Final Report

Genomic approach in searching for genes involved in genetic susceptibility to clinical malaria

Principle Investigator	
Anavaj Sakuntabhai	Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University
Co- investigators	**Mahidol University**
Pratap Singhasivanon	Department of Tropical Hygiene, Faculty of Tropical Medicine
Thanyachai Sura	Department of Medicine, Faculty of Medicine Ramathibodi Hospital
Chayanon Peerapittayamongkol	Department of Biochemistry, Faculty of Medicine Siriraj Hospital
Somporn Krasaesub	Research Center, Faculty of Medicine Ramathibodi Hospital
Objoon Trachoo	Department of Medicine, Faculty of Medicine Ramathibodi Hospita
Waraphon Phimpraphi	PhD student, Faculty of Tropical Medicine
Chairat Turbpaiboon	PhD student, Department of Biochemistry, Faculty of Sciences
Sukunya Thamniam	Department of Medicine, Faculty of Medicine Ramathibodi Hospital
Wathanee Chaiyaratna	Research Center, Faculty of Medicine Ramathibodi Hospital
Advisers	**Mahidol University**
Sornchai Looareesuwan	Dean, Faculty of Tropical Medicine
Ahnond Bunyaratvej	Department of Pathology, Faculty of Medicine Ramathibodi Hospital
Suthat Fuchareon	Thalassemia Research Center
Pranee Fuchareon	Thalassemia Research Center
Praphon Wilairat	Department of Biochemistry, Faculty of Sciences

\

Index

Table Index

Introduction

Malaria remains the most important human parasitic disease worldwide, causing over 170 million clinical cases per year, resulting in over a million die. Most of the treatments available have limited efficacy and side effects and the emergence of drug-resistant strains. Vaccines are being developed with varying degree of success. Host genetic factors also play an important role in susceptibility and clinical manifestations of infectious diseases. The mechanisms of natural protective immunity to malaria are not well understood, nor are the pathophysiological mechanisms of the disease.

Most of the studies reported to date were based on case/control study of severe malaria and have tested genes based on biological functions or the distorted allelic distribution in the regions with high endemicity of malaria. Some of the genes, which will be identified through our approaches, may have no or little effect on the protection against severe malaria. Therefore, it is virtually impossible to identify these genes by studies conducted on severe malaria. The familial study that will be performed in this project enable us to test candidate regions, such as those homologous to the regions identified in animal models as well as the whole genome search in a systematic way. This has a potential to identify new genes, which could not be detected through candidate gene approach.

Material and Methods

Populations

This study is based on populations from Suanpung district, Ratchaburi province, located near the Thai-Myanmar border, which is an endemic area for malaria in Thailand. The size of the populations is around 6000 with 2800 individuals have been followed up for a number of malaria attacks, type of malaria (*P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* or mixed infection), blood parasitemia and clinical response to treatment since 1994. In the first 2 years, all individuals in the study have been checked for blood parasitemia monthly regardless of their symptoms. After this period, patients in the study came to the clinic when they developed fever. They were checked for the presence of the parasites by investigators who had experience in interpreting the slides and were then received appropriated treatment.

Family structures were established by interview. DNA were extracted using standard phenol/chloroform extraction from 1231 individuals from EDTA whole blood and 295 individuals from cells obtained from buccal swab. In addition, DNA samples were obtained from 271 capillary heparinized blood.

Whole genome amplification

In order to save our DNA samples, we performed whole genome amplification by mean of primer extension pre-amplification method (PEP) [1]. Although this method is efficient, the pre-amplified products have short-lived.

We are also testing the other 2 new methods for whole genome amplification, REPLI-g (Molecular Staging Inc., USA) and GenomiPhi (Amersham Biosciences, USA). The REPLI-g method is based on Multiple Displacement Amplification (MDA) technology [2]. It carries out an isothermal genome amplification using a uniquely processive DNA polymerase with exonuclease-resistant primers. The high fidelity is made possible in part because of the novel properties of the DNA polymerase mix which is capable to polymerize at least 70kb

without dissociating from the genomic DNA template, therefore, results in a large fragment amplified product. Amersham Biosciences bought the license of this enzyme mix and provide it as a GenomiPhi DNA Amplification kit.

SNP discovery

In order to screen candidate genes, polymorphisms of each gene were identified first by literature review or public database. Polymorphisms (mostly single nucleotide polymorphism or SNP) which have been shown to be associated with severe form of malaria and/or have the effect on expression of the gene or the protein the gene encoded were selected. These polymorphisms will be tested in a limited number of population. Polymorphisms which have frequency more than 5% in our population were subjected for large scale genotyping in the whole population (method below).

The other way to identify SNP is direct sequencing the entire coding sequences and regulatory region of the gene. Currently, we have another similar project to search for genes involved in clinical malaria in Senegalese population. Genome screening linkage analysis is finished. We are now investigating a few candidate genes which are in the region which showed linkage. Screening for SNPs was performed by direct sequencing using Big Dye Terminator sequencing system in a ABI3700. We then analysed the electropherograms using Genalys software developed by Centre National de Genotypage (ref). We also included 32 selected Suanpung population according to their susceptibility to malaria infection during this large scale sequencing. The gene which we investigated by this way is SLC4A1 (AE1). Table 1 lists all SNPs studied, their origin and methods for genotyping.

Table 1 SNPs and mutation studied, origin anfm method for genotyping

Gene or locus	SNP name	Origin	Method for Genotyping
β-globin	SNP1	sequencing	PCR-RFLP/TaqMan
β-globin	SNP2	sequencing	TaqMan
β-globin	SNP3	sequencing	TaqMan
β-globin	HbE	literature	TaqMan
β-globin	IVS-1	literature	TaqMan
β-globin	4bpdel	literature	TaqMan
α-globin	HbCS	literature	TaqMan
ABO	ABO-297	database	TaqMan
Duffy blood group	duffAg	literature	TaqMan
ICAM1	ICAM1-kilifi	literature	PCR-RFLP
ICAM1	ICAM1-1405	database	PCR-RFLP
IL-10	IL-10-819	literature	PCR-RFLP
IL-10	IL-10-1082	literature	PCR-RFLP
TNF-α	TNF-308	literature	TaqMan
TNF-α	TNF-238	literature	TaqMan

Mutation screening

Screening for mutation of those individuals whom hematological investigation suggested abnormalities in globin genes or G6PD were performed by direct sequencing using Big Dye Terminator system in a ABI3700 and analyzed by the Genalys program.

SNP validation

Polymorphisms identified from literature review or public database were validated in 96 or more randomly selected individuals from our population by mean of digestion with restriction enzyme after polymerase chain reaction amplification (PCR-RFLP). Primers used, enzymes used, fragment lenghts and PCR condition for polymorphism are shown in Table 2. Some of them were used for large-scale genotyping.

Table 2 Primers, enzyme, size of PCR products and PCR condition of PCR-RFLP reaction used in the project

SNP name	Forward Primers (5' to 3')	Reverse Primer (5' to 3')	Enzyme	Size Uncut (Allele)	Size Cut (Allele)	Annealing Temp.	Mg Cl_2
ICAM1-kilifi	TGTCCCCCTCAAAAG TCATC	TCATACACCTTCCGG TTGTTC	NlaIII	99+48 (A)	74+48+25 (T)	53	2.5
ICAM1-1405	CTTGAGGGCACCTAC CTCTG	AGGATACAACAGGCG GTGAG	*BstUI*	154 (A)	101+53 (G)	60	1.5
IL-10-819	TCAACTTCTTCCACC CCATC	AGTGAGCAAACTGAG GCACAGACA	*NlaIII*	206 (T)	180+26 (C)	60	1.5
IL-10-1082	ACACTACTAAGGCTT CCTTGGGA	GATGGGGTGGAAGAA GTTGA	*EcoNI*	126 (G)	106+20 (A)	60	1.5
TNF-308	GAGGCAATAGGTTTT GAGGGCCAT	GGGACACACAAGCAT CAAG	*NcoI* (art)	147 (A)	126 + 21 (G)	63	1 5
TNF-238	AGAAGACCCCCCTCG GAACC	TCTCGGTTTCTTCTC CATCG	*HpaII* (art)	116 (A)	97 + 19 (G)	60	2

art. =artificail site

Large scale SNP typing

After confirmation of polymorphisms in our population, the SNPs were then studied in the whole population by mean of 3 methods listed below according to their priority.

1. TaqMan™ assay

This technique is developed by Applied Biosystems (Foster City, USA) (www.appliedbiosystems.com). It has been used to detect amplified product in real time PCR. They were then later develop this technique for allelic discrimination or SNP typing. For allelic discrimination using TaqMan™ assay, there were 2 TaqMan probes which are different in dye colours and different in nucleotide at the polymorphic site of the SNP. We performed this assay through Assay-by-Design™ service that designs, synthesizes primers and probes for SNP genotyping.

We have modified some parts of the protocol in order to minimize the cost of genotyping. The protocol is as followed. Genomic DNA (1 ng) was dried in optical 96-well reaction plates. A 5 μl reaction mix containing 1x TaqMan® Universal PCR Master Mix and 1x primers and probes Mix (Assay-by-Design) was added to each well. Amplification was performed using ABI Prism 7000 Sequence Detection System. The reaction mixture was

heated at 95°C for 10 min to activate the modified DNA polymerase, followed by 40 cycles of denaturation, 15 sec. at 92°C, and annealing/extension 1 min. at 60°C. Endpoint fluorescence measurements were done during a 1 min. incubation at 60°C, and analysis was performed by the ABI Prism 7000 SDS software.

