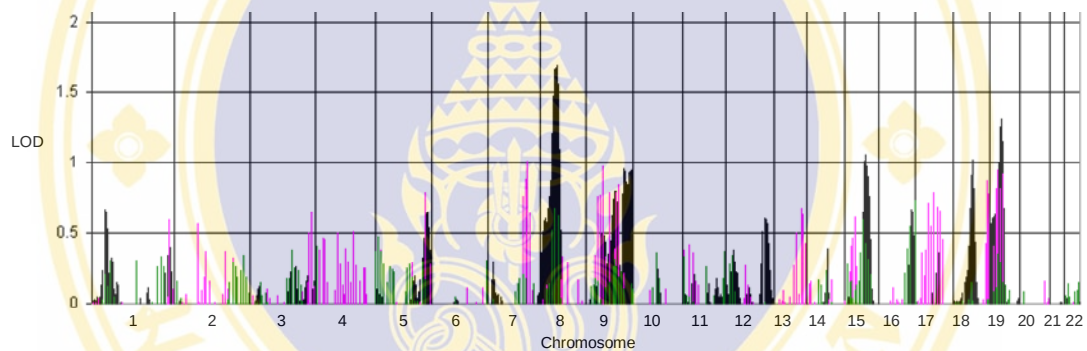
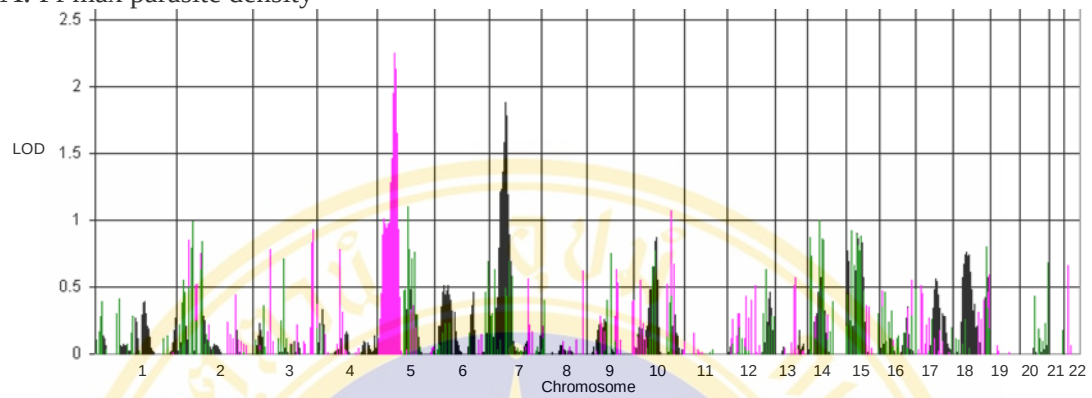


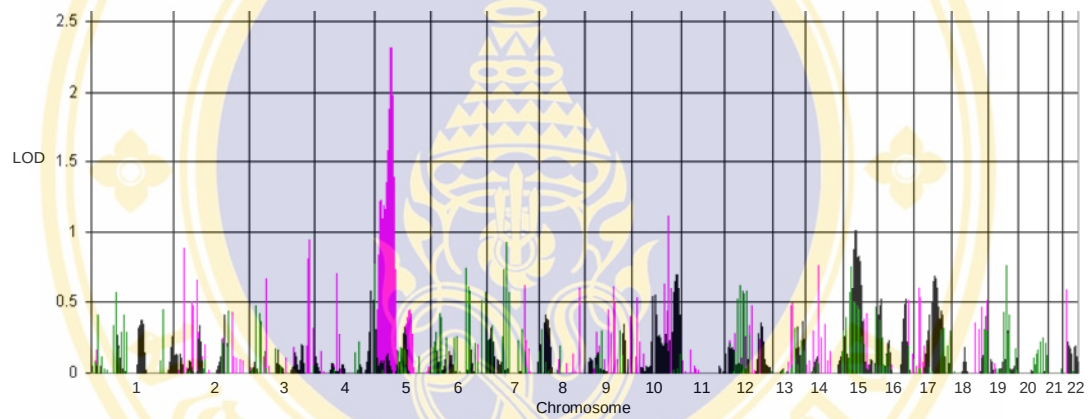
Figure 11. Linkage analysis results of number and proportion of clinical visits (cont.).

(A) Number of visits Pf+. (B) Proportion of visits Pf+. (C) Number of non-malaria fever visits. (D) Number of visits Pv+. (E) Proportion of visits Pv+. Pink, green, and black columns represent the results from Merlin regression, Merlin variance component, and SOLAR variance component respectively.

A. Pf max parasite density



B. Pf mean parasite density



C. Pf all parasite densities

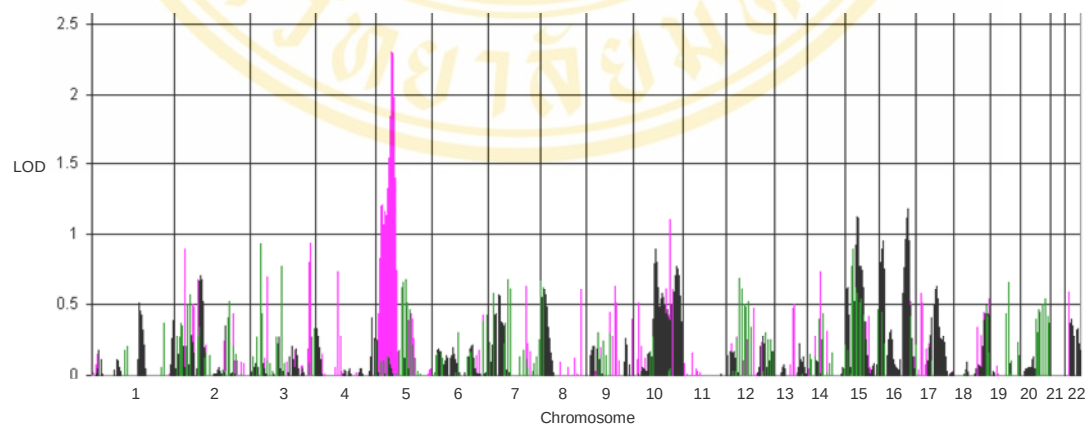
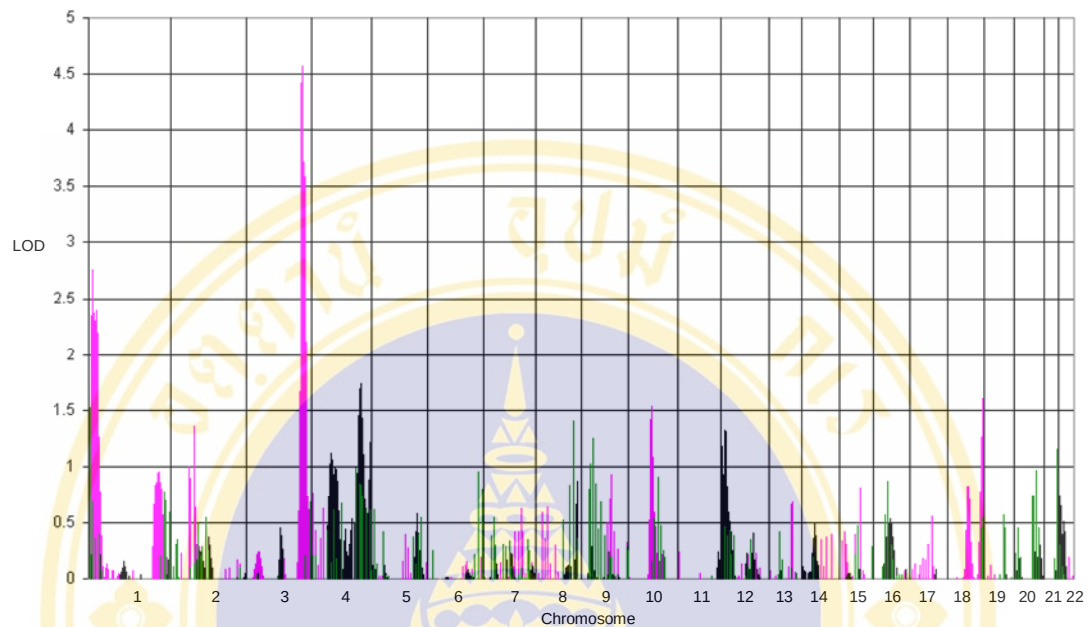


Figure 12. Linkage analysis results of maximum, mean, and all *P. falciparum* parasite density (A) Pf max parasite density. (B) Pf mean parasite density. (C) Pf all parasite densities. Pink, green, and black columns represent the results from Merlin regression, Merlin variance component, and SOLAR variance component respectively.

