RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2025.035

ScienceAsia 51 (2): 2025: ID 2025035: **1–9**

Genetic analysis of glutathione synthetase of *Plasmodium falciparum*, a potential candidate for antimalarial drug development

Phumin Simpalipan[a], **Sittiporn Pattaradilokrat**[b], **Vorthon Sawaswong**[c,d], **Korawich Uthayopas**[e], **Kittima Phutthawong**[b], **Pongchai Harnyuttanakorn**[b,*]

[a] Zoology PhD Program, Faculty of Science, Chulalongkorn University, Bangkok 10330 Thailand
[b] Department of Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330 Thailand
[c] Center of Excellence in Systems Microbiology, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok 10330 Thailand
[d] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400 Thailand
[e] School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane QLD 4072 Australia

*Corresponding author, e-mail: pongchai.h@chula.ac.th

Received 14 Feb 2024, Accepted 25 Feb 2025
Available online 1 May 2025

ABSTRACT: Glutathione synthetase (GS) is an enzyme involved in the synthesis of glutathione, a molecule protecting the parasite from heme-induced cell damage and potentially affecting antimalarial drug responses. Currently, the design of *in silico* assay for chemical screen is impeded due to the lack of complete gene sequences from endemic populations of malaria parasites. This study aimed to investigate the genetic diversity of the *gs* gene in *P. falciparum* populations in Thailand and worldwide. The catalog of the *gs* gene of *P. falciparum* was generated, consisting of sequences of 223 *P. falciparum* isolates from 15 countries. Population genetic analyses were conducted using the sequences of *gs* in the database to reveal the allelic and nucleotide diversity. In addition, the neutrality tests were performed to determine the evidence of natural selection. The analysis of the *gs* sequences in the database identified 55 unique haplotypes, characterized by 41 SNPs and 1 indel mutation. The most common *gs* haplotype was the *gs1* which was found in all *P. falciparum* populations. Sequence analysis also revealed low genetic diversity at the *gs* locus, likely due to negative selection. In conclusion, the present study establishes the catalog of *P. falciparum gs* gene and provides basic genetic diversity information for antimalarial drug design and for monitoring the effects of antimalarial drugs on parasite populations.

KEYWORDS: antimalarial drug, genetic diversity, glutathione synthetase, malaria, negative selection, *Plasmodium falciparum*

INTRODUCTION

Human malaria is one of the life-threatening infectious diseases in 87 countries worldwide [1]. Countries in Southeast Asia and South Asia are considered as the important malaria hotspots due to the emergence and spread of multi-drug resistance malaria [2] High failure rates of artemisinin treatment are reported [3], suggesting the possibility of clinical artemisinin resistance. This highlights the need of research to identify novel antimalarial targets from biological resources or to screen new combinations of antimalarial compounds to accompany a new method for malaria treatment [4–6]. Because the asexual blood stage malaria parasite is exclusively responsible for clinical symptoms of the disease, enzymes in metabolic pathways essential for the blood stage development could be novel sources of antimalarial targets.

During the development of the malaria parasites in an erythrocyte of a mammalian host, the malaria parasites must be exposed to environmental and metabolic stress, such as hydrogen peroxide (H_2O_2), and nitric oxide (NO), produced during immune responses by the hosts [7]. The stress negatively affects RBC membrane integrity and impact on parasite survival. To alleviate the adverse oxidative stress, the parasite employs glutathione (GSH) system that acts as an antioxidant and maintains its redox status [8]. GSH is the most abundant antioxidant tripeptide (γ-L-glutamyl-L-cysteinyl-glycine) that acts as a co-factor for detoxification enzymes, such as glutathione peroxidases (GPx) and glutathione S-transferases (GST). *P. falciparum* maintains high levels (2 mM) of reduced GSH in the cytosol [9]. The *de novo* biosynthetic pathway for GSH establishes a constant source of GSH in the cytoplasm [10]. GSH in *P. falciparum* can be synthesized using two ATP-dependent reactions. In the *de novo* biosynthesis pathway, γ-glutamylcysteine synthetase (γ-GCS, EC 6.3.2.2) catalyzes the conjugation of L-glutamate and L-cysteine, forming γ-L-glutamyl-L-cysteine. Subsequently, glutathione synthetase (GS, EC.6.3.2.3) catalyzes the conjugation of γ-L-glutamyl-L-cysteine and glycine, resulting in the production of GSH. Disruption of the *P. berghei* gamma-glutamylcysteine synthetase or glutathione reductase genes resulted in a significant reduction of GSH in intraerythrocytic stages, thereby negatively affecting the parasite development [11]. In addition, GSH also mediates the detoxification of

antimalarial drugs [12]. An increase in GSH levels could lead to an increase resistance to antimalarial drugs, such as chloroquine and artemisinin in rodent and human malaria [12]. This suggests the GSH-dependent detoxification system are not only related to the heme detoxification but also may be associated with the antimalarial drug resistance [13, 14]. Thus, antioxidant capabilities of the malaria parasite are of considerable significance.

Currently, compounds that interfere the GSH synthesis have been identified for *P. falciparum*, including D,L-buthionine-(S, R)sulphoximine (BSO) and methylene blue, which are an inhibitor of γ-GCS, and glutathione reductase (GR), respectively [15]. Nevertheless, there are no report of specific inhibitors for *P. falciparum* GS. Discovery of the novel inhibitors targeting the GS may provide intervention strategies against *Plasmodium* and related parasites. According to the genome sequence of *P. falciparum*, there is one copy of the gene encoding GS (*gs*) (PF3D7_0512200) in the genome [16]. The *gs* gene contains one exon, spanning 1,968 bp. in length. The deduced polypeptide of GS contains 655 amino acids, with an estimate size of 70 kDa. The orthologs could be identified in other malaria parasites of human, including *P. vivax*, and animal models, such as *P. knowlesi* and *P. yoelii* [17–19]. RNA sequencing and proteomic expression analyses showed that GS was synthesized throughout the asexual blood stage cycle and the highest expression levels were detected in trophozoite [20]. Recently, a 3D-model of GS was constructed using *de novo* modeling method, facilitating the discovery of potential novel GS inhibitors [21]. However, the model was based on the sequence of the *P. falciparum* 3D7 strain [16]. Currently, genome sequence data of *P. falciparum* has been obtained from wild isolates from several studies, but there is no report that focus on the analysis of the haplotype (allelic) diversity of the *gs* gene in the parasite populations and explore the roles of GS as a potential drug target [22, 23]. Here, we generated the *gs* sequences of *P. falciparum* in Thailand and combined them with the hundreds of sequences from public resources [24], thereby presenting the first catalog of the complete *gs* sequences of *P. falciparum*. Therefore, the goals of the present study are to investigate the extent of the sequence diversity of *P. falciparum gs* gene in Thailand. The outcome of this study will further facilitate *in silico* assay development for inhibitor screening and epidemiological studies to monitor the susceptibility of parasite against antimalarial drugs.