Primers and probes used for genotyping of our SNPs is shown in Table 3. We found that this technique is the most reliable, efficient, fast and economical method for SNP typing at the moment.

Table 3 Primers and probes used in TaqMan assay

SNP name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	TaqMan probe – FAM (Allele)	TaqMan probe- VIC (Allele)
TNF-308	GAAATGGAGGCAATAG GTTTTGAG	GTAGGACCCTGGAGGC TGAAC	CCGTCCTCATGCC	CCGTCCCCATGCC
TNF-238	TCAGTCAGTGGCCCAG AAGAC	CCCTCACACTCCCCAT CCT	CCCTGCTCTGATTC	CTGCTCCGATTCC
HbE	GCAAGGTGAACGTGGA TGAAG	GGTCTCCTTAAACCTG TCTTGTAACC	TGGTGGTAAGGCC	TTGGTGGTGAGGCC
IVS-1	GGTGAACGTGGATGAA GTTGGT	GCCCAGTTTCTATTGG TCTCCTTAA	CTGGGCAGTTTGG	TGGGCAGGTTGGTAT
4bpdel	GCTGGTGGTCTACCCT TGGA	ACAGCATCAGGAGTGG ACAGATC	AGAGGTTGAGTCCTTT	CCAGAGGTTCTTTG
HbCS	TGGCTTCTGTGAGCAC CGT	CCATCGGGCAGGAGGA A	AGCTTGACGGTATTT	CAGCTTAACGGTATTT
ABO-297	TGGCTGGCTCCCATTG TC	CCTGAACTGCTCGTTG AGGAT	CGATGTTGAATGTGC	CGATGTTGAACGTGC
Duffy	CTGATGGCCCTCATTA GTCCTT	GCTGGGACGGCTGTCA	CCAAGGTAAGAGCC	CTTCCAAGATAAGAGCC

1. PCR-RFLP

Assay-by-Design™ service is a service that guarantee the success of primers and probes they design. There were some cases that primers and probes cannot be designed or passed the quality control of this service. The 2nd method of choice is PCR-RFLP. This method is not that efficient when we perform large scale genotyping, however, it is more economical (depending on the price of the enzyme) and reliable than other methods available. If the SNP does not change restriction site, artificial site will be created during PCR amplification.

2. SNaPshot™

This technique is developed by Applied Biosystems (Foster City, USA) based on primer extension (PE) assay. The reaction is based on annealing a detection primer to the nucleic acid sequence immediately 3' of the nucleotide position to be analyzed and to extend this primer with a single labeled dideoxynucleotide that is complementary to the nucleotide to be detected using DNA polymerase [3-5]. Each dideoxynucleotide is labeled with 4 different fluorescent dyes to indicate the SNP allele, which can be then scored by electrophoresis on a fluorescence-based DNA sequencer. Each SNP is differentiated by size of the primers used in PE reaction, which results in different size of PE product. This technique is the last choice for us because it is very expensive when performed as simplex and the results can be difficult to interpret. We have not used this method for SNP typing in our study.

Bioinformatics and Statistics

Transformation of phenotypes

Because genetic statistics that we will use for linkage and association study assume normal distribution of the phenotypes. We, therefore, transformed the phenotypes by controlling for other confounding factors using multivariate regression analysis. The residual phenotypes were then estimated for each individual. The transformed phenotypes were tested for normal distribution by estimation of skewness and kurtosis and Shapiro-Wilk test for normality. All the phenotypes analyses were performed with STATA version 7.

Errors detection

Usually, SNP typing has some errors. Errors in genotyping is the major cause of false negative and false positive which makes analysis unreliable. We have 2 steps to detect genotyping errors.

1. Mendelian inheritance We first checked for inconsistency for Mendelian inheritance in the family of each marker using PedCheck program [6].

2. Haplotype analysis and recombination detection In principle, no recombination between SNP within the gene should not be detected in the population. We therefore performed haplotype analysis using modified Simwalk2 program [7]. Recombination between SNPs within the same gene indicate genotyping errors.

The results of genotyping of those individuals which whom the program indicate errors were re-interpreted. If controversy still exists, re-genotyping of those individuals were performed from stock DNA.

Haplotype analysis

Haplotypes were estimated by Simwalk2 [7]. Linkage disequilibrium coefficient and association study of SNPs were estimated by GOLD [8].

Analyses of heritability

Genetic contribution to the phenotypes were performed by estimation of the heritability, using variance-component model in which 2 alternative variance models were compared [9]. The significance of a polygenic component in the heritability of each phenotype was examined by the comparison of the likelihood of a null model that included only environmental variance, Ve, with the likelihood of a full model that included both environmental and polygenic variance, Vg. Twice the difference in $\log_e$ likelihood of the two models yields a test statistic that is asymptotically distributed as a ½:½ mixture of a χ^2_1 variable and a point mass of zero. When the null hypothesis was rejected, heritability, *h2*, was then estimated as Vg/(Ve+Vg). Analyses were performed with the SOLAR program (version 1.7.3; for download binaries, see www.sfbr.org/sfbr/public/software/solar/).

Linkage studies

The pedigree-based variance-component linkage analysis was used to estimate the genetic variance attributable to the region around a specific genetic marker. [9]. This approach is based on specify the expected genetic covariance between arbitrary relatives as a function of IBD relationships at a quantitative-trait locus (QTL). IBD status for each locus was estimated for the linkage studies as a multipoint fashion using Markov Chain Monte Carlo methods by use of LOKI (version 2.4.5; for download binaries, see

www.loki.homeunix.net/, [10]. Linkage analyses were run by use of the SOLAR program. For each phenotype, we tested the null hypothesis that the genetic variance due to 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 QTL is estimated. Environmental and residual genetic variance were included in both models. The difference between the two $\log_{10}$ likelihood produces a LOD score that is equivalent of the classical LOD score of linkage analysis. Twice the difference in $\log_e$ likelihood of the two models yields a test statistic that is asymptotically distributed as a ½:½ mixture of a χ^2_1 variable and a point mass of zero.

Association studies

Transmission disequilibrium test (TDT) was estimated in the presence of linkage by use of variance components with orthogonal model described by Abecasis [11]. The orthogonal model extends the approach proposed by Fulker in 1999 [12]. The Fulker model involves maximum likelihood modeling of the sib ship data. Linkage is modeled in the co-variances structure while association parameter are modeled on the mean. The algorithm partitions the gene effect to be between-family (b) and within-family (w) components. A test of within-family association parameter would yield a test for association while controlling for stratification. Abecasis extends this approach to create the orthogonal model that was designed to accommodate any number of offspring and optionally to include parental genotypes if available. Twice the natural log of the likelihood data based on the 2 models under the full model and null hypothesis is asymptotic distributed as a $\chi 2$ statistic, with df equal to the number of parameters being tested. In addition, empirical significance levels were calculated from 1000 Monte Carlo permutations. LOKI was used to calculate IBD for association studies.

Results

Phenotypes and Estimation of Heritability

Phenotypes related to clinical malaria attacks during year 1998-2002 were investigated for genetic effects. The criteria for diagnosis of clinical malaria attacks were fever and fever associated symptoms (headache, vomiting, subjective sensation of fever) with the presence of parasites in blood smear. Species of the causative parasites were specified.

During 1998-2002, there were 2713 clinical malaria attacks from 1443 observed individuals. Of which, 1643 attacks were due to *Plasmodium falciparum*, 849 attacks due to *P. vivax*. Prevalence of clinical attacks for each year are shown in Table 4. Number of clinical attacks per individuals ranged from 1 to 16 for *P. falciparum*, 1 to 14 for *P. vivax* during 5 years of observation. There were significant differences of number of clinical malaria attacks of *P. falciparum* with age group, sex, and hamlet whereas clinical *vivax* malaria attacks was associated with age and year of study.

Table 4 Prevalence of clinical malaria attacks during year 1998-2002

Phenotypes	1998	1999	2000	2001	2002	Total
All clinical attacks	517 (1-12)	850 (1-8)	516 (1-4)	530 (1-7)	300 (1-5)	2713 (1-16)
Clinical falciparum attacks	358 (1-12)	424 (1-8)	336 (1-4)	335 (1-7)	190 (1-5)	1643 (1-14)
Clinical vivax attacks	146 (1-5)	241 (1-5)	164 (1-4)	189(1-5)	109 (1-5)	849 (1-5)

The prevalence of malaria clinical attacks during 5 years was analysed by Poisson regression models including variables for the effect of i) age, ii) sex, iii) hamlet and iv) year of survey (dummy variable: 1998 to 2002).

The expected number of malaria attacks for each individual was estimated according to these models. For each individual i, an Anscombe residual [13] was calculated as a function of the sum of the recorded malaria attacks y_i and the sum of the expected malaria attacks μ_i, the sums being calculated over all the observed periods of the individual that were eligible for the analysis. The Anscombe residual is $a_i = 1.5\ (y_i^{2/3} - \mu_i^{2/3}) / \mu_i^{1/3}$. Among the residuals for Poisson regression models, it is the closest to normality, then standardized to mean zero and variance 1 [13]. The residual number of clinical attacks were then used for further genetic analysis. After transformation, the data was tested for normality (Table 5).