A. Pv max parasite density



B. Pv mean parasite density

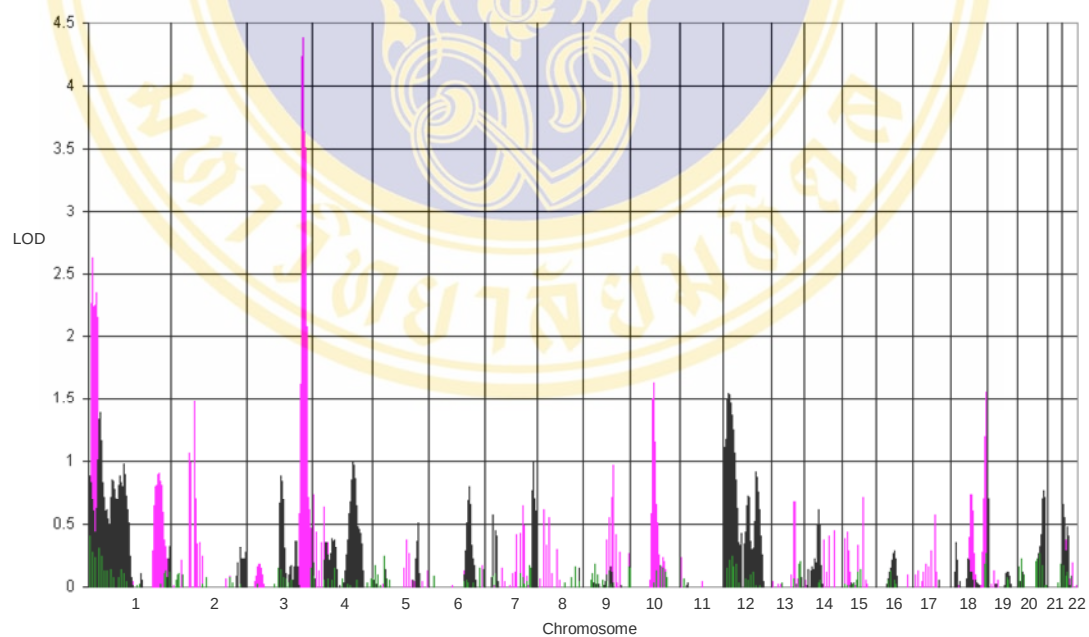


Figure 13. Linkage analysis results of maximum, mean, and all *P. vivax* parasite density (A) Pv max parasite density. (B) Pv mean parasite density. (C) Pv all parasite densities. Pink, green, and black columns represent the results from Merlin regression, Merlin variance component, and SOLAR variance component respectively.

C. Pv all parasite densities

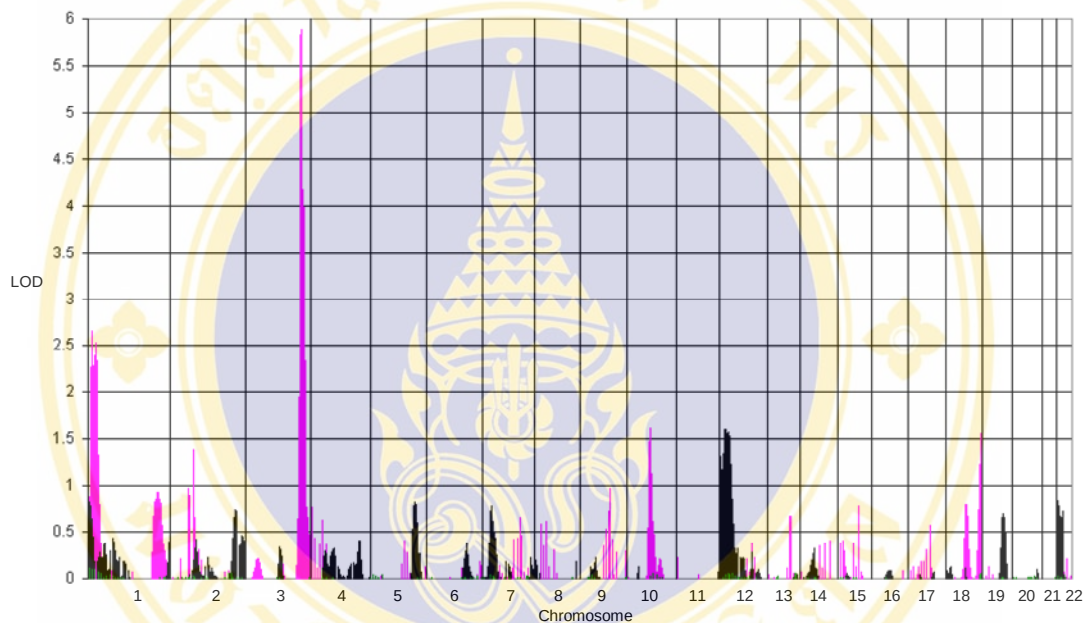


Figure 13. Linkage analysis results of maximum, mean, and all *P. vivax* parasite density (cont.). (A) Pv max parasite density. (B) Pv mean parasite density. (C) Pv all parasite densities. Pink, green, and black columns represent the results from Merlin regression, Merlin variance component, and SOLAR variance component respectively.

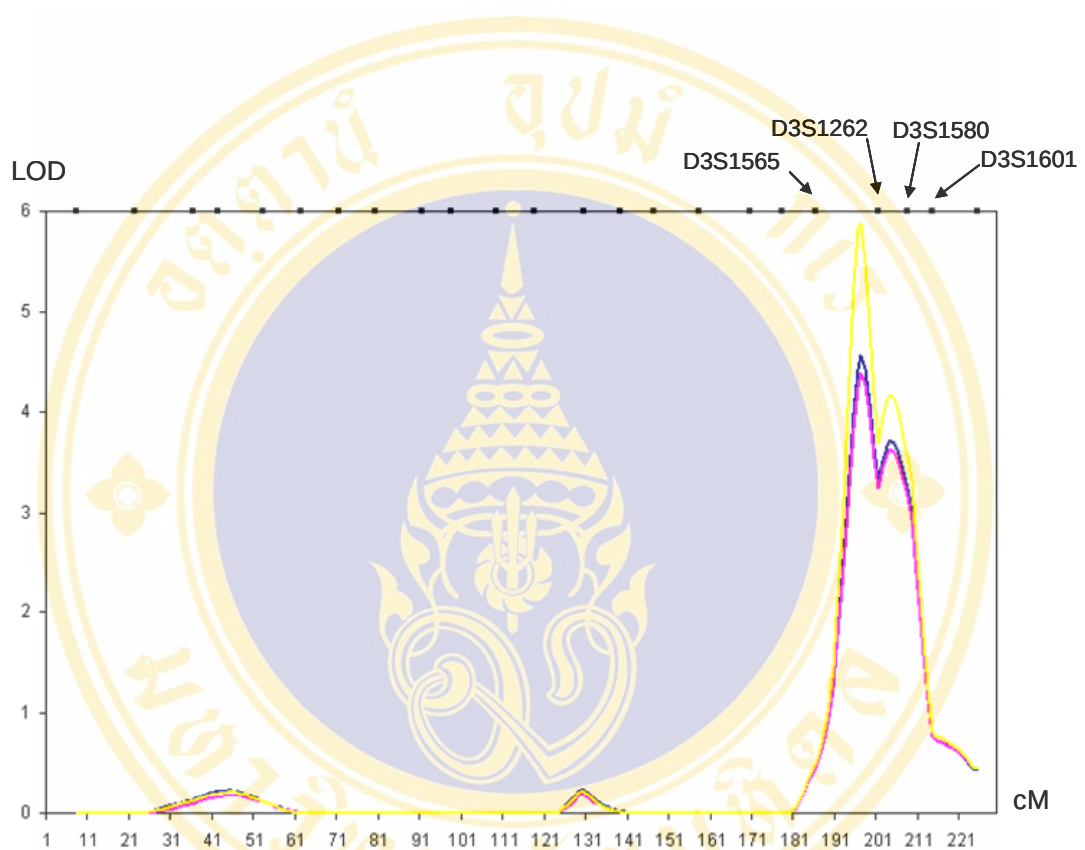


Figure 14. Linkage analysis results on chromosome 3 by Merlin regress with 3 *P. vivax* parasite density phenotypes; maximum, mean, and all *P. vivax* parasite density which are represented in blue, pink, and yellow respectively. The markers flanking the peaks are also shown in the figure.

Table13. The summary of the chromosome positions, methods of analyses, and the phenotypes of which the LOD score is higher than 1.5.

Phenotype	Chromosome	cM	LOD	Method
Number of visits Pv+	8	59	1.72	Solar variance component
Proportion of visits Pv+	8	59	1.69	Solar variance component
Pf max parasite density	5	59	2.25	Regression
	7	58	1.88	Solar variance component
Pf mean parasite density	5	59	2.31	Regression
Pf all parasite densities	5	59	2.30	Regression
Pv max parasite density	1	15	2.75	Regression
	1	28	2.39	Regression
	3	197	4.57	Regression
	4	173	1.74	Solar variance component
	10	81	1.53	Regression
	18	126	1.60	Regression
Pv mean parasite density	1	15	2.62	Regression
	1	28	2.35	Regression
	3	197	4.39	Regression
	10	81	1.63	Regression
	12	17	1.55	Solar variance component
	18	126	1.55	Regression
Pv all parasite densities	1	15	2.65	Regression
	1	28	2.53	Regression
	3	197	5.89	Regression
	10	81	1.61	Regression
	12	20	1.61	Solar variance component
	18	126	1.55	Regression

CHAPTER 6

DISCUSSION

To our knowledge, this research is the first autosome-wide linkage analysis on clinical malaria in Asian population, aiming to identify human genetic variants that are responsible for susceptibility and resistance to the disease. Our clinical malaria longitudinal cohort conducted in a western province of Thailand focusing on phenotype analyses of number of clinical attack and parasite density after controlling for environmental effects.

Malaria endemicity in Suan Phung was homogeneous and transmission of *P. falciparum* and *P. vivax* was perennial, with anophelines biting throughout the year regardless of seasonal fluctuation. Approximately 47% of the population living in 7 hamlets of the study area experienced at least 1 clinical malaria attack during the study period. Total parasite prevalence remained constant throughout 8 years despite a few fluctuations in the relative contribution of the 2 main parasite species; *P. falciparum* (66%) and *P. vivax* (32%). More than half of malaria clinical attacks (54%) occurred in children less than 15 years old, supporting previous observation that children between 1 and 15 years of age had higher risk of malarial infection than adults (1) and that malaria-specific immunity was acquired with repeated exposure and increasing age in endemic areas (2). The number of the attack was higher in children between the age of 1 and 15 than in children less than 1 year old. This may be the result of infants' relatively high malaria resistance during the first 3 months of their lives (3).