MATERIALS AND METHODS

Origins and cultivation of the malaria parasite

A total of 61 samples of the malaria parasite *P. falciparum* were obtained from a repository at Malaria Research Laboratory, Department of Biology, Faculty of Science, Chulalongkorn University (Bangkok, Thailand) (Table S1). The origins of the parasites are published elsewhere [25, 26]. The species of the malaria parasites was identified using a microscopic examination by experienced parasitologists. These isolates were also confirmed using genotyping methods with microsatellites and sequencing of antigen-coding genes and were maintained as frozen stabilates kept in −80 °C [25, 27]. The study has been reviewed and approved by the Research Ethics Review Committee for Research involving Human Research Participants, Chulalongkorn University (COA No. 041/2016) and the Ethics Committee for Research in Human Subjects, Department of Disease Control, Ministry of Public Health (No. FWA00013622).

Preparation of genomic DNA

Genomic DNA of *P. falciparum* was extracted using a standard phenol-chloroform purification procedure [25, 26]. Genomic DNA was re-suspended in Tris-EDTA (TE) buffer and stored at −20 °C for gene amplification.

Amplification of the gs gene and DNA sequencing

Nucleotide sequences of *P. falciparum* 3D7 strain chromosome 5 (NCBI ID: NC_004326.2) were retrieved from the National Center for Biotechnology Information. The *gs* gene of *P. falciparum* was located on the genomic position 20929 and 28456 on *P. falciparum* chromosome 5 (XM_001351643) [16]. Primers for Polymerase Chain Reaction (PCR) and DNA sequencing were designed using Primer3Plus online software [28]. Sequences for PCR and DNA sequencing of the gs gene were shown in Table S2.

PCR reactions for amplification of *P. falciparum gs* gene were prepared in a 50-µl volume containing 2 mM $MgCl_2$, 200 mM dNTPs, 0.5 µM forward primer (GS1F) and reverse primer (GS1941R), 10–100 ng of DNA templates, 1 U of *Taq* polymerase (Biotechrabbit, Germany) in 1 × reaction buffer. PCR conditions were 95 °C for 5 min, followed by 35 cycles of 95 °C 40 s, 58 °C 40 s, 60 °C 2 min, and followed by a final extension of 60 °C 8 min. Gene amplification was performed in Veriti™ 96 well thermal cycler (Applied Biosystems, USA). PCR products were analyzed using standard agarose gel electrophoresis and were submitted to commercial Sanger DNA sequencing (Bioneer, The Republic of Korea). Eight sequence reactions were performed using 2 PCR primers and additional 6 DNA sequencing primers, including GS712F, GS1316F, GS516R, GS1074R, GS1560R and GS1941R (Table S2). DNA sequence texts and electropherograms were inspected and were assembled using MEGA7 program [29].

Retrieval of nucleotide sequences from public databases

We searched nucleotide sequences of the *gs* gene of *P. falciparum* in two public databases: PlasmoDB (http://plasmodb.org/plasmo) and NCBI (http://www.ncbi.nlm.nih.gov) websites. As of December 2024, a total of 202 and 13 nucleotide sequences were in the PlasmoDB and NCBI databases [24]. Of these, 162 sequences had unambiguous nucleotide reads and were used to construct the global database of *gs* of *P. falciparum* (Table S1).

Nucleotide sequence analysis

The number of single nucleotide polymorphic sites, alleles, and polypeptide sequences were determined from nucleotide alignments using BioEdit sequence editor (version 7.2.6) and DnaSP software (version 6) [30]. Nucleotide diversity (π) was referred to the number of nucleotide differences per nucleotide site between pair of sequence. Sliding window plots of π, with a window length of 90 nucleotides and a step size of 3 nucleotides, were constructed to illustrate SNP distributions. Haplotype diversity index (H) was extrapolated from a relative frequency of each haplotype and the number of sequences in the alignment dataset. Log likelihood score of nucleotide substitution models for DNA sequence alignments were determined from 4×4 nucleotide substation rate matrix combining with rate variation among nucleotide sites and equal/unequal base frequencies in jModelTest program. The best-fit nucleotide substitution model was selected using Bayesian information criteria.

Departures from the neutrality of molecular evolution were calculated using three neutrality tests, Tajima's D, Fu and Li's D^* and Fu and Li's F^* computational modules in DnaSP 6 software. D, D^* and F^* values were considered to be statistically significant when p value was less than 0.05. The mean number of non-synonymous substitutions per non-synonymous sites (d_N) and synonymous substitutions per synonymous sites (d_S) was estimated using Nei and Gojobori method combined with the best-fit nucleotide substitution model.

Divergence of alleles, detecting in alignment of each gene, among *P. falciparum* populations in Thailand as well as the population differentiation between parasites in Africa, South America and Asia were analysed by Wright's fixation index (F_{st}) in DnaSP 6.10.1 software. Significant differences of F_{st} variance for any pairs of parasite populations were deemed at p value of less than 0.05.

Phylogenetic analysis

Multiple sequence alignment of 55 haplotypes of the *gs* gene from *P. falciparum* and the ortholog gene from *P. reichenowi* (outgroup [31]) was generated by ClustalW algorithm in BioEdit 7.2. The best-fit nucleotide substitution model was determined from ClustalW aligned sequences, using log likelihood calculation and Bayesian information criteria selection strategy in MEGA11 program. Evolutionary relationships of the sequences were traced by maximum likelihood (ML) using RAxML with GTR+G+I substitution model [32]. The robustness of ML tree topology was challenged with 200 replicates autoMRE Bootstrap resampling method. The final ML tree with bootstrap value were visualized by iToL web server [33].