Table 5 Mean, Standard Deviation (SD), Skewness and Kurtosis, Shapiro-Wilk test for normality of the transformed phenotypes

Phenotypes	Mean	SD	Skewness	Kurtosis	Shapiro-Wilk test			
					W	V	Z	P value
All clinical attacks	-0.157	1.237	1.421	6.089	0.905	84.012	11.134	$< 10^{-5}$
Clinical falciparum attacks	-0.419	1.166	1.173	4.835	0.919	70.892	10.707	$< 10^{-5}$
Clinical vivax attacks	-0.392	1.065	1.638	6.276	0.859	124.410	12.120	$< 10^{-5}$

In order to estimate the heritability, we performed a test using variance-component model in which 2 alternative variance models were compared [9]. The results of genetic effect were shown in Table 6.There were strong genetic effects for number of clinical malaria attacks of *P. falciparum* both analysis as quantitative trait (h^2=0.24, $p < 10^{-7}$) and as qualitative trait (h^2=0.34, $p < 10^{-7}$). However, there was no genetic effects for number of clinical malaria attacks for *P. vivax* when analyzing as quantitative trait but show nearly significance when analyzing as qualitative trait (h^2=0.07, p =0.06). These phenotypes will be used for linkage and association study for both candidate genes approach and systematic genome screening.

These evidences supported the previous findings of genetic effects in SriLakan populations [14] that there is a genetic effect for the intensity of clinical *falciparum* attacks whereas genetic effect for clinical *vivax* malaria attacks confer absolute susceptibility or refractoriness to the infection.

Table 6 Estimation of heritablility of the phenotypes

Phenotypes	P value	Heritability	Standard Error
P. falciparum-qualitative	$< 10^{-7}$	0.34	0.06
P. vivax-qualitative	0.06	0.07	0.03
All clinical attacks	$< 10^{-7}$	0.36	0.05
Clinical falciparum attacks	$< 10^{-7}$	0.24	0.05
Clinical vivax attacks	NS	NE	NE

NS = not significant, NE = not estimated.

Whole genome amplification

Apart from 1231 individuals whom we obtained their DNA samples from EDTA whole blood, there were 295 individuals, mostly young children from whom we collected cells from buccal swab. In addition, we obtained more DNA samples from 271 individuals using capillary heparinized blood. Therefore, quantity and quality of DNA samples from those individuals will be less than those whom we have EDTA whole blood. In order to screen the whole genome of these individuals, we tested several methods for amplification the whole genome and then used for multiplex microsatellite typing systems that were performed at the Centre National de Genotypage.

There were 3 methods tested, 1) Primer Extension Pre-amplification [1] Method developed by Molecular Staging Inc. (REPLI-g™) and 3) GenomiPhi™ DNA amplification kit (Amersham Biosciences).

We compared the 3 methods with 3 types of DNA samples, 1) extracted from ETDA whole blood by phenol/chloroform, 2) extracted from buccal swab cells using phenol/chloroform method and 3) extracted from capillary heparinized blood using phenol/chloroform method. There were 8 samples from each type of DNA.

We tested the amplified DNA samples using 4 PCR reactions for the 2 methods, REPLI-g and GenomiPhi DNA amplification kit, and multiplex microsatellite typing system used at CNG for all 3. The amplified DNA samples from EDTA whole blood performed well with no difference detected with the 2 methods when tested with 4 PCR reactions. DNA samples extracted from buccal swab performed poorly with all 3 methods whereas GenomiPhi™ DNA amplification kit performed better than REPLI-g™ when using DNA samples extracted from capillary heparinized blood. However, PEP is the best method for multiplex microsatellite typing.

The problem of PEP method is short-lived (3 months) of amplified DNA samples and low-yield. Currently, we are working with Dr Ivo Gut at CNG to improve the yield and storage of amplified DNA samples.

Study of β-globin locus

β-thalassemia trait is the most common type of hemoglobinopathies in the population with prevalence around 10%. By direct sequencing 8 unrelated individuals who showed abnormal hemoglobin typing results, 3 types of known mutations of β-thalassemia have been identified i.e. 1) splice site mutation of intron 1 (nt1, G>T), 2) 4 bp-deletion of codon 41/42 and 3) HbE (Glu26Lys). Primers used for direct sequencing the entire β-globin gene is shown in Table 7.

Table 7 Primers used for direct sequencing β-globin gene.

Region amplified	Forward primer (5' to 3')	Reverse primer (5' to 3')	Annealing temp.	Product size (bp)
Exon1	ACTCCTAAGCCAGTGCCAGA	CAGCATCAGGAGTGGACAGA	60	506
Exon2	GCACTGACTCTCTCTGCCT	AACGATCCTGAGACTTCCACA	60	411
Exon3	GAGTCCAAGCTAGGCCCTTT	TTTGCAGCCTCACCTTCTTT	60	450

Intron2	GACCAAATCAGGGTAATTTTGCAT	AGTGATACTTGTGGGCCAGG	60	508

Apart from β-thalassemia trait, hereditary persistent of fetal hemoglobin (HPFH) were found in 90 individuals in whom hemoglobin typing were done. Percentage of HbF ranged from more than 1 % to 32.9 %. We have identified 3 individuals with 29-30% of HbF carried 27 kb deletion of the South East Asia type of HPFH. This finding has been published in Hemoglobin 2003 [15]. Primers used for gap PCR is shown in Table 8.

Table 8 Primers used in Gap-PCR

Name of Deletion	Forward primer-1 (5' to 3')	Reverse primer (5' to 3')	Forward primer-2 (5' to 3')	Annealing temp	Normal product size (bp)	Deleted prodcut size (bp)
SEA HPFH	TGGTATCTGCAG CAGTTGCC	AGCCTCATGGT AGCAGAATC	ATTGTTGAGTTG CAGGATCG	58	565	376

Table 9 Allele frequency and Association study for Linkage Disequilibrium of 3 SNPs and 3 mutations studied in β-globin locus.

Name of SNP	Allele frequency	SNP1	SNP2	SNP3	HbE	IVS1	4bpdel
SNP1	0.06		0.936	0.301	10^{-4}	$<10^{-5}$	0.37
SNP2	0.49	0.000		$<10^{-5}$	0.74	0.0007	0.76
SNP3	0.49	0.003	0.609		0.65	0.003	0.477
HbE	0.004	0.111	0.000	0.001		0.586	0.786
IVS1	0.03	0.144	0.023	0.017	0.000		0.59
4bpdel	0.01	0.001	0.000	0.001	0.000	0.000	

p value is shown at the right upper half of the table.
delta^2 for linkage disequilibrium is shown at the left lower half

In addition to the 3 mutations of β-globin gene that have been identified through individuals whom hemoglobin typing suggested β-thalassemia trait, we studied 3 other polymorphisms in this region. These polymorphisms were identified during direct sequencing of the entire β-globin gene with high frequency. Minor allele frequency for the 3 SNPs and 3 mutations and their linkage disequilibrium is shown in Table 9.

In addition to association study, linkage analysis was performed for each SNP and each phenotype by mean of variance component using the QTDT program. The results of association study and linkage analysis are shown in Table 10. The highest association is found between SNP2 with residual number of clinical *falciparum* attacks (p = 0.005). SNP3 which is in tightly linkage disequilibrium with SNP2 also showed marginally significant (p = 0.042). Frequencies of HbE and 4bp deletion mutation of β-thalassemia are too low for statistical analysis. The most common mutation of β-thalassemia found in this population, IVS1, showed no significant association with all the phenotypes tested. Combined analysis of 4bp deletion and IVS1 did not show significant association either. However, there was a significant association of SNP1 with residual number of clinical *falciparum* malaria attacks (p

= 0.032). This SNP also showed association with IVS1 mutation. Linkage analysis did not reveal significant linkage of all these SNPs with all phenotypes studied.

Table 10 Association study and linkage analysis of SNPs

Name of gene or locus	Name of SNP	Allele Freq.	All clinical attacks		Clinical falciparum attacks		Clinical vivax attacks	
			Asscoiation study	Linkage analysis	Asscoiation study	Linkage analysis	Asscoiation study	Linkage analysis
β-globin	SNP1	0.06	NS	NS	0.032	NS	NS	NS
β-globin	SNP2	0.49	NS	NS	0.005	NS	NS	NS
β-globin	SNP3	0.49	NS	NS	0.042	NS	NS	NS
β-globin	HbE	0.004	NT	NT	NT	NT	NT	NT
β-globin	IVS1	0.03	NS	NS	NS	NS	NS	NS
β-globin	4bpdel	0.01	NT	NT	NT	NT	NT	NT
β-globin	IVS1+4bpdel	0.04	NS	NS	NS	NS	NS	NS
ABO	ABO297	0.38	NS	NS	0.051	NS	NS	NS
ICAM1	ICAM1-1405	0.30	NS	NS	NS	NS	NS	NS
IL-10	IL10-819	0.36	NS	NS	NS	NS	NS	NS
IL-10	IL10-1082	0.03	NS	NS	NS	NS	NS	NS

NS = not significant, NT = not tested.

Study of α-globin locus

Hemoglobin constant spring was studied in the whole population by mean of TaqMan assay. The prevalence is 0.004 in Suanpung population which is too low for association study.

For genotyping of α-globin genes deletion, we directly sequenced PCR product amplified from homozygote of α-thalassemia in order to identify the break point of the α-globin gene deletion. The breakpoint was identified for α-thal1 SEA type. We are designing a gap PCR reaction for genotyping this deletion in the whole population.