P. falciparum predominated in children between 5 and 9 years old, while *P. vivax* predominated in children between 1 and 4 years old. These morbidity data demonstrated that exposure to parasites and clinical immunity development occurred much earlier for *P. vivax* than for *P. falciparum*. Supporting this observation, in the studies in South Pacific and Vanuatu where these 2 malaria species coexist, it was

found that *P. vivax* was responsible for most malaria morbidity during the first two years of life (4, 5). The early development of clinical immunity to *P. vivax* could be explained by the rapid development of strong homologous and partial heterologous immunity to *P. vivax* (6).

For all phenotypes considered, genetic effects explained more of the phenotypic variability than any environmental effect discussed (age group, gender, hamlet, and date). The effect of bednet use was only performed on a small data set for which data was available. It was found to be insignificant and therefore excluded from further analyses. The genetic contributions varied according to the phenotype. Interestingly, the genetic contribution to asexual parasite density was higher for *P. vivax* than for *P. falciparum* (31.9-42% and 19.2-27.8% respectively). This difference could not be explained by house effect because, when included, house had insignificant contribution (or marginal significant) to the total variation of these phenotypes (Table 14 in Appendix). This insignificant contribution of house effect to asexual parasite density phenotypes supports the hypothesis that house can contribute to the tendency to become infected (or ill), but has no impact upon the parasite once the infection has started. Nevertheless, the house effect is sometimes considered inseparable from genetic effect and thus was not included in this study. Moreover, for both species, genetics contributed more to parasite density phenotypes than the clinical attacks (or parasite prevalence). This implies that genetics plays a relatively larger role in determining the outcome of infection once the human is infected than in the probability of infection. Furthermore, it's worth noting that date significantly contributed to the asexual parasitemia of both *P. falciparum* and *P. vivax*. A possible explanation is that, during specific times of the year, there could be some selections of the parasite in which the virulent strains were selected and thus led to high parasitemia in infected individuals. The association of age, genetics and asexual parasite density suggests that immunological mechanisms may be involved. Variation in key immunological mechanisms has been hitherto studied particularly for *P. falciparum* infections and, from our findings here, it is likely that the human genetic basis to parasite control lies in variation in immune responses, although novel globin or erythrocyte mutations may be implicated.

Data analyses by variance component (VC) and Haseman-Elston (HE) linkage methods revealed several prominent linkage results. The most significant ones were on chromosome 3q27 in HE analyses with *P. vivax* parasite density phenotypes. Although this linkage could not be confirmed in VC analyses by either program, it was robust to variations in phenotype definition. To be specific, analyzing the data as maximum, mean, or all parasite density yields approximately the same linkage signals in this region. The fact that no signal was detected by VC analyses could be explained by the very high residual kurtosis of the phenotypes (2.8, 23.9, and 36.7 for maximum, mean, and all *P. vivax* parasite density respectively). Kurtosis defines the degree of peakedness of a distribution relative to a normal distribution (Figure 16 in Appendix). High kurtosis indicates that a trait distribution is not well suited for VC which is less robust than HE methods for non-normality of the trait distribution. The results obtained implies that human genetics plays an important role in determining individual tolerance of infection as well as the overall outcome of human-parasite interaction. In other words, maximum parasite density represents how well ones fight the parasite, and it is incredibly unlikely that the same person would be infected by parasites with the same genotype twice so mean parasite density gives a mean value of how a person responds to an array of parasite genotypes. Thus, other than human genetics, we also support parasite genetics as another factor underlying susceptibility/resistance to the disease.

Several other regions with evidence for linkage were identified. They, although not statistically significant, still provide some clues to the discovery of causative polymorphisms. As has been shown from previous studies of other complex diseases, the underlying causative variants can be found in regions with weaker linkage signals (7). The lack of linkage between the previously identified regions on chromosome 5q31-5q33 (8, 9) and *P. falciparum* parasite density, and those on chromosome 6p21.3 (10, 11) and 10p15.3-10p14 (12) and malaria clinical episodes can be explained by the fact that, here, our study population is Asian which is genetically different from the African in those studies. Therefore, it is not surprising if the genetic control of clinical malarial phenotypes of the 2 populations lies within different genetic variants, although sharing of the same underlying polymorphisms is feasible.

Thalassemia and G6PD are good examples of erythrocyte mutations where their geographical distribution overlaps with that of *P. falciparum* malaria (13, 14). This provides evidence that the genetic polymorphisms causing these abnormalities have been selected to high frequencies because they have protected against malarial infection. In this study, no linkage signals were obtained on either chromosome 11 or 16 where beta and alpha-thalassemia reside respectively. This could reflect the relatively small number of diseased individuals in the population. On the other hand, it has been found that alpha-thalassemia was associated with protection against severe and fatal malaria but had no effect on other clinical malarial phenotypes in Kenyan children (15). Nevertheless, we still cannot exclude these genes from being among the candidate genes contributing to resistance/susceptibility to clinical malaria. G6PD lies on chromosome X which has different pattern of inheritance from autosomal chromosomes. As different genetic linkage algorithm and program must be applied, we did not include chromosome X in this analysis. In addition, several SNPs of these genes are being tested for association with clinical malarial phenotypes in the population.

In the present study, the candidate genes were defined as genes that were involved in red blood cell structure/metabolisms or immunological responses. The linkage region on chromosome 3q27 comprised several genes that could be classified as functional candidates namely dendritic cell-specific lysosome-associated membrane glycoprotein (LAMP3), kelch-like protein 6 (KLHL6), histidine-rich glycoprotein (HRG), and B-cell lymphoma 6 protein (BCL6).

LAMP3 encodes type I integral membrane protein which is expressed in thymus, lymph node, lung, and weakly in spleen, as well as in dendritic cell lines. Expression of LAMP3 is also detected by immunohistochemical analysis in interdigitating cells, but not macrophages, in T-cell areas of lymphoid tissue. De Saint-Vis *et al.* (16) proposed that LAMP3 is a marker specific for mature dendritic cells and that it may be important in the processing of exogenous antigens.

KLHL6 encodes a BTB-kelch protein with a lymphoid tissue-restricted expression pattern. KLHL6 is expressed in the B-lymphocyte lineage and is strongly upregulated in germinal center (GC) B-cells. Kroll *et al.* (17) found that KLHL6 is

important in B-cell antigen receptor signal transduction and formation of the full germinal center response in mice.

HRG encodes serum histidine-rich glycoprotein which has several functions by simultaneous interaction with numerous ligands including certain complement proteins and immunoglobulin. Chang et al. found that HRG bound to complement C8, C9, factor D, and S-protein as well as regulated C9 dimerization (18). Also, HRG was found to interact with FcγRI and block monomeric IgG binding to monocyte (19).

BCL6 functions as a transcriptional repressor and is necessary for germinal center formation. Cattoretti et al. (20) found that BCL6 is specifically regulated during B-cell differentiation and suggested a role for BCL6 in germinal center development or function.

Linkage analysis of mouse malaria infected by *P. chabaudi* revealed several regions on chromosome 3, 5, 9, 11, and 17 with evidence of linkage to parasite density. Our linkage region on chromosome 3q27 can be referred to its mouse syntenic ones on chromosome 3 and 16 (NCBI, <http://www.ncbi.nlm.nih.gov/Homology/>). Nevertheless, the linkage region on mouse chromosome 3 is syntenic to human chromosome 4q21-4q25, not 3q27. This difference in linkage regions between the previous study in mouse and this study can be explained by the fact that *P. chabaudi* produces an infection in mice that shares many similarities with human *P. falciparum* but not *P. vivax* malaria. Thus, the parasite density of *P. chabaudi* and *P. vivax* could be controlled by different genetic variants.

The only known human genetic influence on resistance to *P. vivax* infection, as far as we know, is the effect of Duffy blood group antigen on the inability of *P. vivax* merozoites to invade Duffy negative red blood cells (21-23). However, the Duffy negative blood group is mostly found in West Africa (24). We have screened for the mutation in 132 villagers and found that none of them had the negative blood group, thus, the antigen is likely of little importance here. This study gives a new insight to *P. vivax* infection; the genetic control of the parasite density. Moreover, to our knowledge, this is the first systematic evidence that *P. vivax*, as well as *P. falciparum*, had exerted selective pressure on human genome supporting the similar idea proposed earlier (13).

In this study, the original pedigrees were partitioned into sub-pedigrees using the simulated annealing algorithm. Since linkage analysis using an exact approach, as implemented in Merlin, imposes strict constraints on the complexity of the pedigree it can analyze, it was necessary to break the original pedigrees into sub-pedigrees with maximum complexities of 20 bits. For the estimation of IBD matrices by Loki and linkage analyses by SOLAR, pedigrees of arbitrary size and complexity can be used. However, the estimation of IBD matrices using a stochastic (random) method such as in Loki can face problems with convergence, particularly for large pedigrees. We therefore also partitioned the pedigree into structures with maximum complexities of 100 bits to perform analysis with Loki. Nevertheless, although the 100 bits maximum permitted complexity pedigrees used to run Loki had 96.9% of original structure retained, it is worth analyzing the data using the original pedigrees and comparing both results. Further study should then be performed on the linked region either by positional candidate gene or fine mapping approach. In positional candidate gene approach, the most promising candidate genes within the mapped region will be selected for direct sequencing in the whole population to search for polymorphisms. These polymorphisms will then be tested for association with the phenotypes. For fine mapping approach, denser genetic markers (for example, every cM) will be selected and genotyped. Then, linkage analyses will be performed to narrow down the region linked to the phenotypes.