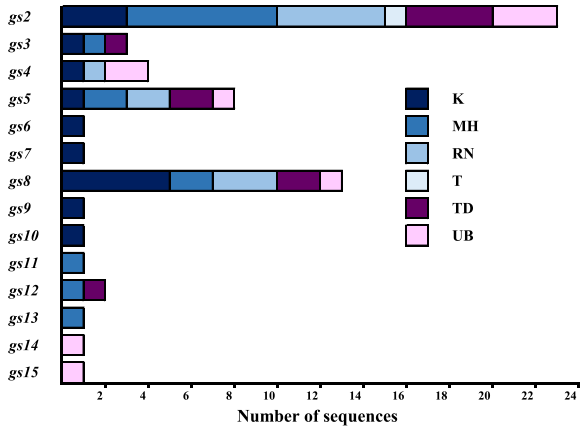

Fig. 1 Frequencies and distribution of the 14 *gs* haplotypes of *P. falciparum* from Thailand. Abbreviation: K, Kanchanaburi; MH, Mae Hong Son; RN, Ranong; T, Tak; TD, Trat; UB, Ubon Ratchathani. *gs1* was detected in *P. falciparum* strain 3D7 [16], but was not detected in any *P. falciparum* populations in Thailand.

RESULTS

Genetic diversity of the gs gene in *P. falciparum* from Thailand

PCR products of the *gs* gene of the malaria parasite *P. falciparum* was amplified from genomic DNA extracted from 61 samples. DNA sequencing generated nucleotide sequences, with polymorphic size between 1944 and 1980 bp, corresponding to nucleotide positions of 16 to 1968 (1953 bp in length) of *gs* of *P. falciparum* strain 3D7. The sequence alignment of 61 *gs* fragments revealed 9 SNPs, at positions 31, 432, 439, 441, 888, 1126, 1438, 1908, and 1956. All SNPs were dimorphic. Of these, 7 sites at nucleotide positions 31 (GTT/TTT), 439 and 441 (GAC/AAT), 1126 (CTA/ATA), 1438 (AAA/GAA), 1908 (AAT/AAG), and 1956 (TTA/TTT) resulted in non-synonymous amino acid substitutions in 6 codons, corresponding to the amino acid positions 11 (V/F), 147 (D/N), 376 (L/I), 480 (K/E), 636 (K/N) and 652 (L/F) of the GS sequence, respectively. Allele frequencies at SNP positions at 31, 439, 441, 888, 1438, 1908 were 60:1, while allele frequencies at SNP positions at 1126 and 1956 were 17:44 and 58:3, respectively. The other two SNPs loci at positions 432 (GAT/GAC) and 888

Table 1 Nucleotide diversity indices and neutrality tests on the GS gene of *P. falciparum* in a global sequence database.

Origin	N	Genetic variation			Tests of neutrality			Tests of positive selection	
		allele	H	π	*D*	*D**	*F**	*dN*	*dS*
Thailand*	61	14	0.7984±0.0371	0.0049±0.0025	−1.6711	−3.4679**	−3.3875**	0.0004±0.0003	0.0003±0.0001
Vietnam	3	3	1.0000±0.2720	0.00103±0.0004	ND	ND	ND	ND	ND
Malaysia	1	1	ND	ND	ND	ND	ND	ND	ND
India	1	1	ND	ND	ND	ND	ND	ND	ND
Asia	**66**	**18**	**0.8118±0.0363**	**0.0050±0.0025**	**−1.9353***	**−3.1350***	**−3.2233****	**0.0005±0.0003**	**0.0007±0.0003**
Senegal	47	15	0.8862±0.0250	0.0012±0.0007	−0.6932	0.1135	−0.1846	0.0008±0.0004	0.0032±0.0015
Gambia	52	23	0.9465±0.0143	0.0014±0.0008	−0.7627	−1.0593	−1.1345	0.0009±0.0004	0.0034±0.0018
Uganda	6	5	0.9333±0.1217	0.00158±0.0011	−0.7353	−0.6825	−0.7463	0.0015±0.0006	0.0018±0.0012
Mali	21	11	0.8762±0.0581	0.0012±0.0007	−0.6014	−0.5222	−0.6325	0.0010±0.0005	0.0019±0.0009
Ghana	1	1	ND	ND	ND	ND	ND	ND	ND
Mozambique	1	1	ND	ND	ND	ND	ND	ND	ND
Tanzania	1	1	ND	ND	ND	ND	ND	ND	ND
*P. falciparum 3D7**	1	1	ND	ND	ND	ND	ND	ND	ND
Africa	**130**	**42**	**0.9215±0.0135**	**0.0014±0.0008**	**−1.8050***	**−3.5120****	**−3.3811****	**0.0009±0.0004**	**0.0032±0.0016**
Brazil	3	2	0.6000±0.1753	0.0006±0.0005	ND	ND	ND	0.0004±0.0004	0.0018±0.0018
French Guiana	22	4	0.7100±0.0625	0.0006±0.0005	0.4711	1.0955	1.0624	0.0003±0.0002	0.0023±0.0016
El Salvador	1	1	ND	ND	ND	ND	ND	ND	ND
Honduras	1	1	ND	ND	ND	ND	ND	ND	ND
South America	**27**	**5**	**0.7607±0.0522**	**0.0007±0.0005**	**0.1151**	**1.1543**	**0.9881**	**0.0004±0.0003**	**0.0021±0.0014**
Overall	**223***	**55**	**0.9291±0.0080**	**0.0029±0.0016**	**−1.9209***	**−4.8715****	**−4.3251****	**0.0009±0.0004**	**0.0025±0.0011**

Table 2 Variation of the repetitive region in GS of *P. falciparum*. The nucleotide and amino acid sequences of tandem repeats and the downstream franking region were aligned. The repetitive sequences could be classified as 5 types, named R1, R2, R3, R4 and R5. R1 was detected in *P. falciparum* strain 3D7.