Study of G6PD locus

G6PD deficiency was detected in 15% of the populations. None of the common G6PD mutations in Thailand; G6PD Viangchan (291 Val>Met) [16]; G6PD Mahidol (163 Gly>Ser) [17] and G6PD Canton (459 Arg>Leu) [18] have been detected in the populations. We also directly sequenced the entire coding regions of G6PD in 12 unrelated individuals and their mother but have not found any causative mutation. Primers used for direct sequencing the entire coding regions of G6PD is shown in Table 11. We are waiting for the results of genome screening that are being performed at CNG. Genome screening linkage analysis will be performed to localize the area where the gene responsible G6PD in this popualtion resides. New gene or new mutation of the G6PD gene is expected to be identified.

Table 11 Primers used for direct sequencing the entire coding regions of G6PD

Region amplified	Forward primer (5' to 3')	Reverse primer (5' to 3')	Annealing temp.	Product size (bp)
Exon2	GCCGTTCACAAGGAGTGATT	CAGGCACTTCCTGGCTTTTA	50	307
Exon3-4	GCTTGTGGCCCAGTAGTGAT	AGGAGAGGAGGAGAGCATCC	60	471
Exon5	TCAAAGAGAGGGGCTGACAT	GTTTCGTGGAGCAACGCT	60	431
Exon6-7	TGCAGCTGTGATCCTCACTC	CTGCAGGGTGACTGGCTCT	60	591
Exon8	GGAAGTGAGTCTTGCAGCTTG	GGTTGAGGACACCTGCTCTG	58	308
Exon9	CCTGAGGGCTGCACATCT	GACCAGTGCGTGAGTGTCTC	55	368
Exon10-11	ACTGGAGCTCCCACTGAGAC	ACCCCATAGCCCACAGGTAT	60	544
Exon12	GGCCTCCCAAGCCATACTA	CCACTTGTAGGTGCCCTCAT	60	290
Exon13-1	TTATGGCAGGTGAGGAAAGG	CACAGGCAGATTCTCTCACG	55	600
Exon13-2	CAAGCACTCGAGACCATCCT	GGGGTCAGAACCAGAAGTGA	55	549

Study of SLC4A1 (AE1) polymorphisms

SLC4A1 or AE1 or band 3 is the major glycoprotein of the erythrocyte membrane and mediates exchange of chloride and bicarbonate across the phospholipid bilayer. Senescent cell antigen (SCA), an aging antigen, is an epitope that appears on old cells and marks them for removal by the immune system. The aging antigen is generated by the clustering of protein band 3. Besides its role in the removal of senescent and damaged cells, SCA also appears to be involved in the removal of erythrocytes in hemolytic anemia and the removal of malaria-infected erythrocytes. In addition, deletion of amino acid residues 400-408 of SLC4A1 results in Melanesian ovalocytosis [19]. Ovalocytic erythrocytes from Melanesians are resistant to invasion by malaria parasites. In our Senegalese populations, genome screening linkage analysis revealed linkage of number of clinical *falciparum* attacks to chromosome 17q21-22 (unpublished data) which is an area that SLC4A1 resides. Therefore, we performed direct sequencing of the entire coding region and its promoter including intron 3 which is the region that regulate expression of the kidney isoform of the genes. Primers used for direct sequencing are shown in Table 12.

Table 12 Primers used for direct sequencing the entire SLC4A1

Region amplified	Forward primer (5' to 3')	Reverse primer (5' to 3')	Annealing temp.	Product size (bp)
Promotor	GTGAATGGTCTTGCAGTGGC	AAGAGCTGGCTCCTGGACAC	60°C	592
Promotor	GAGCTGACATTGTTTCAGGT	CAGGGTCCCTTGGTGAAGT	60°C	692
Exon1	AGCTGTCCAGATGTGGGTAA	CATACCATCTGCTGCCATTG	60°C	474
Exon2 and 3	GGGCAGCAGCTATTTCTGAGAG	AACTTTAATCCCAATCTCCAGCAC	60°C	1229

Exon4 and 5	GGGAATATAAGGGGCTGACC	CCTCTATCCCCTTGCTCCTC	60°C	615
Exon6	TGGGAGATAAGGGAGTGGTG	CTAGCAGTTGGTTGGCCACT	60°C	488
Exon7	GCCTCCTAGAGCTGCGTAGA	GAGATGGGAGCCATAGTGGA	60°C	407
Exon8	TCTACCCCAGTCCCTTGATG	CTGCTTGTGGTCGGTTTTTC	60°C	319
Exon9 and 10	AAAAACCAACCACAAGCAGG	GCCAGGTAGGATAGCAGCAG	60°C	713
Exon11	CTGCTGCTATCCTACCTGGC	ATGTGATGGGAGACAGAGGC	60°C	707
Exon12	CCCATTCCCATCAGACAATC	TCATTTCCAGGAGCCCATAG	60°C	434
Exon13	TCTATGGGCTCCTGGAAATG	CTGGGTATAGCGGGAGATGA	60°C	454
Exon14	TGCTGGTGTTTGAGGAAGC	AACCTCCCGTGTGCATTAAC	60°C	376
Exon15	GTGGATGGATGGGTAGATGG	GGAATTGGGAATGGGAATCT	60°C	519
Exon16	TTAGATGCTGATGGATCCCC	GTAGTCCCAGCTGGCTTCAG	60°C	322
Exon17	CCAAGTGCCTCCAACCTAAC	CTAGTCGGGAGGGCCACAC	60°C	405
Exon18	GCTACAAGGACACCAAGTATGGAG	AGAAGGCCTCGGAGTGGAG	60°C	575
Exon19 and 20	GCAACCTGGGCTGAGAGTG	CATGCTCCCAGCTCTTGTG	60°C	508
Exon21	TCCACAGGGTGACTCAGGTC	TGGAGTTGAGGATAATGGCTCTC	60°C	1481

We identified 20 SNPs (Table 13), 6 in the coding region, 2 change amino acid and 4 are synonymous. One SNP is at the 5 UTR, 2 at 3 UTR and 10 are in intron. Ten (50 %) are newly identified SNPs which were not found in the public SNA database. Table 13 showed p value of association study between SNPs identified (linkage disequilibrium) in Suanpung populations. 10 SNPs form haplotype block. At the moment, we are preparing large scale SNP typing for the whole populations.

Table 13 Position, Predicted Consequence, Allele Frequency and p-value from Association study for Linkage Disequilibrium of SNPs identified in SLC4A1 (AE1) gene

SNP* identified	Position or consequence	Allele Freq.	AE1_2	AE1_3	AE1_4	AE1_5	AE1_6	AE1_7	AE1_8	AE1_9	AE1_10	AE1_11	AE1_12	AE1_13	AE1_14	AE1_15	AE1_16	AE1_17	AE1_18	AE1_19	AE1_20
AE1_1	In 3	0.071	0	1	1	0.001	0.485	0.438	0.68	0.653	0.065	0.644	0.651	0.658	0.166	0.008	0.046	0.658	0.438	0.024	0.007
AE1_2	In 3	0.071		1	1	0.001	0.485	0.438	0.68	0.653	0.065	0.644	0.651	0.658	0.166	0.008	0.046	0.658	0.438	0.024	0.007
AE1_3	Ex 4 (5'UTR)	0.375			0	1	0.284	1	0	0.418	1	0.276	1	0.4	0.06	0.247	1	0.106	1	0	0.276
AE1_4	In 4	0.318				0.444	0.507	0.472	0	0.478	1	0.225	0.459	0.472	0.112	0.048	1	0.059	0.472	0	0.378
AE1_5	In 5	0.192					0.327	0.007	0.326	1	1	0.524	1	0.524	0.079	0.01	0.103	0.364	0.007	0.784	0.194
AE1_6	Ex 6 (Leu/Leu)	0.042						0.497	0.822	1	1	0.698	1	0.698	0.286	0.417	0.698	0.581	0.497	0.038	0.001
AE1_7	In 6	0.1							0.364	0.759	0.729	0.729	0.734	0.739	0.298	0.881	0.734	0.636	0	0.109	0.329
AE1_8	In 10	0.25								0.524	0.524	0.144	1	0.524	0.143	0.847	0.524	0.144	0.364	0	0.269
AE1_9	Ex 11 (Ala/Ala)	0.042									0.829	0.829	1	0.829	0.621	0.586	0.829	0.829	0.759	0.358	0.586
AE1_10	Ex 12 (Arg/His)	0.036										0.843	1	0.843	0.589	0.122	0.84	0.778	0.729	0.379	0.097
AE1_11	In 12	0.036											1	0.843	0.589	0.553	0.84	0.778	0.729	0.379	0.589
AE1_12	In 17	0.033												0.846	0.596	0.561	0.843	0.783	0.734	0.39	0.111
AE1_13	Ex 18 (Ala/Ala)	0.033													0.552	0.569	0.846	0.786	0.739	0.36	0.577
AE1_14	In 18	0.133														0.094	0.544	0.583	0.298	0.019	0.081
AE1_15	In 18	0.107															0.097	0.569	0.881	0.074	0.126
AE1_16	Ex 20 (Val/Ile)	0.038																0.783	0.734	0.357	0.092
AE1_17	Ex 21 (Tyr/Tyr)	0.067																	0.636	0.193	0.428
AE1_18	Ex 21 (3'UTR)	0.1																		0.109	0.329
AE1_19	Ex 21 (3'UTR)	0.367																			0.001

Study of ABO blood group

One polymorphism of the gene coding for ABO blood group has been shown to be marginally associated with number of clinical malaria attacks in Senegalese population (unpublished data). This SNP located in exon 6 but do not change amino acid. We found nearly significant association (p=0.051, Table 10) of this SNP with residual number of clinical *falciparum* malaria attacks without linkage.