CHAPTER 7

CONCLUSION

This research was conducted in the cohort of Karen population from Suan Phung district, Ratchaburi province, located near the Thai-Myanmar border of Thailand. Phenotypes related to clinical malaria attacks and parasite density have been investigated in 3,481 participants living in 7 hamlets during year 1994-2005. Statistical analyses and model fitting were applied to the raw data and Pearson's residual of each phenotype was then calculated. Family-based genetic variance component analysis was used to determine the relative contributions of genetic and environmental factors to the total variation of malarial phenotypes of *P. falciparum* and *P. vivax*. From the result, it appeared that genetics contributed more to parasite density phenotypes than to clinical attacks (parasite prevalence) for both species and genetic contribution to asexual parasite density of *P. vivax* was higher than that of *P. falciparum*. Variance component (VC) and Haseman-Elston (HE) approach of linkage analyses were then performed on the phenotypes in which the genetic effect was significantly estimated. The most significant results were identified in HE analyses with all 3 *P. vivax* parasite density phenotypes on chromosome 3q27.1-3q27.3 at 197 cM (between markers D3S1565-D3S1262) with the LOD score of 4.39-5.89 (p -value $< 1 \times 10^{-5}$). The 1 LOD score support interval encompassed a 14 cM/ 6.4 Mb distance containing 62 annotated or hypothetical genes. Functional candidates include genes encoding dendritic cell-specific lysosome-associated membrane glycoprotein (LAMP3), kelch-like protein 6 (KLHL6), histidine-rich glycoprotein (HRG), and B-cell lymphoma 6 protein (BCL6). No support for our data can be derived from previous linkage study in mouse malaria. This study provides a systematic evidence of the genetic control of *P. vivax* parasite density as well as the selective pressure that this parasite species had exerted on human genome.

REFERENCES

1. WHO. WHO Expert Committee on Malaria. World Health Organ Tech Rep Ser. 2000;892:i-v, 1-74.
2. Haldane JBS. Disease and evolution. Ric Sci Suppl 1949;A(19):68-76.
3. Hill AV. Malaria resistance genes: a natural selection. Trans R Soc Trop Med Hyg. 1992 May-Jun;86(3):225-6, 32.
4. Weatherall DJ. Host genetics and infectious disease. Parasitology. 1996;112 Suppl:S23-9.
5. Burt RA. Genetics of host response to malaria. Int J Parasitol. 1999 Jun;29(6):973-9.
6. Garcia A, Marquet S, Bucheton B, Hillaire D, Cot M, Fievet N, et al. Linkage analysis of blood *Plasmodium falciparum* levels: interest of the 5q31-q33 chromosome region. Am J Trop Med Hyg. 1998 Jun;58(6):705-9.
7. Rihet P, Traore Y, Abel L, Aucan C, Traore-Leroux T, Fumoux F. Malaria in humans: *Plasmodium falciparum* blood infection levels are linked to chromosome 5q31-q33. Am J Hum Genet. 1998 Aug;63(2):498-505.
8. Jepson AP, Banya WA, Sisay-Joof F, Hassan-King M, Bennett S, Whittle HC. Genetic regulation of fever in *Plasmodium falciparum* malaria in Gambian twin children. J Infect Dis. 1995 Jul;172(1):316-9.
9. McGuire W, Knight JC, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D. Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles. J Infect Dis. 1999 Jan;179(1):287-90.
10. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, et al. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991 Aug 15;352(6336):595-600.
11. Hill AV, Elvin J, Willis AC, Aidoo M, Allsopp CE, Gotch FM, et al. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. Nature. 1992 Dec 3;360(6403):434-9.

12. Burgner D, Xu W, Rockett K, Gravenor M, Charles IG, Hill AV, et al. Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet*. 1998 Oct 10;352(9135):1193-4.
13. Kun JF, Mordmuller B, Lell B, Lehman LG, Luckner D, Kremsner PG. Polymorphism in promoter region of inducible nitric oxide synthase gene and protection against malaria. *Lancet*. 1998 Jan 24;351(9098):265-6.
14. Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, et al. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis*. 2008 Jan 15;46(2):165-71.
15. White NJ. *Plasmodium knowlesi*: the fifth human malaria parasite. *Clin Infect Dis*. 2008 Jan 15;46(2):172-3.
16. WHO. Severe *falciparum* malaria. World Health Organization, Communicable Diseases Cluster. *Trans R Soc Trop Med Hyg*. 2000 Apr;94 Suppl 1:S1-90.
17. Gupta S, Day KP. A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasite Immunol*. 1994 Jul;16(7):361-70.
18. Rogier C, Commenges D, Trape JF. Evidence for an age-dependent pyrogenic threshold of *Plasmodium falciparum* parasitemia in highly endemic populations. *Am J Trop Med Hyg*. 1996 Jun;54(6):613-9.
19. WHO. Thailand-Malaria Situation in SEAR Countries.; 2004.
20. Malaria control program in Thailand, Ministry of Public Health, Thailand.; 2002.
21. Khoury M, Beaty T, Cohen B. *Fundamentals of Genetic Epidemiology*. Oxford: Oxford University Press; 1993.
22. Strachan T, Read A. *Human Molecular Genetics 2*. Singapore: John Wiley & Sons (Asia) Pte Ltd; 1999.
23. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. 1998 May 15;280(5366):1077-82.
24. Hardy GH. Mendelian proportions in a mixed population. *Science*. 1908 Jul 10;28(706):49-50.
25. King MC, Lee GM, Spinner NB, Thomson G, Wrensch MR. Genetic epidemiology. *Annu Rev Public Health*. 1984;5:1-52.

26. Beaty TH, Yang P, Munoz A, Khoury MJ. Effect of maternal and infant covariates on sibship correlation in birth weight. *Genet Epidemiol.* 1988;5(4):241-53.
27. del Junco D, Luthra HS, Annegers JF, Worthington JW, Kurland LT. The familial aggregation of rheumatoid arthritis and its relationship to the HLA-DR4 association. *Am J Epidemiol.* 1984 May;119(5):813-29.
28. Sattin RW, Rubin GL, Webster LA, Huezo CM, Wingo PA, Ory HW, et al. Family history and the risk of breast cancer. *Jama.* 1985 Apr 5;253(13):1908-13.
29. Feinleib M, Garrison RJ, Fabsitz R, Christian JC, Hrubec Z, Borhani NO, et al. The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am J Epidemiol.* 1977 Oct;106(4):284-5.
30. Cohen BH. Chronic obstructive pulmonary disease: a challenge in genetic epidemiology. *Am J Epidemiol.* 1980 Aug;112(2):274-88.
31. Linet MS, Van Natta ML, Brookmeyer R, Khoury MJ, McCaffrey LD, Humphrey RL, et al. Familial cancer history and chronic lymphocytic leukemia. A case-control study. *Am J Epidemiol.* 1989 Oct;130(4):655-64.
32. Khoury MJ, Erickson JD, James LM. Etiologic heterogeneity of neural tube defects. II. Clues from family studies. *Am J Hum Genet.* 1982 Nov;34(6):980-7.
33. Lilienfeld AM. Formal discussion of: genetic factors in the etiology of cancer: an epidemiologic view. *Cancer Res.* 1965 Sep;25(8):1330-5.
34. Phillips PH, Linet MS, Harris EL. Assessment of family history information in case-control cancer studies. *Am J Epidemiol.* 1991 Apr 15;133(8):757-65.
35. Weiss KM, Chakraborty R, Majumder PP, Smouse PE. Problems in the assessment of relative risk of chronic disease among biological relatives of affected individuals. *J Chronic Dis.* 1982;35(7):539-51.
36. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet.* 1999 Dec;23(4):397-404.
37. Forabosco P, Falchi M, Devoto M. Statistical tools for linkage analysis and genetic association studies. *Expert Rev Mol Diagn.* 2005 Sep;5(5):781-96.

38. Morton NE. Sequential tests for the detection of linkage. *Am J Hum Genet.* 1955 Sep;7(3):277-318.
39. Ott J. *Analysis of Human Genetic Linkage.* Third Edition ed: The Johns Hopkins University Press, MD, USA; 1999.
40. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. *Hum Hered.* 1971;21(6):523-42.
41. Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA.* 1987 Apr;84(8):2363-7.
42. Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet.* 1996 Jun;58(6):1323-37.
43. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet.* 1995 Nov;11(3):241-7.
44. Whittemore AS, Halpern J. Probability of gene identity by descent: computation and applications. *Biometrics.* 1994 Mar;50(1):109-17.
45. Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet.* 1972 Mar;2(1):3-19.
46. Sham PC, Purcell S, Cherny SS, Abecasis GR. Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *Am J Hum Genet.* 2002 Aug;71(2):238-53.
47. Almasy L, Warren DM. Software for quantitative trait analysis. *Hum Genomics.* 2005 Sep;2(3):191-5.
48. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet.* 1994 Mar;54(3):535-43.
49. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* 1998 May;62(5):1198-211.
50. Hopper JL, Mathews JD. Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet.* 1982 Oct;46(Pt 4):373-83.
51. Nyholt DR. GENEHUNTER: your 'one-stop shop' for statistical genetic analysis? *Hum Hered.* 2002;53(1):2-7.
52. Dudbridge F. A survey of current software for linkage analysis. *Hum Genomics.* 2003 Nov;1(1):63-5.

53. Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet.* 1997 Sep;61(3):748-60.
54. Fulker DW, Cardon LR. A sib-pair approach to interval mapping of quantitative trait loci. *Am J Hum Genet.* 1994 Jun;54(6):1092-103.
55. Elston RC, Buxbaum S, Jacobs KB, Olson JM. Haseman and Elston revisited. *Genet Epidemiol.* 2000 Jul;19(1):1-17.
56. Palmer LJ, Jacobs KB, Elston RC. Haseman and Elston revisited: the effects of ascertainment and residual familial correlations on power to detect linkage. *Genet Epidemiol.* 2000 Dec;19(4):456-60.
57. Forrest WF. Weighting improves the "new Haseman-Elston" method. *Hum Hered.* 2001;52(1):47-54.
58. Xu X, Weiss S, Xu X, Wei LJ. A unified Haseman-Elston method for testing linkage with quantitative traits. *Am J Hum Genet.* 2000 Oct;67(4):1025-8.
59. Wang T, Elston RC. Two-level Haseman-Elston regression for general pedigree data analysis. *Genet Epidemiol.* 2005 Jul;29(1):12-22.
60. de Andrade M, Amos CI, Thiel TJ. Methods to estimate genetic components of variance for quantitative traits in family studies. *Genet Epidemiol.* 1999;17(1):64-76.
61. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet.* 1996 Jun;58(6):1347-63.
62. Kruglyak L, Lander ES. Faster multipoint linkage analysis using Fourier transforms. *J Comput Biol.* 1998 Spring;5(1):1-7.
63. Lange K, Cantor R, Horvath S, et al. 'Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets'. *Am J Hum Genet.* 2001;69 (Suppl.):504.
64. Lange K, Weeks D, Boehnke M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol.* 1988;5(6):471-2.
65. Stringham HM, Boehnke M. Identifying marker typing incompatibilities in linkage analysis. *Am J Hum Genet.* 1996 Oct;59(4):946-50.

66. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002 Jan;30(1):97-101.
67. Neale MC. 'The use of Mx for association and linkage analysis'. *GeneScreen.* 2000;1:107-11.
68. Neale MC, Boker SM, Xie G, Maes HH. *Mx: Statistical Modeling.* 6th ed: Department of Psychiatry, VCU, Richmond, VA.; 2003.
69. Posthuma D, de Geus EJ, Boomsma DI, Neale MC. Combined linkage and association tests in mx. *Behav Genet.* 2004 Mar;34(2):179-96.
70. S.A.G.E. *Statistical Analysis for Genetic Epidemiology*—computer program package available from Statistical Solutions Ltd, Cork, Ireland.; 2004.
71. Province MA, Rao DC. General purpose model and a computer program for combined segregation and path analysis (SEGPATH): automatically creating computer programs from symbolic language model specifications. *Genet Epidemiol.* 1995;12(2):203-19.
72. Province MA, Rice TK, Borecki IB, Gu C, Kraja A, Rao DC. Multivariate and multilocus variance components method, based on structural relationships to assess quantitative trait linkage via SEGPATH. *Genet Epidemiol.* 2003 Feb;24(2):128-38.
73. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet.* 2005 Aug;77(2):171-92.
74. Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet.* 1995 Jun;10(2):224-8.
75. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med.* 1976 Aug 5;295(6):302-4.
76. Hamblin MT, Di Rienzo A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet.* 2000 May;66(5):1669-79.

77. Hamblin MT, Thompson EE, Di Rienzo A. Complex signatures of natural selection at the Duffy blood group locus. *Am J Hum Genet.* 2002 Feb;70(2):369-83.
78. Zimmerman PA, Woolley I, Masinde GL, Miller SM, McNamara DT, Hazlett F, et al. Emergence of FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea. *Proc Natl Acad Sci U S A.* 1999 Nov 23;96(24):13973-7.
79. Williams TN, Maitland K, Bennett S, Ganczakowski M, Peto TE, Newbold CI, et al. High incidence of malaria in alpha-thalassaemic children. *Nature.* 1996 Oct 10;383(6600):522-5.
80. Ganczakowski M, Town M, Bowden DK, Vulliamy TJ, Kaneko A, Clegg JB, et al. Multiple glucose 6-phosphate dehydrogenase-deficient variants correlate with malaria endemicity in the Vanuatu archipelago (southwestern Pacific). *Am J Hum Genet.* 1995 Jan;56(1):294-301.
81. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature.* 2002 Oct 24;419(6909):832-7.
82. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, et al. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science.* 2001 Jul 20;293(5529):455-62.
83. Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, et al. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature.* 1995 Jul 20;376(6537):246-9.
84. Gilles HM, Fletcher KA, Hendrickse RG, Lindner R, Reddy S, Allan N. Glucose-6-phosphate-dehydrogenase deficiency, sickling, and malaria in African children in South Western Nigeria. *Lancet.* 1967 Jan 21;1(7482):138-40.
85. Luzzatto L, Usanga FA, Reddy S. Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites. *Science.* 1969 May 16;164(881):839-42.
86. Usanga EA, Luzzatto L. Adaptation of Plasmodium falciparum to glucose 6-phosphate dehydrogenase-deficient host red cells by production of parasite-encoded enzyme. *Nature.* 1985 Feb 28-Mar 6;313(6005):793-5.

87. Facer CA. Merozoites of *P. falciparum* require glycophorin for invasion into red cells. *Bull Soc Pathol Exot Filiales*. 1983 Nov;76(5):463-9.
88. Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, Zimmerman PA, et al. *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat Med*. 2003 Jan;9(1):87-92.
89. Hayes RJ, Marsh K, Snow RW. Case-control studies of severe malaria. *J Trop Med Hyg*. 1992 Jun;95(3):157-66.
90. Allen SJ, Bennett S, Riley EM, Rowe PA, Jakobsen PH, O'Donnell A, et al. Morbidity from malaria and immune responses to defined *Plasmodium falciparum* antigens in children with sickle cell trait in The Gambia. *Trans R Soc Trop Med Hyg*. 1992 Sep-Oct;86(5):494-8.
91. Sokhna CS, Rogier C, Dieye A, Trape JF. Host factors affecting the delay of reappearance of *Plasmodium falciparum* after radical treatment among a semi-immune population exposed to intense perennial transmission. *Am J Trop Med Hyg*. 2000 Feb;62(2):266-70.
92. Stirnadel HA, Stockle M, Felger I, Smith T, Tanner M, Beck HP. Malaria infection and morbidity in infants in relation to genetic polymorphisms in Tanzania. *Trop Med Int Health*. 1999 Mar;4(3):187-93.
93. Pasvol G, Weatherall DJ, Wilson RJ. Cellular mechanism for the protective effect of haemoglobin S against *P. falciparum* malaria. *Nature*. 1978 Aug 17;274(5672):701-3.
94. Shear HL, Roth EF, Jr., Fabry ME, Costantini FD, Pachnis A, Hood A, et al. Transgenic mice expressing human sickle hemoglobin are partially resistant to rodent malaria. *Blood*. 1993 Jan 1;81(1):222-6.
95. Williams TN, Mwangi TW, Roberts DJ, Alexander ND, Weatherall DJ, Wambua S, et al. An immune basis for malaria protection by the sickle cell trait. *PLoS Med*. 2005 May;2(5):e128.
96. Agarwal A, Guindo A, Cissoko Y, Taylor JG, Coulibaly D, Kone A, et al. Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood*. 2000 Oct 1;96(7):2358-63.

97. Mockenhaupt FP, Ehrhardt S, Cramer JP, Otchwemah RN, Anemana SD, Goltz K, et al. Hemoglobin C and resistance to severe malaria in Ghanaian children. *J Infect Dis.* 2004 Sep 1;190(5):1006-9.
98. Rihet P, Flori L, Tall F, Traore AS, Fumoux F. Hemoglobin C is associated with reduced *Plasmodium falciparum* parasitemia and low risk of mild malaria attack. *Hum Mol Genet.* 2004 Jan 1;13(1):1-6.
99. Modiano D, Luoni G, Sirima BS, Simporé J, Verra F, Konate A, et al. Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature.* 2001 Nov 15;414(6861):305-8.
100. Fairhurst RM, Baruch DI, Brittain NJ, Ostera GR, Wallach JS, Hoang HL, et al. Abnormal display of PfEMP-1 on erythrocytes carrying haemoglobin C may protect against malaria. *Nature.* 2005 Jun 23;435(7045):1117-21.
101. Chotivanich K, Udomsangpetch R, Pattanapanyasat K, Chierakul W, Simpson J, Looareesuwan S, et al. Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe *P. falciparum* malaria. *Blood.* 2002 Aug 15;100(4):1172-6.
102. Modiano G, Morpurgo G, Terrenato L, Novelletto A, Di Rienzo A, Colombo B, et al. Protection against malaria morbidity: near-fixation of the alpha-thalassemia gene in a Nepalese population. *Am J Hum Genet.* 1991 Feb;48(2):390-7.
103. Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, et al. High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature.* 1986 Jun 19-25;321(6072):744-50.
104. Williams TN, Wambua S, Uyoga S, Macharia A, Mwacharo JK, Newton CR, et al. Both heterozygous and homozygous alpha+ thalassemias protect against severe and fatal *Plasmodium falciparum* malaria on the coast of Kenya. *Blood.* 2005 Jul 1;106(1):368-71.
105. Mockenhaupt FP, Ehrhardt S, Gellert S, Otchwemah RN, Dietz E, Anemana SD, et al. Alpha(+)-thalassemia protects African children from severe malaria. *Blood.* 2004 Oct 1;104(7):2003-6.
106. Allen SJ, O'Donnell A, Alexander ND, Alpers MP, Peto TE, Clegg JB, et al. alpha+-Thalassemia protects children against disease caused by other

- infections as well as malaria. *Proc Natl Acad Sci U S A*. 1997 Dec 23;94(26):14736-41.
107. Cattani JA, Gibson FD, Alpers MP, Crane GG. Hereditary ovalocytosis and reduced susceptibility to malaria in Papua New Guinea. *Trans R Soc Trop Med Hyg*. 1987;81(5):705-9.
 108. Foo LC, Rekhraj V, Chiang GL, Mak JW. Ovalocytosis protects against severe malaria parasitemia in the Malayan aborigines. *Am J Trop Med Hyg*. 1992 Sep;47(3):271-5.
 109. Allen SJ, O'Donnell A, Alexander ND, Mgone CS, Peto TE, Clegg JB, et al. Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. *Am J Trop Med Hyg*. 1999 Jun;60(6):1056-60.
 110. Genton B, al-Yaman F, Mgone CS, Alexander N, Paniu MM, Alpers MP, et al. Ovalocytosis and cerebral malaria. *Nature*. 1995 Dec 7;378(6557):564-5.
 111. Bienzle U, Ayeni O, Lucas AO, Luzzatto L. Glucose-6-phosphate dehydrogenase and malaria. Greater resistance of females heterozygous for enzyme deficiency and of males with non-deficient variant. *Lancet*. 1972 Jan 15;1(7742):107-10.
 112. Baum J, Ward RH, Conway DJ. Natural selection on the erythrocyte surface. *Mol Biol Evol*. 2002 Mar;19(3):223-9.
 113. Mayor A, Bir N, Sawhney R, Singh S, Pattnaik P, Singh SK, et al. Receptor-binding residues lie in central regions of Duffy-binding-like domains involved in red cell invasion and cytoadherence by malaria parasites. *Blood*. 2005 Mar 15;105(6):2557-63.
 114. Wang HY, Tang H, Shen CK, Wu CI. Rapidly evolving genes in human. I. The glycoporphins and their possible role in evading malaria parasites. *Mol Biol Evol*. 2003 Nov;20(11):1795-804.
 115. Pattanapanyasat K, Yongvanitchit K, Tongtawe P, Tachavanich K, Wanachiwanawin W, Fucharoen S, et al. Impairment of *Plasmodium falciparum* growth in thalassemic red blood cells: further evidence by using biotin labeling and flow cytometry. *Blood*. 1999 May 1;93(9):3116-9.