Type	nt. (aa.)	Repetitive sequence	nt. (aa.)
R1	412 (138)	GAT AAT AGT GAT AAT AGT GAT AAT AGT GAC AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT — — — — — — — — TTT D N S D N S D N S D N S N N S N N S N N F	483 (161)
R2	412 (138)	GAT AAT AGT GAT AAT AGT GAT AAT AGT GAC AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT — — — — — — TTT D N S D N S D N S D N S N N S N N S N N S N N F	492 (164)
R3	412 (138)	GAT AAT AGT GAT AAT AGT GAT AAT AGT GAC AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT TTT D N S D N S D N S D N S N N S N N S N N S N N S N N S N N S N N F	510 (170)
R4	412 (138)	GAT AAT AGT GAT AAT AGT GAT AAT AGT GAC AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT — — — — — — — — — — — — TTT D N S D N S D N S D N S N N S N N S N N F	474 (158)
R5	412 (138)	GAT AAT AGT GAT AAT AGT — — — GAC AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT AGT AAT AAT — — — — — — TTT D N S D N S D N S N N S N N S N N S N N F	483 (161)

nt. = nucleotide position; aa. = amino acid position.

(CAT/CAC) resulted in synonymous amino acid substitutions, corresponding to the amino acid positions 144 (D) and 296 (H). Average nucleotide diversity of the 61 *gs* sequences of *P. falciparum* was estimated to be 0.0049 ± 0.0025 (Table 1). This indicates low genetic diversity in the *gs* gene *P. falciparum* population in Thailand.

The nucleotide sequence alignment also revealed the size polymorphism within this gene. Insertion/deletion mutations (indel) were detected in two regions, corresponding to the nucleotide positions 412 and 480 of *gs* of *P. falciparum* 3D7. The indel mu-

Table 3 F_{st} values of *gs* alleles between *P. falciparum* in 5 different populations in Thailand.

	UB	TD	MH	RN
K	−0.04659 NS	−0.07063 NS	0.06062 NS	−0.05566 NS
RN	−0.07286 NS	−0.05818 NS	0.07914 NS	
MH	0.01402 NS	−0.03008 NS		
TD	−0.06647 NS			

K, Kanchanaburi; MH, Mae Hong Son; RN, Ranong; TD, Trad; UB, Ubon Ratchathani; NS, non-significant.

tations encoded variable lengths of two tripeptide repeats (DNS or NNS), resulting in 5 sequence patterns, namely R1, R2, R3, R4 and R5 (Table 2). Of the five patterns, R3 was the longest sequence, containing 4 DNS repeats joined with 6 NNS repeats, while the shortest was R4 containing 4 DNS repeats and 2 NNS repeats. Sequence of GS of *P. falciparum* 3D7 had R1 pattern, which contained 4 DNS repeats and 3 NNS repeats. For *P. falciparum* in Thailand, the repeats patterns R1, R2 and R3 were the most common, with frequencies of 29, 22 and 8 respectively. These results demonstrated that SNP and indel mutations contributed to the polymorphism in the *gs* gene of *P. falciparum*.

The comparison of *gs* in *P. falciparum* in Thailand revealed 14 distinct haplotypes, named *gs2* to *gs15*, with the estimated haplotype index (*Hd*) of 0.7984 ± 0.0371 (Table 1). The *gs1* haplotype was named after the *gs* allele of *P. falciparum* 3D7, while the *gs* sequences of K1CB1 and T9/94/RC17 clones were named as haplotypes *gs2* and *gs3*, respectively (Table S1). The distribution and the frequency of the 14 *gs* haplotypes were shown in Fig. 1. Analysis of haplotype distributions identified 3 haplotypes, *gs2*, *gs5* and *gs8*, which were distributed all sampling sites, with frequencies of 22, 8 and 13. Eight others were unique or region-specific haplotypes and were found in single parasite isolate each, including *gs6*, *gs7*, *gs9*, and *gs10* in *P. falciparu*m from Kanchanaburi Province, *gs11* and *gs13* in *P. falciparum* from Mae Hong Son, and *gs14* and *gs15* in *P. falciparum* from Ubon Ratchathani. As a result, a total of 9, 1, 7, and 4 haplotypes were detected in *P. falciparum* from Kanchanaburi, Tak, Mae Hong Son and Ranong, located along the border of Thailand and Myanmar respectively, while 5 and 6 haplotypes were found in *P. falciparum* from Trat and Ubon Ratchathani, located on the borders of Thailand-Laos and Thailand-Cambodia, respectively.

To determine the genetic differentiation between *P. falciparum* populations, the Wright's fixation index (F_{st}) was analyzed as described in *Materials and Methods*. F_{st} values were not significant among all pairs of *P. falciparum* populations in Thailand (Table 3). This result indicates the genetic homogeneity in the *gs* alleles among different *P. falciparum* populations in Thailand.

Genetic diversity of the *gs* locus in *P. falciparum* worldwide

To compare the genetic diversity of *gs* of *P. falciparum* in Thailand and other parasite populations, the *gs* sequences of 212 isolates of *P. falciparum* were retrieved from the PlasmoDB and NCBI databases (Table S1). These sequences were originally from natural *P. falciparum* isolates from 15 populations worldwide. Of these, 162 sequences contained unambiguous nucleotide reads and were included in a subsequent analysis. After combined with the 61 *gs* sequences of *P. falciparum* from Thailand generated in the present study, this resulted in a database comprising 223 *gs* sequences.

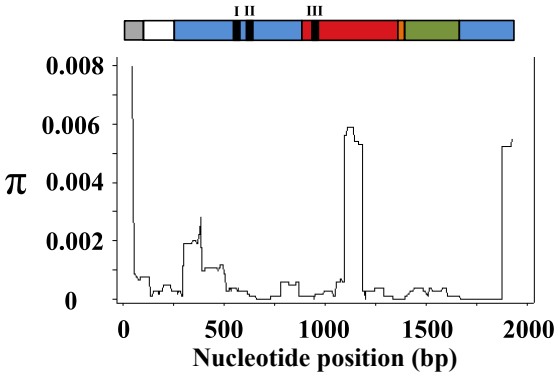

Fig. 2 Sliding window plot of average pairwise nucleotide diversity index (π) of the *gs* gene of *P. falciparum*. The π values were plotted by sliding window of 90 bp and step size of 3 bp. Nucleotide positions 16-1968 were analyzed. Domains in *gs* were divided into dimerization unit (grey), C-terminal domain (blue), N-terminal domain (red), helical linker (orange), and lid domain (green).