Study of Duffy blood group

Duffy blood group is an important receptor of *P. vivax* to enter red blood cells[20]. Mutation at the GATA binding site of promoter of the gene coding for Duffy blood group result in null expression of this receptor on red blood cell membrane. This mutation is identified in 100 % of the African populations in consistent with lack of *P. vivax* infection in this region. Therefore, this polymorphism is a strong candidate for susceptibility to *P. vivax* infection. We have screened this polymorphism in populations in Suanpung villages by mean of TaqMan assay. We could not detect this mutation in 92 individuals studied.

However, Duffy blood group is a strong candidate for *P. vivax* susceptibility. There may be another variant of Duffy blood group which affect adhesion and invasion of *P. vivax*. We plan to direct sequence the entire coding and regulatory region of this gene in our population.

Study of TNF α polymorphisms

Two polymorphisms located at the TNF promoter region [21, 22] were studied in Suanpung population. One is the SNP at the position 308 nucleotide 5' to the A of the start ATG codon. This SNP has been reported to be associated with severe form of malaria infection in Africa. The frequency of this polymorphism in Suanpung population was studied using 330 individuals and found to be 0.01. This polymorphism was identified at frequency of 0.05 in Karen who admitted because of severe form of malaria at the Hospital of Tropical Disease, Faculty of Tropical Medicine, Mahidol University. This polymorphism may be implicated in a severe form of the disease in Karen populations as in African populations. However, Karen admitted in the hospital may be genetically different from Karen at Suanpung villages. In addition, the finding is well-known. We, therefore, do not have plan to publish this result.

The other polymorphism of TNF promoter studied is polymorphism at the position of 238 nucleotide 5' to the A of the start ATG codon. This polymorphism has been reported to have an effect on expression of the cytokine. The frequency of this polymorphism is 0.05 in 360 individuals from Suanpung. The frequency is too low to have statistical power for association analysis. We do not plan to genotype this polymorphism in the whole populations.

Study of ICAM1 polymorphisms

One variant of ICAM-1 (ICAM-1 kilifi; ICAM1-179; Lys29Met) has been reported to result in decrease adhesion with parasitized red blood cell however children who are homozygotes of ICAM-1 kilifi showed higher prevalence of cerebral malaria in Kenya [23]. One SNP of ICAM-1 which result in a non-synonymous change (ICAM1-1405; HGBASE;SNP000002435, Glu422Lys) was selected from the public database. Again ICAM-1 kilifi cannot be identified in Suanpung population whereas 0.06 of Karen admitted because of severe malaria at the hospital carried this mutation. We do not have plan to publish this result for the same reason as TNF-308.

The ICAM1-1405 is highly polymorphic in Suanpung population. From Senegalese study, this polymorphism showed marginal association with maximum parasite density of *P. falciparum* during clinical malaria attacks. We therefore, performed association study and linkage analysis with the phenotypes we have. No significant association or linkage has been detected (Table 10).

Study of IL-10 polymorphisms

Two SNPs in the promoter region of IL-10 (IL10-819; C819T and IL10-1082; G1082A) from the literature [24] has been reported to affect expression of the gene were also tested. We studied these 2 polymorphisms in our populations using PCR-RFLP. We could not detect significant association or linkage of the 2 SNPs with any of the phenotypes studied.

Discussion

We have confirmed previous finding of genetic effect on intensity of clinical *falciparum* malaria attacks and absolute susceptibility to clinical *vivax* malaria attacks. We have screened for association of 5 genes (γ-globin, β-globin, ABO, ICAM-1 and IL-10) using 11 SNPs. There are 2 loci that need further investigation. We obtained significant association with β-globin locus with clinical *falciparum* attacks. We also found nearly significant association with ABO blood group with the same phenotype.

For β-globin locus, our findings are surprising. We have confirmed the role of β-globin locus on susceptibility to malaria infection. However, there were no significant association of mutations causing β-thalassemia with number of clinical malaria attacks for both species either when analyzed them separately or combined. However, there are other 20 individuals whom hematological investigation suggested β-thalassemia trait but do not carry these 2 mutations (IVS1 and 4 bp deletion). We plan to sequence these individuals and include them in our analysis of β-thalassemia.

Our present results suggested that β-thalassemia mutations themselves may not play an important role in susceptibility to clinical malaria but it may be the other polymorphism which is in linkage disequilibrium with them. We identified 3 polymorphisms this region which showed association with number of clinical falciparum malaria attacks. Two polymorphisms are in linkage disequilibrium with each other, the other polymorphism is in linkage disequilibirum with the most frequent polymorphism causing β-thalassemia. We hypothesized that β-thalassemia mutations themselves may not be the polymorphism which confer resistance to clinical malaria. It is possible that 2 of the 3 polymorphisms we identified in this region plays an important role. These polymorphisms may modify clinical severity of β-thalassemia which then help those patients survive better. The finding that 1 of these polymorphisms is in linkage disequilibrium with the most common mutation for β-thalassemia in this population supported this hypothesis.

One polymorphism of ABO blood group showed nearly significant association. This polymorphism also showed marginal significant association with clinical *falciparum* attacks in Senegalese population (unpublished data). The findings of association of the same SNPs with the same phenotype in 2 genetically different populations re-enforced the results. Therefore, the ABO locus will be one of the region that will be explored as the SNP we tested is unlikely to have functional effect as it does not change amino acid.

We also investigated polymorphisms of 5 other genes (α-globin, Duffy blood group, SLC4A1, G6PD and TNF-α). We sequenced the whole entire coding sequences of 2 genes

(SLC4A1 and G6PD). We are going to investigate further the role of polymorphisms of SLC4A1. Mutation responsible for G6PD deficiency is still not identified. Results from genome screening linkage analysis will suggest us where the mutation should be in the whole genome. Identification of new mutation of G6PD could be published by its own. Then the role of G6PD could be studied with the phenotypes.

The role of α-globin will be studied after we re-design the gap PCR to detect those common gene deletion. The alternative way to study this locus is to use microsatellites within this region which are already included in the genome screening set. Duffy blood group and TNF-α will also be studied by finding new polymorphisms after we have more evidence for linkage of these loci with genome screening linkage analysis.

Conclusion

By investigating 10 genes, we now identify 2 loci which are worth further investigating their role in genetic susceptibility to clinical malaria. Although there is no evidence for association of other genes with the phenotypes studied, we still cannot exclude the role of those genes unless we fully investigate all the polymorphisms within the gene. However, full study for each gene is very expensive and time consuming. We are waiting for more evidence of the role of those loci from genome screening results before further investigating.

Future plan

Investigation of β-globin locus

The plan to further investigate the role of β-globin locus are

1. Correction of family structure using genome screening linkage analysis results.
2. Correction of genotyping errors for all SNPs in this locus after correction of family structures by
   a. Check for Mendelian inheritance
   b. Check for recombination within the locus using all SNPs and microsatellites in the region.
3. Re-analysis for both association study and linkage analysis.
4. If the results are confirmed, direct sequencing of the entire region will be performed to identify all SNPs.
5. Genotyping of selected SNPs which are predicted to affect expression of the gene or function of the protein will be studied in the entire population.
6. Functional study to investigate the role of these polymorphisms in susceptibility to clinical malaria will be performed both *ex vivo* and *in vitro*. For example, the role of these polymorphisms in expression of γ and β globins during malaria infection. Response of malaria parasites in *in vitro* culture using red blood cells from individuals who are heterozygote and homozygote for these polymorphisms.

There are 2 important parts of the project that we are now performing in order to further investigate genetic susceptibility to clinical malaria. There will be more publications coming because the project is still ongoing but with grants from diffferent sources.

Phenotyping (supported by the National Research Council)

At the moment, we have very good data concerning number of clinical malaria attacks of each species during 1998 to 2002. Although the analysis showed highly significant genetic effect on intensity of *P. falciparum* infection and absolute susceptibility to *P. vivax*, however, there are numbers of other interesting phenotypes related to malaria and other infections which are prevalence in this population. Genotyping of the whole genome is due to finish in August 2003 which will give us genotyping data of the whole genome of the population. We can then perform linkage analysis study to localize area of genes responsible for those phenotypes without performing genome screening again which a very expensive and demanding procedure. We, therefore, plan to collect more phenotypes. This part of the project receiving funding from the National Research Council as a part of budget for research of Mahidol University. The phenotypes that we plan to collect are parasitemia during asymptomatic period (both trophozoites and gametocytes), helminthic infestations, transmission ability, inducible high hemoglobin F.

Genome screening (supported by Centre National de Genotypage)

Genome screening using 400 microsatellites is now performing at CNG. It is due to finish in August 2003. Problems of genome screening are the complexity of the family structure, mismatched samples, inconsistency of the transmission of the marker alleles which can also due to wrong information obtained by interview, mislabeling, data entering errors etc.

At the moment we identified less than 5% of samples which showed errors. Family structure will be validated after genome screening finish by IBS_check program developed by Simon Health at CNG. This program will calculate the actual allele sharing (IBS) between each pair of individuals within family and outside family (unrelated individuals). Z statistic was used to test for significant difference between observed IBS and IBS expected from the relationship defined by interview. In this manner, we can detect those whom we have wrong information, even for the one who should be unrelated. For those who are wrong, the data will be checked. If correct information cannot be obtained because of ethical problem, these individuals will be removed from our analysis.