116. Roberts DJ, Craig AG, Berendt AR, Pinches R, Nash G, Marsh K, et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature*. 1992 Jun 25;357(6380):689-92.
117. Aitman TJ, Cooper LD, Norsworthy PJ, Wahid FN, Gray JK, Curtis BR, et al. Malaria susceptibility and CD36 mutation. *Nature*. 2000 Jun 29;405(6790):1015-6.
118. Omi K, Ohashi J, Patarapotikul J, Hananantachai H, Naka I, Looareesuwan S, et al. CD36 polymorphism is associated with protection from cerebral malaria. *Am J Hum Genet*. 2003 Feb;72(2):364-74.
119. Pain A, Urban BC, Kai O, Casals-Pascual C, Shafi J, Marsh K, et al. A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet*. 2001 May 12;357(9267):1502-3.
120. Fernandez-Reyes D, Craig AG, Kyes SA, Peshu N, Snow RW, Berendt AR, et al. A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum Mol Genet*. 1997 Aug;6(8):1357-60.
121. Kun JF, Klabunde J, Lell B, Luckner D, Alpers M, May J, et al. Association of the ICAM-1Kilifi mutation with protection against severe malaria in Lambarene, Gabon. *Am J Trop Med Hyg*. 1999 Nov;61(5):776-9.
122. Bellamy R, Kwiatkowski D, Hill AV. Absence of an association between intercellular adhesion molecule 1, complement receptor 1 and interleukin 1 receptor antagonist gene polymorphisms and severe malaria in a West African population. *Trans R Soc Trop Med Hyg*. 1998 May-Jun;92(3):312-6.
123. Rowe JA, Moulds JM, Newbold CI, Miller LH. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature*. 1997 Jul 17;388(6639):292-5.
124. Cockburn IA, Mackinnon MJ, O'Donnell A, Allen SJ, Moulds JM, Baisor M, et al. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. *Proc Natl Acad Sci U S A*. 2004 Jan 6;101(1):272-7.

125. Zimmerman PA, Fitness J, Moulds JM, McNamara DT, Kasehagen LJ, Rowe JA, et al. CR1 Knops blood group alleles are not associated with severe malaria in the Gambia. *Genes Immun.* 2003 Jul;4(5):368-73.
126. Nagayasu E, Ito M, Akaki M, Nakano Y, Kimura M, Looareesuwan S, et al. CR1 density polymorphism on erythrocytes of falciparum malaria patients in Thailand. *Am J Trop Med Hyg.* 2001 Jan-Feb;64(1-2):1-5.
127. Bennett S, Allen SJ, Olerup O, Jackson DJ, Wheeler JG, Rowe PA, et al. Human leucocyte antigen (HLA) and malaria morbidity in a Gambian community. *Trans R Soc Trop Med Hyg.* 1993 May-Jun;87(3):286-7.
128. Koch O, Rockett K, Jallow M, Pinder M, Sisay-Joof F, Kwiatkowski D. Investigation of malaria susceptibility determinants in the IFNG/IL26/IL22 genomic region. *Genes Immun.* 2005 Jun;6(4):312-8.
129. Koch O, Awomoyi A, Usen S, Jallow M, Richardson A, Hull J, et al. IFNGR1 gene promoter polymorphisms and susceptibility to cerebral malaria. *J Infect Dis.* 2002 Jun 1;185(11):1684-7.
130. Aucan C, Walley AJ, Hennig BJ, Fitness J, Frodsham A, Zhang L, et al. Interferon-alpha receptor-1 (IFNAR1) variants are associated with protection against cerebral malaria in the Gambia. *Genes Immun.* 2003 Jun;4(4):275-82.
131. Wilson JN, Rockett K, Jallow M, Pinder M, Sisay-Joof F, Newport M, et al. Analysis of IL10 haplotypic associations with severe malaria. *Genes Immun.* 2005 Sep;6(6):462-6.
132. Morahan G, Boutlis CS, Huang D, Pain A, Saunders JR, Hobbs MR, et al. A promoter polymorphism in the gene encoding interleukin-12 p40 (IL12B) is associated with mortality from cerebral malaria and with reduced nitric oxide production. *Genes Immun.* 2002 Nov;3(7):414-8.
133. Luoni G, Verra F, Arca B, Sirima BS, Troye-Blomberg M, Coluzzi M, et al. Antimalarial antibody levels and IL4 polymorphism in the Fulani of West Africa. *Genes Immun.* 2001 Nov;2(7):411-4.
134. Kun JF, Mordmuller B, Perkins DJ, May J, Mercereau-Puijalon O, Alpers M, et al. Nitric oxide synthase 2(Lambarene) (G-954C), increased nitric oxide production, and protection against malaria. *J Infect Dis.* 2001 Aug 1;184(3):330-6.

135. Burgner D, Usen S, Rockett K, Jallow M, Ackerman H, Cervino A, et al. Nucleotide and haplotypic diversity of the NOS2A promoter region and its relationship to cerebral malaria. *Hum Genet.* 2003 Apr;112(4):379-86.
136. Hobbs MR, Udhayakumar V, Levesque MC, Booth J, Roberts JM, Tkachuk AN, et al. A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children. *Lancet.* 2002 Nov 9;360(9344):1468-75.
137. Kwiatkowski D. Malarial toxins and the regulation of parasite density. *Parasitol Today.* 1995 Jun;11(6):206-12.
138. Flori L, Sawadogo S, Esnault C, Delahaye NF, Fumoux F, Rihet P. Linkage of mild malaria to the major histocompatibility complex in families living in Burkina Faso. *Hum Mol Genet.* 2003 Feb 15;12(4):375-8.
139. Jepson A, Sisay-Joof F, Banya W, Hassan-King M, Frodsham A, Bennett S, et al. Genetic linkage of mild malaria to the major histocompatibility complex in Gambian children: study of affected sibling pairs. *Bmj.* 1997 Jul 12;315(7100):96-7.
140. McGuire W, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D. Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature.* 1994 Oct 6;371(6497):508-10.
141. Meyer CG, May J, Luty AJ, Lell B, Kremsner PG. TNFalpha-308A associated with shorter intervals of *Plasmodium falciparum* reinfections. *Tissue Antigens.* 2002 Apr;59(4):287-92.
142. Wattavidanage J, Carter R, Perera KL, Munasingha A, Bandara S, McGuinness D, et al. TNFalpha*2 marks high risk of severe disease during *Plasmodium falciparum* malaria and other infections in Sri Lankans. *Clin Exp Immunol.* 1999 Feb;115(2):350-5.
143. Aidoo M, McElroy PD, Kolczak MS, Terlouw DJ, ter Kuile FO, Nahlen B, et al. Tumor necrosis factor-alpha promoter variant 2 (TNF2) is associated with pre-term delivery, infant mortality, and malaria morbidity in western Kenya: Asembo Bay Cohort Project IX. *Genet Epidemiol.* 2001 Nov;21(3):201-11.

144. Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N, Marsh K, et al. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet.* 1999 Jun;22(2):145-50.
145. Flori L, Delahaye NF, Iraqi FA, Hernandez-Valladares M, Fumoux F, Rihet P. TNF as a malaria candidate gene: polymorphism-screening and family-based association analysis of mild malaria attack and parasitemia in Burkina Faso. *Genes Immun.* 2005 Sep;6(6):472-80.
146. Shi YP, Nahlen BL, Kariuki S, Urdahl KB, McElroy PD, Roberts JM, et al. Fcgamma receptor IIa (CD32) polymorphism is associated with protection of infants against high-density *Plasmodium falciparum* infection. VII. Asembo Bay Cohort Project. *J Infect Dis.* 2001 Jul 1;184(1):107-11.
147. Cooke GS, Aucan C, Walley AJ, Segal S, Greenwood BM, Kwiatkowski DP, et al. Association of Fcgamma receptor IIa (CD32) polymorphism with severe malaria in West Africa. *Am J Trop Med Hyg.* 2003 Dec;69(6):565-8.
148. Omi K, Ohashi J, Patarapotikul J, Hananantachai H, Naka I, Looareesuwan S, et al. Fcgamma receptor IIA and IIIB polymorphisms are associated with susceptibility to cerebral malaria. *Parasitol Int.* 2002 Dec;51(4):361-6.
149. Mackinnon MJ, Gunawardena DM, Rajakaruna J, Weerasingha S, Mendis KN, Carter R. Quantifying genetic and nongenetic contributions to malarial infection in a Sri Lankan population. *Proc Natl Acad Sci U S A.* 2000 Nov 7;97(23):12661-6.
150. Abel L, Cot M, Mulder L, Carnevale P, Feingold J. Segregation analysis detects a major gene controlling blood infection levels in human malaria. *Am J Hum Genet.* 1992 Jun;50(6):1308-17.
151. Garcia A, Cot M, Chippaux JP, Ranque S, Feingold J, Demenais F, et al. Genetic control of blood infection levels in human malaria: evidence for a complex genetic model. *Am J Trop Med Hyg.* 1998 Apr;58(4):480-8.
152. Rihet P, Abel L, Traore Y, Traore-Leroux T, Aucan C, Fumoux F. Human malaria: segregation analysis of blood infection levels in a suburban area and a rural area in Burkina Faso. *Genet Epidemiol.* 1998;15(5):435-50.