Nucleotide sequence analysis of the global *gs* database revealed 41 SNP sites. Of these, 28 and 13 SNPs resulted in non-synonymous and synonymous mutations, respectively (Table S3). Four SNPs at nucleotide positions 18 (GAT/GAC), 21 (GAG/GAA), 1126 (CTA/ATA) and 1908 (AAG/AAT) were detected in the majority of *P. falciparum* isolates, with frequencies of 80:20, 80:20, 66:34, and 37:63. Other SNPs had the frequency of $< 5\%$. A sliding window plot of the nucleotide diversity index showed that SNPs were clustered in 3 regions, the N-terminal sequence (nucleotide position 18–24), the central region containing repetitive sequence (nucleotide positions 1096–1198) and the C-terminal region (nucleotide positions 1908–1956) (Fig. 2). The nucleotide diversity of *gs* varied according to the geographical locations of the parasites. The nucleotide diversity indices of *gs* were 0.0050 ± 0.0025, 0.0014 ± 0.0008, and 0.0007 ± 0.0005 in *P. falciparum* populations from Asia, Africa and South America, respectively (Table 1). The overall π was 0.0029 ± 0.0016. This data indicates that the sequence diversity of *gs* was higher in *P. falciparum* populations in Asia than those in Africa and South America.

In addition, the sequence analysis identified additional 40 haplotypes, named *gs16* to *gs55*, in the global *gs* sequence database (Table S4). In these haplotypes, the previously described 5 types of the tripeptide repeats were detected. The majority (188 samples, 84.3%) of these sequences contained R1 tripeptide repeats. While R2 and R3 tripeptide repeats were found

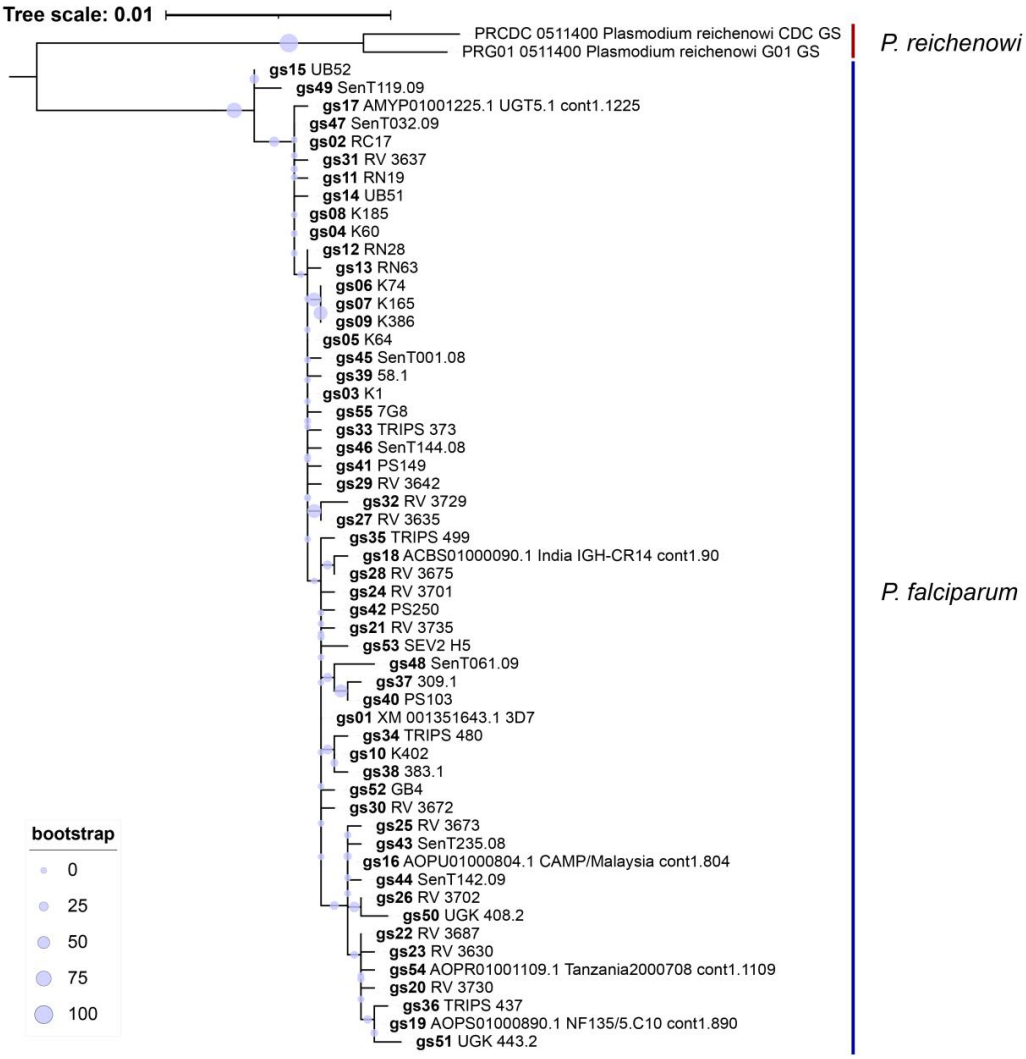

Fig. 3 Maximum likelihood tree of the 55 haplotypes of GS- encoding genes in *P. falciparum*. The *gs* gene of *P. reichenowi* was used as an outgroup. The taxa were named following the haplotypes and their representative strain names, respectively. Bootstrap values were shown as the circular sign on the clades which represent the degree of bootstrap value based on its size. Scale bar indicated nucleotide substitution per site.

Table 4 F_{st} values of *gs* alleles between *P. falciparum* in 6 malaria endemic countries.

