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