Correct information and genotyping is the most important part of genetic study. The data presented in this reports analyzed from uncorrected data because genome screening is not finished. The results may be different after correction.

Publication

We have published one article entitled in hemoglobin. The article is attached.

Problems

1. **Family structures** The biggest problem that we have in this project is the complexity of family structures. We have 1 family consisted of 888 individuals. The size of family exceed capacity of most programs used for genetic analysis. There are only 2 programs which have been designed for complex family structure that are available at the moment, Loki and Solar. We have been in close contact of the authors of the 2 programs. There should not be any problem to increase size of family that can be analyzed by the programs. However,

after genome screening and correction of family structure, the size of families studied are expected to be smaller.

2. **SNP database** At the moment, we have to rely on SNP database in the public domain. Information of those SNP mostly came from Caucasian or Japanese. Therefore, it takes time and money to validate those SNP which are in the database in our population. In addition some SNPs in Thai population cannot be found in the public domain. For example half of our SNPs in the SLC4A1 (AE1) gene are newly identified. Using public SNP database we may miss an important SNP in Thai population. The Thailand SNP database project would facilitate the discovery and reduce cost of genotyping.

References

1. Zhang, L., *et al.*, Whole Genome Amplification from a Single Cell: Implications for Genetic Analysis. *Proc. Natl. Acad. Sci. USA*, 1992. **89**: p. 5847-5851.

2. Dean, F., *et al.*, Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA*, 2002. **99**(8): p. 5261-6.

3. Syvanen, A.C., From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat*, 1999. **13**(1): p. 1-10.

4. Sokolov, B.P., Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res*, 1990. **18**(12): p. 3671.

5. Kuppuswamy, M.N., *et al.*, Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes. *Proc Natl Acad Sci U S A*, 1991. **88**(4): p. 1143-7.

6. O'Connell, J. and D. Weeks, PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*, 1998. **63**(1): p. 259-66.

7. Sobel, E. and K. Lange, Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet.*, 1996. **58**(6): p. 1323-37.

8. Abecasis, G. and W. Cookson, GOLD--graphical overview of linkage disequilibrium. *Bioinformatics*, 2000. **16**: p. 182-3.

9. Almasy, L. and J. Blangero, Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*, 1998. **62**: p. 1198-1211.

10. Heath, S., Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet.*, 1997. **61**(3): p. 748-60.

11. Abecasis, G., L. Cardon, and W. Cookson, A general test of association for quantitative traits in nuclear families. *Am J Hum Genet*, 2000. **66**(1): p. 279-92.

12. Fulker, D., *et al.*, Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet.*, 1999. **64**(1): p. 259-67.

13. Cameron, A.C. and P.K. Trivedi, *Regression analysis of count data*. Econometric society monographs ; no. 30. 1998, Cambridge, UK ; New York, NY: Cambridge University Press. xvii, 411.

14. Mackinnon, M.J., *et al.*, Quantifying genetic and nongenetic contributions to malarial infection in a Sri Lankan population. *Proc Natl Acad Sci U S A*, 2000. **97**(23): p. 12661-6.

15. Trachoo, O., *et al.*, Molecular characterization of hereditary persistence of fetal hemoglobin in the Karen people of Thailand. *Hemoglobin*, 2003. **27**(2): p. 97-104.

16. Nuchprayoon, I., S. Sanpavat, and S. Nuchprayoon, Glucose-6-phosphate dehydrogenase (G6PD) mutations in Thailand: G6PD Viangchan (871G>A) is the most common deficiency variant in the Thai population. *Hum Mutat*, 2002. **19**(2): p. 185.

17. Vulliamy, T.J., *et al.*, G6PD mahidol, a common deficient variant in South East Asia is caused by a (163)glycine----serine mutation. *Nucleic Acids Res*, 1989. **17**(14): p. 5868.

18. Stevens, D.J., *et al.*, G6PD Canton a common deficient variant in South East Asia caused by a 459 Arg----Leu mutation. *Nucleic Acids Res*, 1990. **18**(23): p. 7190.

19. Jones, G.L., *et al.*, Human erythrocyte Band-3 has an altered N terminus in malaria-resistant Melanesian ovalocytosis. *Biochim Biophys Acta*, 1990. **1096**(1): p. 33-40.

20. Miller, L.H., *et al.*, The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med*, 1976. **295**(6): p. 302-4.

21. Knight, J.C., *et al.*, A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet*, 1999. **22**(2): p. 145-50.

22. Kwiatkowski, D., *et al.*, TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria. *Lancet*, 1990. **336**(8725): p. 1201-4.

23. Fernandez-Reyes, D., *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. **6**(8): p. 1357-60.

24. Turner, D., *et al.*, An investigation of polymorphism in the interleukin-10 gene promoter. *Eur J Immunogenet*, 1997. **24**(1): p. 1-8.

120021542.HEM27.2.RI.040103

HEMOGLOBIN
Vol. 27, No. 2, pp. 97–104, 2003

ORIGINAL ARTICLE

Molecular Characterization of Hereditary Persistence of Fetal Hemoglobin in the Karen People of Thailand

Objoon Trachoo, M.D.,[1,*] Thanyachai Sura, M.D.,[1,*] Anavaj Sakuntabhai,[3] Pratap Singhasivanon,[4] Srivicha Krudsood,[4] Waraphon Phimpraphi,[5] Somporn Krasaesub,[2] Suporn Chanjarunee,[6] and Sornchai Looareesuwan[5]

[1]Department of Medicine, Division of Medical Genetics and Molecular Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[2]Research Center, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[3]Genetic of Infectious Diseases and Autoimmunes, Institute Pasteur, Paris, France
[4]Department of Tropical Hygiene and [5]Department of Clinical Tropical Medicine, Faculty of Tropical Medicine, Hospital for Tropical Disease, Mahidol University, Bangkok, Thailand
[6]Department of Medicine, Division of Hematology, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

ABSTRACT

Hereditary persistence of fetal hemoglobin (HPFH) is the condition whereby a continuously active γ-globin gene expression leads to elevated fetal hemoglobin (Hb F) levels in adult life [Stamatoyannopoulos G, Grosveld F. Hemoglobin switching. In: Stamatoyannopoulos G, Majerus PW, Perlmutter RM, Varmus H, eds. The

*Correspondence: Objoon Trachoo, M.D. and Thanyachai Sura, M.D., MRCP, Department of Medicine, Division of Medical Genetics and Molecular Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Rama 6 Road, Rajtevi, Bangkok 10400, Thailand; Fax: +66-2201-1374; E-mail: objootra@ccme.or.th; E-mail: ratsr@mahidol.ac.th.

DOI: 10.1081/HEM-120021542

0363-0269 (Print); 1532-432X (Online)
www.dekker.com

Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001:135–182; Wood WG. Hereditary persistence of fetal hemoglobin and $\delta\beta$ thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, eds. Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. Cambridge: Cambridge University Press, 2001:356–388; and Weatherall DJ, Clegg JB. Hereditary persistence of fetal hemoglobin. In: Weatherall DJ, Clegg JB, eds. The Thalassaemia Syndromes. Oxford: Blackwell Scientific Publishers, 1981:450–507]. The condition is caused either by mutation of the β- and γ-globin genes, or the γ-gene controlled region on other chromosomes. Several families with this condition have been reported from Vietnam, Cambodia and China, and the Southeast Asian mutation (or HPFH-6), a 27 kb deletion, was demonstrated. Here we report on a mother and her daughter of the Karen ethnic group with high levels of Hb F, living in the Suan Pueng District on the border of Thailand and Myanmar. Genotyping showed a heterozygosity for the 27 kb deletion of the β-globin gene. Their conditions have been confirmed by gap polymerase chain reaction (PCR) with three oligonucleotide primers recently developed by Xu et al. [Xu X-M, Li Z-Q, Liu Z-Y, Zhong X-L, Zhao Y-Z, Mo Q-H. Molecular characterization and PCR detection of a deletional HPFH: application to rapid prenatal diagnosis for compound heterozygotes of this defect with β-thalassemia in a Chinese family. Am J Hematol 2000; 65:183–188.], and a DNA sequencing method. Thus far there has been no official report of the HPFH-6 anomaly from Thailand. The compound heterozygosity of β-thalassemia (thal) and hereditary persistence of Hb F causes the phenotype of thalassemia intermedia; in contrast, homozygotes for this anomaly show only mild microcytic anemia. Hence, genetic counseling for hereditary persistence of Hb F carriers is needed for family planning.

Key Words: Hereditary persistence of fetal hemoglobin (HPFH); Deletional thalassemias; β-Globin gene cluster; Karen; Thailand.

INTRODUCTION

Hereditary persistence of fetal hemoglobin (HPFH) and deletional thalassemias are conditions characterized by the continuation of γ-globin gene expression in adult life, resulting in elevation of fetal hemoglobin (Hb F) levels (1–3). In normal adults, Hb F can be found in the range of 0–1% without other γ-globin gene mutations (1,5,6). The deletional HPFH mutations and deletional thalassemias, i.e., β-, $\delta\beta$- and $^{A}\gamma\delta\beta$-thalassemias are distinguished by the phenotype of heterozygous individuals and F-cell distribution (1,2,7).