	THA	SEN	GAM	UGA	MAL
SEN	0.34427*				
GAM	0.34193*	0.01328 NS			
UGA	0.36799*	0.08721 NS	0.0368 NS		
MAL	0.34284*	−0.0177 NS	0.01369*	0.11789 NS	
FG	0.36994*	0.17923*	0.17115*	0.26586*	0.13486*

THA, Thailand; SEN, Senegal; GAM, The Gambia; MAL, Mali; UGA, Uganda; FG, French Gu.

exclusively in *P. falciparum* populations in Thailand, R4 and R5 could be found in *P. falciparum* isolates from India and Senegal, respectively (Table S1). The global distribution pattern of the haplotypes revealed that there was a total of 18, 41 and 5 haplotypes in *P. falciparum* populations in Asia, Africa and South America, with the haplotype indices of 0.8118 ± 0.0363, 0.9215 ± 0.0135 and 0.7607 ± 0.0522 (Table 1). The *gs1*, *gs2* and *gs3* haplotypes were the most common haplotypes, with frequency of 15.2, 14.3 and 10.8. Three other haplotypes, *gs8*, *gs16* and *gs55*, were minor haplotypes, which were identified in 13, 15 and 11 parasites, respectively. A large majority of the haplotypes were identified in single or a few parasites, with frequency of $< 5\%$. (Table S4). Phylogenetic analysis of 55 haplotypes revealed that all haplotypes form a monophyletic group (Fig. 3). This supports the view that the sequences of *gs* was highly conserved.

In order to determine the genetic difference between *P. falciparum* populations in different continents, the Wright's F_{st} statistics was calculated. In this test, the populations with $n > 10$ were included in the analysis (Table 4). The F_{st} values were significant between Thailand and all countries as well as between pairs of French Guiana and all countries. Significant F_{st} values were not detected between any pairs of parasites in Africa. This indicated that the population sub-division in three continents, Asia, Africa and South America. To further confirm this finding, the F_{st} analysis was performed using alleles of the parasite populations in Asia, Africa and South America. F_{st} values between Asia-Africa, Asia-South America and South America-Africa were 0.3482 ($p < 0.05$ 0.2472 ($p < 0.05$) and 0.1306 ($p < 0.05$), respectively, supporting the allelic differentiation among populations in Asia, Africa and South America.

Evidence of negative selection

Having shown that the genetic diversity of *gs* gene in *P. falciparum* was low, a next goal is to investigate if the *gs* locus was subjected to negative selection. In this study, three neutrality tests, including Tajima's *D* test, Fu and Li's *D** test and Fu and Li's *F** test, were performed.

The *D*, *D** and *F** values of 67 *gs* sequences in *P. falciparum* in Asia showed significant and negative values of -1.9353 ($p < 0.05$), -3.1350 ($p < 0.01$) and -3.2233 ($p < 0.01$) respectively (Table 1). Furthermore, the *gs* sequences of *P. falciparum* from populations in Africa and South America were also analyzed by neutrality testes. *D*, *D** and *F** values from *gs* in *P. falciparum* in Africa were negative and significant, whereas neutrality tests of *gs* in *P. falciparum* in South America showed positive but not significant values. Nevertheless, when the neutrality tests were performed using all available sequences, *D*, *D** and *F** values showed significant and negative values of -1.9209 ($p < 0.05$), -4.8715 ($p < 0.01$) and -4.3251 ($p < 0.01$), respectively. This result suggested that *gs* of *P. falciparum* were under negative selection. To further determine the signature of negative selection, d_N/d_S ratios were also calculated using the *gs* sequences from *P. falciparum* populations in Thailand and other countries. In all cases d_N/d_S of *gs* from all *P. falciparum* population were less than 1 (or almost close to 0) (Table 1). Taken together, these results suggested evidence of negative selection in the *gs* gene of *P. falciparum*.

DISCUSSION

The primary objective of the present study was to investigate the sequence diversity of *gs* gene in *P. falciparum*. The study revealed for the first time the molecular basis of sequence polymorphism in the *gs* gene in *P. falciparum* in populations worldwide. The *gs* gene sequences could be classified into 55 haplotypes, which were characterized by 41 SNPs and 5 types of indel mutations (short amino acid repeats of DNS and NNS). We firstly demonstrated that there were 3 major haplotypes of *gs*, named *gs1*, *gs2* and *gs3*. While the *gs2* was the most abundant in *P. falciparum* in Thailand, the *gs1* and *gs3* variants were more common in Africa. This finding was consistent with the Wright (F_{st}) statistics analysis as well as data from previous studies indicating that there were different population structures between *P. falciparum* in Asia and Africa [34]. Despite the high haplotypic diversity, the levels of sequence variation (π) were very low. This result was in agreement with a previous analysis which showed that the genes encoding housekeeping enzymes, such as lactase dehydrogenase (*ldh*) exhibit low genetic diversity [35]. The level of genetic diversity of *gs* was lower than the antigen encoding genes, including merozoite surface protein-3 (*msp-3*), apical membrane antigen 1 (*ama-1*) and glutamate-rich protein (*glurp*) [36–38]. The low levels of sequence variations could be essential for the maintenance of the protein functions. Consistent with the above finding, phylogenetic analysis also showed the monophelytic clade of 55 haplotypes of GS. The low sequence diversity was likely due to the functional constraints, which was maintained by negative selection. The evidence of negative selection was demonstrated in this study using three neutrality tests and d_N/d_S ratio of <1. Because GS is an enzyme involved in detoxification process of the malaria parasite and the protein conformation could be important for catalytic activities [6]. To date, the functions of GS was reported by a study using the malaria parasite *P. berghei*. In one study, the *gs* gene of the rodent malaria parasite *P. berghei* was cloned and expressed in *E. coli* [39]. The protein was found to contain binding sites of ATP and the substrate binding sites. Although there is no report of the gene deletion of *gs* in any human or rodent malaria parasites, attempt to knockout one of the gene

in the GSH pathway, the γ-*gcs* gene, of *P. falciparum* was proven to be unsuccessful. This suggests that the sequences of GS reported here could be useful for gene editing technology to identify conserved region for functional studies.

Additional direct benefits of the sequence analysis of the *gs* gene in our database include the identification of the major haplotypes of *gs* in parasite populations in different countries. Among the 55 haplotypes identified herein, the three major haplotype *gs1*, *gs2* and *gs3* were the most prevalent types in *P. falciparum* population and should be chosen for *in silico* drug screening. Currently, a 3D model of *Plasmodium* GS was constructed by *de novo* modelling method using the *P. falciparum* strain 3D7, with the *gs1* haplotype [21]. The assay should also be applicable to include the two other major haplotypes to ensure that the compounds could target parasites with diverse background. A library of GSH analogues could be retrieved from Ligand-info database [21], and this could provide novel sources for discovery of new inhibitors using molecular docking.