153. Sjoberg K, Lepers JP, Raharimalala L, Larsson A, Olerup O, Marbiah NT, et al. Genetic regulation of human anti-malarial antibodies in twins. *Proc Natl Acad Sci U S A*. 1992 Mar 15;89(6):2101-4.
154. Stirnadel HA, Al-Yaman F, Genton B, Alpers MP, Smith TA. Assessment of different sources of variation in the antibody responses to specific malaria antigens in children in Papua New Guinea. *Int J Epidemiol*. 2000 Jun;29(3):579-86.
155. Stirnadel HA, Beck HP, Alpers MP, Smith TA. Heritability and segregation analysis of immune responses to specific malaria antigens in Papua New Guinea. *Genet Epidemiol*. 1999;17(1):16-34.
156. Stirnadel HA, Beck HP, Alpers MP, Smith TA. Genetic analysis of IgG subclass responses against RESA and MSP2 of *Plasmodium falciparum* in adults in Papua New Guinea. *Epidemiol Infect*. 2000 Feb;124(1):153-62.
157. Ranque S, Safeukui I, Poudiougou B, Traore A, Keita M, Traore D, et al. Familial aggregation of cerebral malaria and severe malarial anemia. *J Infect Dis*. 2005 Mar 1;191(5):799-804.
158. Luxemburger C, Price RN, Nosten F, Ter Kuile FO, Chongsuphajaisiddhi T, White NJ. Mefloquine in infants and young children. *Ann Trop Paediatr*. 1996 Dec;16(4):281-6.
159. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of malaria in Africa. *PLoS Med*. 2005 Dec;2(12):e340.
160. Krotoski WA, Collins WE, Bray RS, Garnham PC, Cogswell FB, Gwadz RW, et al. Demonstration of hypnozoites in sporozoite-transmitted *Plasmodium vivax* infection. *Am J Trop Med Hyg*. 1982 Nov;31(6):1291-3.
161. Coatney GR, Cooper WC, Ruhe DS, Young MD, Burgess RW. Studies in human malaria. XVIII. The life pattern of sporozoite induced St. Elizabeth strain *vivax* malaria. *Am J Hyg* 1950;51:200-15.
162. Hankey DD, Jones R, Jr., Coatney GR, Alving AS, Coker WG, Garrison PL, et al. Korean *vivax* malaria. I. Natural history and response to chloroquine. *Am J Trop Med Hyg*. 1953 Nov;2(6):958-69.

163. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A*. 1992 Jul 1;89(13):5847-51.
164. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*. 1998 Jul;63(1):259-66.
165. Visscher PM, Haley CS, Heath SC, Muir WJ, Blackwood DH. Detecting QTLs for uni- and bipolar disorder using a variance component method. *Psychiatr Genet*. 1999 Jun;9(2):75-84.
166. de Koning DJ, Pong-Wong R, Varona L, Evans GJ, Giuffra E, Sanchez A, et al. Full pedigree quantitative trait locus analysis in commercial pigs using variance components. *J Anim Sci*. 2003 Sep;81(9):2155-63.
167. Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP, Weeks DE. Mega2: data-handling for facilitating genetic linkage and association analyses. *Bioinformatics*. 2005 May 15;21(10):2556-7.
168. Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc*. 1987;82:605-10.
169. Allison DB, Fernandez JR, Heo M, Beasley TM. Testing the robustness of the new Haseman-Elston quantitative-trait loci-mapping procedure. *Am J Hum Genet*. 2000 Jul;67(1):249-52.
170. Fulker DW, Cherny SS. An improved multipoint sib-pair analysis of quantitative traits. *Behav Genet*. 1996 Sep;26(5):527-32.
171. Sham PC, Purcell S. Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. *Am J Hum Genet*. 2001 Jun;68(6):1527-32.
172. Timmann C, Evans JA, Konig IR, Kleensang A, Ruschendorf F, Lenzen J, et al. Genome-wide linkage analysis of malaria infection intensity and mild disease. *PLoS Genet*. 2007 Mar 23;3(3):e48.
173. Snow RW, Omumbo JA, Lowe B, Molyneux CS, Obiero JO, Palmer A, et al. Relation between severe malaria morbidity in children and level of

- Plasmodium falciparum* transmission in Africa. *Lancet*. 1997 Jun 7;349(9066):1650-4.
174. Snow RW, Nahlen B, Palmer A, Donnelly CA, Gupta S, Marsh K. Risk of severe malaria among African infants: direct evidence of clinical protection during early infancy. *J Infect Dis*. 1998 Mar;177(3):819-22.
175. Levine N, Harper P. Malaria in other insect-borne diseases in the South Pacific Campaign, 1942-1945 Part IV. *Am J Trop Med* 1947;27(supplement 3):119-28.
176. Maitland K, Williams TN, Bennett S, Newbold CI, Peto TE, Viji J, et al. The interaction between *Plasmodium falciparum* and *P. vivax* in children on Espiritu Santo island, Vanuatu. *Trans R Soc Trop Med Hyg*. 1996 Nov-Dec;90(6):614-20.
177. Gunewardena DM, Carter R, Mendis KN. Patterns of acquired anti-malarial immunity in Sri Lanka. *Mem Inst Oswaldo Cruz*. 1994;89 Suppl 2:63-5.
178. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001 May 31;411(6837):603-6.
179. Carter R, Mendis KN. Evolutionary and historical aspects of the burden of malaria. *Clin Microbiol Rev*. 2002 Oct;15(4):564-94.
180. Allison AC. The discovery of resistance to malaria of sickle-cell heterozygotes. *Biochem Mol Biol Educ*. 2002;30:279-87.
181. Wambua S, Mwangi TW, Kortok M, Uyoga SM, Macharia AW, Mwacharo JK, et al. The effect of alpha+-thalassaemia on the incidence of malaria and other diseases in children living on the coast of Kenya. *PLoS Med*. 2006 May;3(5):e158.
182. de Saint-Vis B, Vincent J, Vandenabeele S, Vanbervliet B, Pin JJ, Ait-Yahia S, et al. A novel lysosome-associated membrane glycoprotein, DC-LAMP, induced upon DC maturation, is transiently expressed in MHC class II compartment. *Immunity*. 1998 Sep;9(3):325-36.
183. Kroll J, Shi X, Caprioli A, Liu HH, Waskow C, Lin KM, et al. The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal center formation. *Mol Cell Biol*. 2005 Oct;25(19):8531-40.

184. Chang NS, Leu RW, Rummage JA, Anderson JK, Mole JE. Regulation of complement functional efficiency by histidine-rich glycoprotein. *Blood*. 1992 Jun 1;79(11):2973-80.
185. Blank M, Shoenfeld Y. Histidine-Rich Glycoprotein Modulation of Immune/Autoimmune, Vascular, and Coagulation Systems. *Clin Rev Allergy Immunol*. 2008 Jan 25.
186. Cattoretti G, Chang CC, Cechova K, Zhang J, Ye BH, Falini B, et al. BCL-6 protein is expressed in germinal-center B cells. *Blood*. 1995 Jul 1;86(1):45-53.
187. Chitnis CE, Miller LH. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. *J Exp Med*. 1994 Aug 1;180(2):497-506.
188. Ward RH, Gershowitz H, Layrisse M, Neel JV. The genetic structure of a tribal population, the Yanomama Indians XI. Gene frequencies for 10 blood groups and the ABH-Le secretor traits in the Yanomama and their neighbors; the uniqueness of the tribe. *Am J Hum Genet*. 1975 Jan;27(1):1-30.