Several types of deletional HPFH mutations have been described. The Southeast Asian (SEA) type of HPFH or HPFH-6 (1,2) has a 27 kb deletion, in which the 5′ breakpoint is located between the δ- and β-globin genes, and the 3′ breakpoint is located approximately 2.3 kb downstream from the 3′HS-1 locus control region (LCR) of the β-globin gene. There are reports of the mutation in individuals and families from Vietnam, Cambodia, and Southern China (4,8–10).

The compound heterozygotes for β-thalassemia (thal) with HPFH express the phenotype of thalassemia intermedia; thus, an individual with heterozygous HPFH as well as β-thal needs genetic counseling for family planning (11–13).

MATERIALS AND METHODS

Screening for α- and β-thal genes of 1237 Karen people was performed as part of a study of the effect of the thalassemia gene on clinical malaria infection at Suan Pueng District, Ratchaburi Province of Thailand. Informed consent was obtained by the National Ethics Committee. Blood samples were collected and were processed for complete blood count (CBC), hemoglobin (Hb) typing and DNA studies. The CBCs were performed by standard hematological techniques with an automated cell counter. Hemoglobin typing was determined by high performance liquid chromatography (HPLC). DNA was extracted by the standard phenol–chloroform method, and further characterized the mutation by polymerase chain reaction (PCR)-based and DNA sequencing methods. The PCR was performed using three oligonucleotide primers, FH-1, FH-4, and FH-5, as described by Xu et al. (4) (Fig. 1). The total 25 μL PCR mixture contained 100 ng of genomic DNA; F1
20 μM of each primer, 2 μM of each dNTP; 0.5 units of Taq DNA polymerase enzyme (GIBCO-BRL, New York, NY), and 2.5 mM of $MgCl_2$ in a 10X PCR buffer (GIBCOBRL®). The PCR started with initial denaturation at 95°C for 5 min, 30 cycles of PCR amplification were performed in a thermocycler (Geneamp® model 2700; Applied BioSystems, Foster City, CA) with denaturation at 95°C for 30 seconds, annealing at 58°C for 1 min and extension at 72°C for 70 seconds. The final extension was at 72°C for 8 min. DNA sequencing was performed in both directions, forward and reverse, by a standard automated method in an ABI PRISM™ 377 sequencer (Perkin Elmer Cetus, Foster City, CA). A PCR for α-thal, as published elsewhere (14), was performed to find the common thalassemia mutations in Thailand, to exclude other causes of red blood cell (RBC) anomalies (15,16).

RESULTS

Hematological Data

High Hb F levels of 29.6 and 32.9%, respectively, were found in a mother (A.K.) and her daughter (B.K.). The MCV value of A.K. was normal and slighly decreased in B.K. Both had normal levels of Hb A_2. Their MCH levels were slightly decreased and the RDW range was higher than in the normal population (Table 1). T1

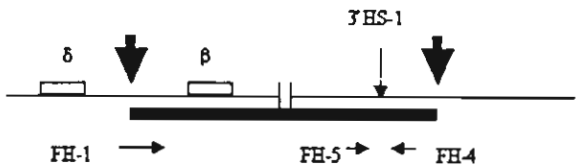

Figure 1. A schematic drawing of the HPFH-6 deletion. Thick arrows indicate the 5′ and 3′ breakpoints, the black bar indicates the deletion region, the thin arrows indicate the direction of the primers, and the dotted arrow indicates the 3′HS-1 site. The primers were FH-1 (5′-TGGTATCTG-CAGCAGTTGCC-3′), FH-4 (5′-AGCCTCATGGTAGCAGAATC-3′) and FH-5 (5′-ATTGTT-GAGTTGCAGGATCG-3′). FH-1 and FH-4 amplify the HPFH-6 deletion region, and FH-5 and FH-4 amplify the normal allele (4).

Table 1. Hematological data and hemoglobin analysis of the Karen family with HPFH-6 from Suan Pueng district, Ratchaburi province, Thailand.

Subject	A.K. (mother)	B.K. (daughter)
Sex–Age	F–31	F–6
Hb (g/dL)	13.3	12.1
PCV (L/L)	0.425	0.374
RBC (10^{12}/L)	5.20	4.85
MCV (fL)	81.7	77.1
MCH (pg)	25.6	24.9
RDW (%)	18.1	17.4
Hb A_2 (%)	3.2	3.1
Hb F (%)	29.6	32.9
α Genotype	αα/αα	αα/αα

Note: Normal laboratory values: Hb 13–16 g/dL for males, 12–15 g/dL for females, 11–15 g/dL for children and pregnant women; RDW 12–15%; MCV 88–99 fL; MCH 27–33 pg; Hb A_2 2.5–3.5%; Hb F in adults 0–1%; the normal α-globin genotype is αα/αα.

Polymerase Chain Reaction Analysis

DNA without a β-globin cluster deletion might reveal one 565 bp fragment specific for the normal allele; whereas, a deletion allele might reveal a 376 bp fragment. The PCR results showed that the normal control had only a single 565 bp fragment, and our subjects had two fragments of 565 and 376 bp, respectively (Fig. 2). These results indicate that A.K. and B.K. were heterozygous for the HPFH-6 mutation. F2

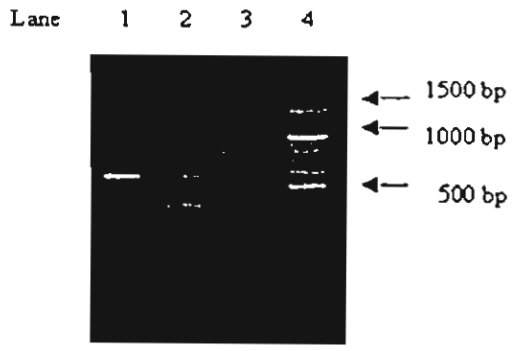

Figure 2. The PCR results of the affected persons, A.K. and B.K. Ethidium bromide-stained 2% agarose gels illustrating the PCR products. Lane 1: normal control; lane 2: A.K.; lane3: B.K.; and lane 4: genetic markers. The normal control showed a single band of 565 bp, while subjects A.K. and B.K. each showing two bands of 565 and 376 bp, were characterized as HPFH-6 heterozygotes.

DNA Sequencing Analysis

The deletion mutation and breakpoints were identified as shown in Fig. 3. The 5′ breakpoint is localized between nucleotides 68322 and 68323 according to the database of GenBank, access number NG_000007 (Homo sapiens genomic β-globin region, HBB@), while the 3′ breakpoint might be localized approximately at nucleotide 96,000, which has been reported between nucleotides 135 and 136 [Fig. 3(A) and 3(B)] of GenBank access number AF042277 (Homo sapiens 3′ breakpoint of a HPFH deletion). The shown DNA sequence data support the SEA deletion mutation according to the study of Motum et al. (8) and Xu et al. (4). Furthermore, nucleotide 207 in GenBank AF042277, located complementary to our reversed primer site, was reviewed and a replacement of C to A was found, similar to the previous report of Xu et al. (4). The modified sequence was submitted to GenBank and has received a new access number, AY156920 (Homo sapiens normal sequence below the 3′ breakpoint of SEA-HPFH). F3

DISCUSSION

We found that the genotype of the Karen people in Suan Pueng with a persistence of Hb F was similar to that of HPFH-6, previously reported in Vietnam, Cambodia, and

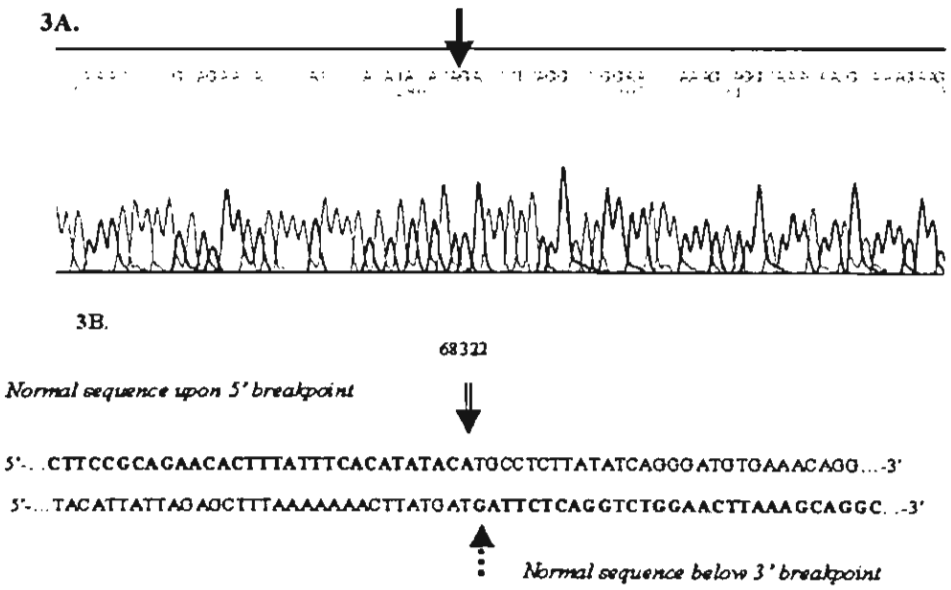

Figure 3. DNA sequencing analysis. (A) DNA sequence of these PCR products: the bold arrow indicates the breakpoint region. (B) Normal sequence around the 5′ and 3′ breakpoints: the bold letters indicate the intact DNA sequence in the PCR results, a double line arrow indicates the 5′ breakpoint at nucleotide 68322 of GenBank access number NG_000007, and the dotted arrow indicates the starting point below the 3′ breakpoint according to GenBank access number AY156920. WEB COLOR

China (4,8–10). Both 5′ and 3′ breakpoints are localized in a region similar to previous publications (4,8,9). After reviewing the reference sequence of the β-globin gene cluster (GenBank NG_000007), we could indicate that the 5′ breakpoint of this 27 kb deletion HPFH is located 1987 bp upstream from the starting point of the β-globin gene or 3761 bp downstream from the δ-globin gene. Approximately 2.3 kb below the 3′HS-1 LCR (17) of the β-globin gene was described as the location of the HPFH-6 3′ breakpoint, the sequence data adjacent to the breakpoint has already been identified by Xu et al. (4), and our data support the sequence of the previous studies (4,8,9). However, the development of a rapid PCR technique, low in cost and producing reliable results, for the detection of the HPFH-6 mutation in Thailand, is beneficial for definite molecular diagnosis.