In addition to the inhibition of GSH synthesis by screening for inhibitors of GS, there were several studies demonstrating that targeting enzymes in the GSH pathway could delay or kill the human malaria parasites. This was due to the fact that the infection with *P. falciparum* caused drastic changes in the GSH metabolism of red blood cells (RBCs). Indeed, infected RBCs lose GSH at a rate 40-fold higher than non-infected RBCs [15]. GSH depletion by BSO inhibited the development of *P. falciparum* with an IC_{50} of 73 μM. The effect of the drug was abolished by supplementation with GSH [15]. In addition, nitrofurantoin, a clinically used antibacterial drug, has been shown to disturb the redox balance by inhibiting GR activity. The drug enhanced the oxidative stress and were able to inhibit growth of intra-erythrocytic stages of both ART-sensitive and ART-resistant strains of *P. falciparum* [40]. Because the depletion of GSH through blocking the enzymes in the GSH synthesis pathway is considered as a chemotherapeutic strategy for malaria, and presents an emerging approach for expediting antimalarial drug development and circumventing resistance, the database of *P. falciparum* GS variants could be readily applicable for aiding design of novel effective inhibitors.

CONCLUSION

Like many housekeeping genes in the *P. falciparum*, glutathione synthetase exhibited extremely low polymorphism, likely due to the functional constraints. Although sequence polymorphism could differentiate the parasites according to the geographical location, such polymorphism is low. This raises a possibility that *gs* could potentially be a target of antimalarial inhibitors. Finding specific inhibitors of GSH production enzymes might be a solution to antimalarial drug resistance in Thailand and the Greater Mekong subregion.

Appendix A. Supplementary data

Supplementary data associated with this article can be found at https://dx.doi.org/10.2306/scienceasia1513-1874.2025.035.

Acknowledgements: This research was supported by a grant from the 100th Anniversary Chulalongkorn University for Doctoral scholarship and Overseas Research Experience Scholarship for Graduate Student to Mr. Phumin Simpalipan. We thank Dr. Robert Butcher for language editing. We also thank Mr. Kritsada Katawutpoonphan and Mr. Sakkarin Sansook for discussion and technical assistance.

REFERENCES

1. Danis M (2023) Malaria today. *Med Trop Sante Int* **3**, 375.
2. Imwong M, Dhorda M, Myo Tun K, Thu AM, Phyo AP, Proux S, Suwannasin K, Kunasol C, et al (2020) Molecular epidemiology of resistance to antimalarial drugs in the Greater Mekong subregion: an observational study. *Lancet Infect Dis* **20**, 1470–1480.
3. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, et al (2014) Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* **371**, 411–423.
4. Sriphana U, Pitchuanchom S, Kongsaeree P, Yenjai C (2013) Antimalarial activity and cytotoxicity of zerumbone derivatives. *ScienceAsia* **39**, 95–99.
5. Maicheen C, Ungwitayatorn J (2019) Antimalarial and β-hematin formation inhibitory activities of chromone derivatives. *ScienceAsia* **45**, 221–228.
6. van der Pluijm RW, Tripura R, Hoglund RM, Pyae Phyo A, Lek D, Ul Islam A, Anvikar AR, Satpathi P, et al (2020) Triple artemisinin-based combination therapies versus artemisinin-based combination therapies for uncomplicated *Plasmodium falciparum* malaria: a multicentre, open-label, randomised clinical trial. *Lancet* **395**, 1345–1360.
7. Hunt NH, Stocker R (1990) Oxidative stress and the redox status of malaria-infected erythrocytes. *Blood Cells* **16**, 499–526.
8. Müller S (2015) Role and regulation of glutathione metabolism in *Plasmodium falciparum*. *Molecules* **20**, 10511–10534.
9. Becker K, Tilley L, Vennerstrom JL, Roberts D, Rogerson S, Ginsburg H (2004) Oxidative stress in malaria parasite-infected erythrocytes: host-parasite interactions. *Int J Parasitol* **34**, 163–189.
10. Krauth-Siegel RL, Müller JG, Lottspeich F, Schirmer RH (1996) Glutathione reductase and glutamate dehydrogenase of *Plasmodium falciparum*, the causative agent of tropical malaria. *Eur J Biochem* **235**, 345–350.
11. Padín-Irizarry V, Colón-Lorenzo EE, Vega-Rodríguez J, Castro Mdel R, González-Méndez R, Ayala-Peña S, Serrano AE (2016) Glutathione-deficient *Plasmodium berghei* parasites exhibit growth delay and nuclear DNA damage. *Free Radic Biol Med* **95**, 43–54.
12. Ginsburg H, Golenser J (2003) Glutathione is involved in the antimalarial action of chloroquine and its modulation affects drug sensitivity of human and murine species of *Plasmodium*. *Redox Rep* **8**, 276–279.