1. Distribution of residuals of the clinical and parasite density phenotypes

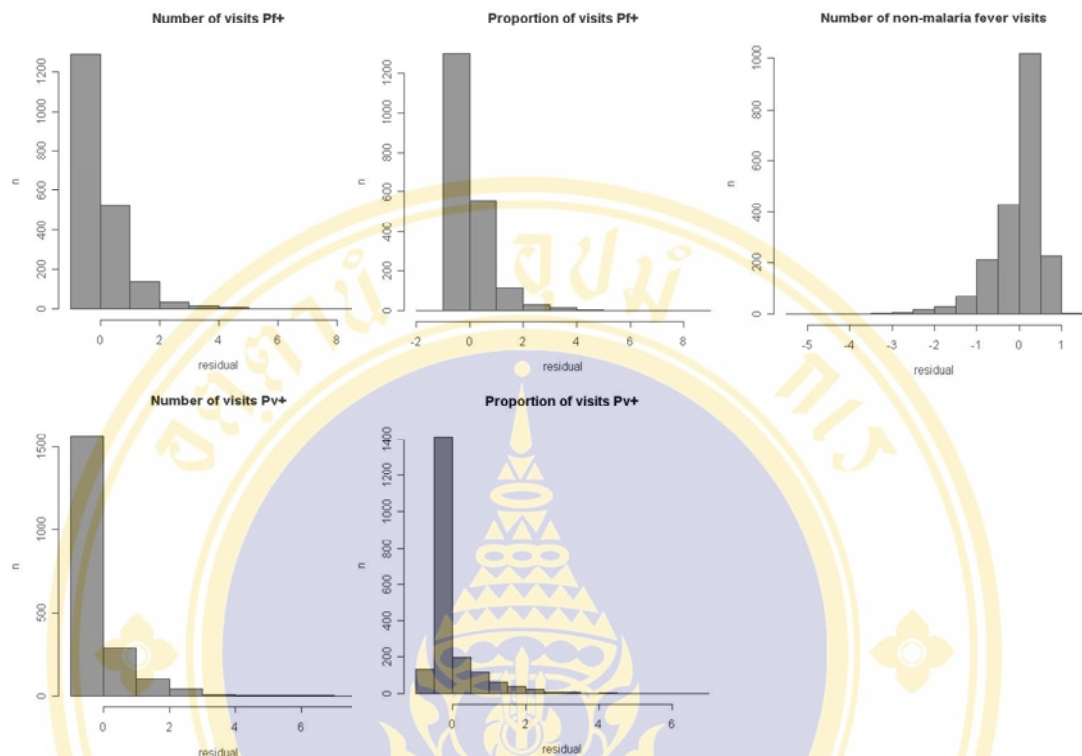


Figure 15. The distributions of residuals of the clinical phenotypes

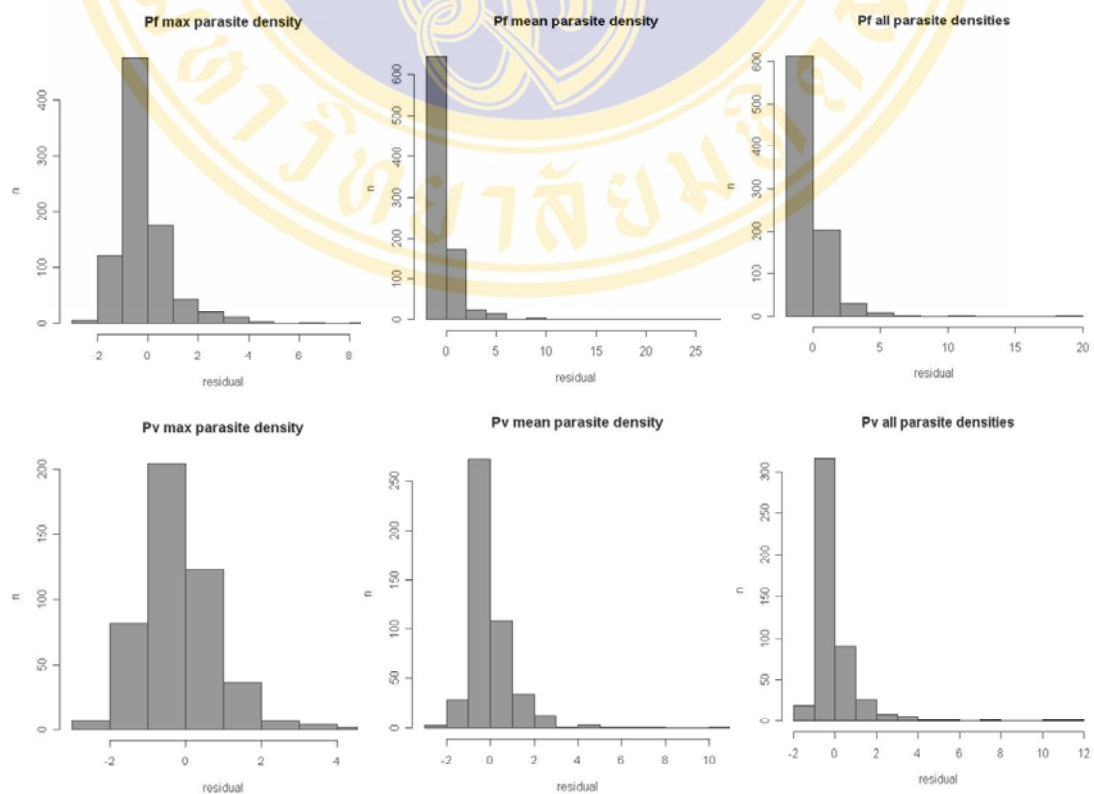


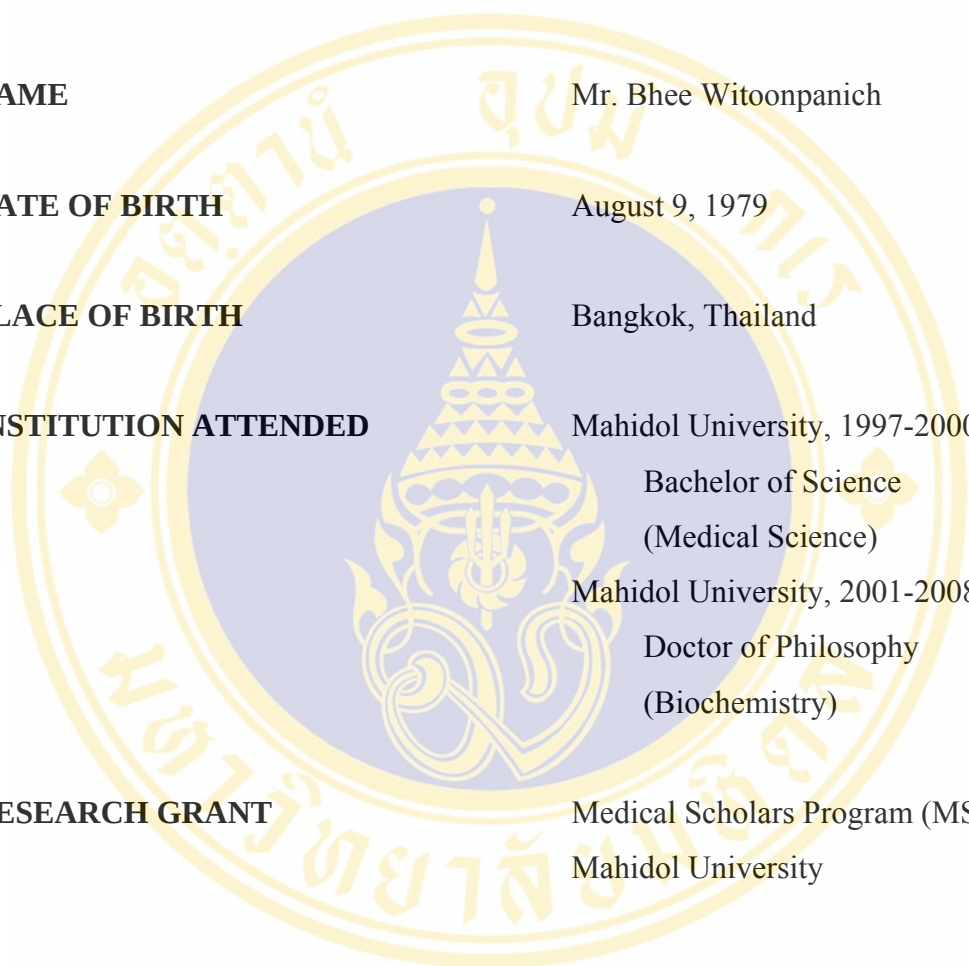
Figure 16. The distributions of residuals of the parasite biological phenotypes

2. Contribution of the house effect to clinical and parasite density phenotypes

Table 14. Contribution (%) of genetic (heritability, h^2) and house (c^2) effects to variability in malaria and non-malaria clinical and biological phenotypes. S.E. is the standard error. NS (Not significant) when p -value >0.05 .

Disease Phenotypes	h^2	S.E.	p -value	c^2	S.E.	p -value
Clinical visits						
Proportion of visits Pf+	3.9	4.8	NS	10.7	2.7	1.35×10^{-5}
Proportion of visits Pf gam+	0	-	NS	5.6	2.9	0.027
Number of visits Pf+	5	4.9	NS	9.8	2.7	4.95×10^{-5}
Number of non-malaria fever visits	4	4.8	NS	10.1	2.8	5.02×10^{-5}
Proportion of visits Pv+	16.6	5.6	0.0009	2.6	2.8	NS
Proportion of visits Pv gam+	11.3	10.5	NS	0.0	-	-
Number of visits Pv+	17.1	5.6	0.0006	2.7	2.9	NS
Parasite biological traits						
Pf mean parasite density	0	-	NS	12.8	4.8	NS
Pf max parasite density	17.4	17.2	NS	3.9	8.8	NS
Pf all parasite densities	0	-	NS	18.5	6.0	0.043
Raw mean Pf gametocyte densities	0	-	NS	35.3	13.9	0.012
Raw maximum Pf gametocyte densities	0	-	NS	31.6	13.8	NS
Pf all gametocyte densities	0	-	NS	6.1	2.4	0.002
Pv mean parasite density	24.7	12.4	0.0153	0.0	-	-
Pv max parasite density	3.3	19	NS	11.4	9.8	NS
Pv all parasite densities	20.2	11.9	0.0363	0.0	-	-
Pv mean gametocyte densities	0	-	NS	0.0	-	-
Pv max gametocyte densities	10.3	37.3	NS	6.7	17.7	NS
Pv all gametocyte densities	12.4	12.6	NS	0.0	-	-

BIOGRAPHY



NAME	Mr. Bhee Witoonpanich
DATE OF BIRTH	August 9, 1979
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTION ATTENDED	Mahidol University, 1997-2000 Bachelor of Science (Medical Science) Mahidol University, 2001-2008 Doctor of Philosophy (Biochemistry)
RESEARCH GRANT	Medical Scholars Program (MSP), Mahidol University
HOME ADDRESS	333/54 Soi Thonglo 17, Sukhumvit 55 Rd., Khlongtoey, Bangkok, Thailand 10110 Tel.: 08-6887-7600 E-mail: novartis21@hotmail.com