The globin chain imbalance in heterozygous HPFH is milder in almost undetectable levels (2) due to the fact that γ-globin chain production in heterozygous deletional HPFH is quantitatively greater than in heterozygous deletional β-, δβ-, or $^{A}\gamma\delta\beta$-thals. Our results showed that the affected subjects had normal Hb levels and slightly decreased MCV and MCH levels because of adequate γ-globin chain compensation. The Hb A_2 level in the HPFH-6 deletion has been shown to be normal or slightly elevated in a previous report (9) because of an intact δ-globin gene; our results have shown a normal level of Hb A_2. The increase of RDW value in HPFH-6 is not clear and was reported in the study of Xu et al. (4). It is well known that RBCs containing Hb F (or F-cells), as found in newborns, have a larger size than normal, and RBCs with a globin chain imbalance, as found in anemias, have a smaller size. The pattern of the RDW graphs in these HPFH carriers was shown to be widely distributed (data not shown). Hence, it is possible that RBCs containing varied quantities of Hb F mixing with some RBCs with a globin chain imbalance might lead to varied red cell sizes and increased RDW values. This phenomenon is similar to the presence of RDW in β-thal carriers but the latter have more globin chain imbalance and a lower amount of F cells (1).

Although the molecular characterization and clinical phenotype of the HPFH-6 mutation are described, to the best of our knowledge, there have been no official reports of this condition in Thailand. We propose that the interpretation of Hb typing without clinical examination and complete blood examination led to the misdiagnosis of HPFH. Percentages of both Hb F and Hb A_2 can be found in varying levels in a number of mutations, both in HPFH and thalassemias, as published elsewhere (1). However, the phenotype should be examined carefully, because each condition leads to a different prognosis (18,19). The compound heterozygotes for β-thal and HPFH might lead to clinical thalassemia intermedia (4,19,20). The other interaction of β- and $(\delta\beta)^{0}$-, $(\delta\beta)^{+}$- or $(^{A}\gamma\delta\beta)^{0}$- might be referred to as a dominantly inherited β-thal condition, which has a different phenotype from other forms of β-thal intermedia in several aspects and a severe thalassemia major could occur (20). On the other hand, HPFH homozygotes show only clinically mild hypochromic and microcytic anemia which does not cause severe anemic symptoms throughout their lifetime (1,2,19). Genetic counseling for each particular condition is different because of disease severity, thus playing an important role in family planning for the couple at risk of high Hb F levels. In Thailand, prenatal diagnosis would be performed in the case of couples at risk of developing a severe thalassemia major. The abortion rate from the technique is 1–3%; therefore, there is no need to diagnose the fetus of a couple with thalassemia intermedia or HPFH homozygotes (13,21).

ACKNOWLEDGMENTS

This research is part of a project titled "Genomic approach in searching for genes involved in genetic susceptibility to clinical malaria." The research was supported in part by grants from Thailand Research Fund (BR-G-44-80-016) and National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand (BT-B-06-MG-14-4507). We are grateful to all Thai-Karen people in Suan Pueng for providing samples used in this study. We thank Dr. N. J. White, Wellcome Trust Foundation, Bangkok, Thailand, for critical review of this paper, Dr. A. Bunyaratvej and R. Wisedpanichkij, Department of Pathology, Ramathibodi Hospital, Bangkok, Thailand, for hematological and α-thal analyses, Dr. C. Tocharoentanaphol, Molecular Laboratory, Ramathibodi Hospital, Bangkok, Thailand, M. Boosabaratana and S. Youngcharoen, Medical Genetics Laboratory, Ramathibodi Hospital, Bangkok, Thailand, for technical assistance in PCR and DNA sequencing, Dr. V. Viprakasit and Dr. P. Vathesatogkit, MRC Molecular Haematology, WIMM, Oxford, UK, for assistance in literature and medical information searches. This paper was accepted as a poster presentation at the 4th HUGO Pacific meeting and 5th Asia-Pacific Conference on Human Genetics held at Ambassador City Jomtien, Pattaya, Cholburi, Thailand on October 27–30, 2002.

REFERENCES

1. Stamatoyannopoulos G, Grosveld F. Hemoglobin switching. In: Stamatoyannopoulos G, Majerus PW, Perlmutter RM, Varmus H, eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001:135–182.
2. Wood WG. Hereditary persistence of fetal hemoglobin and $\delta\beta$ thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, eds. Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. Cambridge: Cambridge University Press, 2001:356–388.
3. Weatherall DJ, Clegg JB. Hereditary persistence of fetal hemoglobin. In: Weatherall DJ, Clegg JB, eds. The Thalassaemia Syndromes. Oxford: Blackwell Scientific Publishers, 1981:450–507.
4. Xu X-M, Li Z-Q, Liu Z-Y, Zhong X-L, Zhao Y-Z, Mo Q-H. Molecular characterization and PCR detection of a deletional HPFH: application to rapid prenatal diagnosis for compound heterozygotes of this defect with β-thalassemia in a Chinese family. Am J Hematol 2000; 65:183–188.
5. Leonova JY, Kazanetz EG, Smetanina NS, Adekile AD, Efremov GD, Huisman THJ. Variability in the fetal hemoglobin level of the normal adult. Am J Hematol 1996; 53:59–65.
6. Craig JE, Rochette J, Sampietro M, Wilkie AOM, Barnetson R, Hatton CSR, Demenais F, Thein SL. Genetic heterogenity in heterocellular hereditary persistence of fetal hemoglobin. Blood 1997; 90:428–434.
7. Calzolari R, McMorrow T, Yannoutsos N, Langeveld A, Grosveld F. Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and $\delta\beta$-thalassemia affects β- but not γ-globin gene expression. EMBO J 1999; 15:949–958.

8. Motum PI, Hamilton TJ, Lindeman R, Le H, Trent RJ. Molecular characterization of Vietnamese HPFH. Hum Mutat 1993; 2:179–184.
9. Dimovski AJ, Divoky V, Adekile AD, Baysal E, Wilson JB, Prior JF, Raven JL, Huisman THJ. A novel deletion of approximately 27 kb including the β-globin gene and the locus control region 3′ HS-1 regulatory sequence: β^0-thalassemia or hereditary persistence of fetal hemoglobin? Blood 1994; 83:822–827.
10. Zeng Y-T, Huang S-Z, Chen B, Liang Y-C, Chang Z-M, Harano T, Huisman THJ. Hereditary persistence of fetal hemoglobin or $(\delta\beta)^0$-thalassemia: three types observed in South Chinese families. Blood 1985; 66:1430–1435.
11. Fucharoen S, Winichagoon P. Hemoglobinopathies in Southeast Asia. Hemoglobin 1978; 11(1):65–88.
12. Fucharoen S, Winichagoon P, Siritanaratkul N, Chowthaworn J, Pootrakul P. α- and β-thalassemia in Thailand. Ann NY Acad Sci 1998; 850:412–414.
13. Jaovisidha A, Ajjimarkorn S, Panburana P, Somboonsub O, Herabutya Y, Rungsiprakarn R. Prevention and control of thalassemia in Ramathibodi Hospital, Thailand. SEA J Trop Med Pub Hlth 2000; 31(3):561–565.
14. Bowden DK, Vickers MA, Higgs DR. A PCR-based strategy to detect the common severe determinants of α thalassemia. Br J Haematol 1992; 81:104–108.
15. Fucharoen S, Winichagoon P, Thonglairuam V. β-Thalassemia associated with α-thalassemia in Thailand. Hemoglobin 1988; 12(5 & 6):581–592.
16. Wang C, Beganyi L, Fernandes BJ. Measurements of red cell parameters in α-thalassemia trait: correlation with the genotype. Lab Hematol 2000; 6:163–166.
17. Fleenor DE, Kaufman RE. Characterization of the DNase I hypersensitive site 3′ of the human β globin gene domain. Blood 1993; 81:2781–2790.
18. Weatherall DJ. Overview: mechanisms for the heterogenity of the thalassemias. Int J Pediatr Hematol Oncol 1994; 4:3–4.
19. Charache S, Clegg JB, Weatherall DJ. The Negro variety of hereditary persistence of fetal hemoglobin is a mild form of thalassaemia. Br J Haematol 1976; 34:527–529.
20. Weatherall DJ. The thalassemias. In: Stamatoyannopoulos G, Majerus PW, Perlmutter RM, Varmus H, eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001:183–226.
21. Fucharoen S, Winichagoon P. Thalassemia in Southeast Asia: problems and strategy for prevention and control. SEA J Trop Med Pub Hlth 1992; 23:647–655.

Received November 22, 2002
Accepted January 7, 2003