13. Kotepui M, Mahittikorn A, Anabire NG, Kotepui KU (2023) Impact of malaria on glutathione peroxidase levels: a systematic review and meta-analysis. *Sci Rep* **13**, 13928.
14. Kotepui M, Kotepui K, Mahittikorn A, Majima HJ, Tangpong J, Yen HC (2023) Association of reduced glutathione levels with *Plasmodium falciparum* and *Plasmodium vivax* malaria: a systematic review and meta-analysis. *Sci Rep* **13**, 16483.
15. Lüersen K, Walter RD, Müller S (2000) *Plasmodium falciparum*-infected red blood cells depend on a functional glutathione *de novo* synthesis attributable to an enhanced loss of glutathione. *Biochem J* **346**, 545–552.
16. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, et al (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511.
17. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, et al (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519.
18. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, et al (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**, 757–763.
19. Pain A, Böhme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, et al (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799–803.
20. Kamaliddin C, Guillochon E, Salnot V, Rombaut D, Huguet S, Guillonneau F, Houzé S, Cot M, et al (2021) Comprehensive analysis of transcript and protein relative abundance during blood stages of *Plasmodium falciparum* infection. *J Proteome Res* **20**, 1206–1216.
21. Kumar YN, Jeyakodi G, Thulasibabu R, Gunasekaran K, Jambulingam P (2015) Screening and identification of inhibitors against glutathione synthetase, a potential drug target of *Plasmodium falciparum*. *Comb Chem High Throughput Screen* **18**, 492–504.
22. Ahouidi A, Ali M, Almagro-Garcia J, Amambua-Ngwa A, Amaratunga C, Amato R, Amenga-Etego L, Andagalu B, et al (2021) An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. *Wellcome Open Res* **6**, 42.
23. Abdel Hamid MM, Abdelraheem MH, Acheampong DO, Ahouidi A, Ali M, Almagro-Garcia J, Amambua-Ngwa A, et al (2023) Pf7: an open dataset of *Plasmodium falciparum* genome variation in 20,000 worldwide samples. *Wellcome Open Res* **8**, 22.
24. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, et al (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* **37**, D539–D543.
25. Pumpaibool T, Arnathau C, Durand P, Kanchanakhan N, Siripoon N, Suegorn A, Sitthi-Amorn C, Renaud F, et al (2009) Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. *Malar J* **8**, 155.
26. Simpalipan P, Pattaradilokrat S, Siripoon N, Seugorn A, Kaewthamasorn M, Butcher RD, Harnyuttanakorn P (2014) Diversity and population structure of *Plasmodium falciparum* in Thailand based on the spatial and temporal haplotype patterns of the C-terminal 19-kDa domain of merozoite surface protein-1. *Malar J* **13**, 54.
27. Cheychom J, Kanchanakhan N, Vijaykadga S, Gaywee J, Harnyuttanakorn P (2015) Cytochrome b mutation and atovaquone susceptibility in *Plasmodium falciparum* isolates from the Thai-Cambodian border during 1990–2010. *ScienceAsia* **41**, 340–344.
28. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71–W74.
29. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870–1874.
30. Rozas R, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* **34**, 3299–3302.
31. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, Shaw KS, Ayouba A, et al (2016) Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun* **7**, 11078.
32. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
33. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293–W296.
34. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, O'Brien J, Djimde A, et al (2012) Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379.
35. Simpalipan P, Pattaradilokrat S, Harnyuttanakorn P (2018) Global sequence diversity of the lactate dehydrogenase gene in *Plasmodium falciparum*. *Malar J* **17**, 16.
36. Lumkul L, Sawaswong V, Simpalipan P, Kaewthamasorn M, Harnyuttanakorn P, Pattaradilokrat S (2018) Unraveling haplotype diversity of the apical membrane antigen-1 gene in *Plasmodium falciparum* populations in Thailand. *Korean J Parasitol* **56**, 153–165.
37. Pattaradilokrat S, Trakoolsoontorn C, Simpalipan P, Warrit N, Kaewthamasorn M, Harnyuttanakorn P (2018) Size and sequence polymorphisms in the glutamate-rich protein gene of the human malaria parasite *Plasmodium falciparum* in Thailand. *Parasit Vectors* **11**, 49.
38. Pattaradilokrat S, Sawaswong V, Simpalipan P, Kaewthamasorn M, Siripoon N, Harnyuttanakorn P (2016) Genetic diversity of the merozoite surface protein-3 gene in *Plasmodium falciparum* populations in Thailand. *Malar J* **15**, 517.
39. Sharma SK, Banyal HS (2011) Characterization of *Plasmodium berghei* glutathione synthetase. *Parasitol Int* **60**, 321–323.
40. Shafi S, Gupta S, Jain R, Shoaib R, Munjal A, Maurya P, Kumar P, Kalam Najmi A, et al (2023) Tackling the emerging artemisinin-resistant malaria parasite by modulation of defensive oxido-reductive mechanism via nitrofurantoin repurposing. *Biochem Pharmacol* **215**, 115756.

Appendix A. Supplementary data

Table S1 List of *P. falciparum* isolates and laboratory strains and genotypes of the *gs* gene. A total of 61 sequences from *P. falciparum* in Thailand were generated in the present study. The other 211 sequences were downloaded from PlasmoDB and NCBI database. EX indicates the 49 sequences with ambiguous nucleotide and were excluded from the analysis.

Continent	Country	Isolate	Haplotype	Indel	Accession ID	Accession ID
					NCBI	PlasmoDB
Asia	Thailand	T9/94RC17	*gs2*	R1	PQ753213	N/A
		K1CB1	*gs3*	R1	PQ753214	N/A
		K60	*gs4*	R3	PQ753215	N/A
		K64	*gs5*	R2	PQ753216	N/A
		K74	*gs6*	R3	PQ753217	N/A
		K165	*gs7*	R2	PQ753218	N/A
		K185	*gs8*	R2	PQ753219	N/A
		K195	*gs8*	R2	PQ753220	N/A
		K205	*gs2*	R1	PQ753221	N/A
		K215	*gs8*	R2	PQ753222	N/A
		K386	*gs9*	R1	PQ753223	N/A
		K389	*gs2*	R1	PQ753224	N/A
		K391	*gs8*	R2	PQ753225	N/A
		K392	*gs2*	R1	PQ753226	N/A
		K402	*gs10*	R1	PQ753227	N/A
		K403	*gs8*	R2	PQ753228	N/A
		RN19	*gs11*	R1	PQ753229	N/A
		RN26	*gs8*	R2	PQ753230	N/A
		RN28	*gs12*	R3	PQ753231	N/A
		RN36	*gs5*	R2	PQ753232	N/A
		RN63	*gs13*	R3	PQ753233	N/A
		RN66	*gs2*	R1	PQ753234	N/A
		RN68	*gs5*	R2	PQ753235	N/A
		RN70	*gs2*	R1	PQ753236	N/A
		RN72	*gs2*	R1	PQ753237	N/A
		RN122	*gs8*	R2	PQ753238	N/A
		RN129	*gs4*	R3	PQ753239	N/A
		RN130	*gs2*	R1	PQ753240	N/A
		RN131	*gs2*	R1	PQ753241	N/A
		RN133	*gs8*	R2	PQ753242	N/A
		MH6	*gs2*	R1	PQ753243	N/A
		MH7	*gs5*	R2	PQ753244	N/A
		MH10	*gs8*	R2	PQ753246	N/A
		MH11	*gs5*	R2	PQ753247	N/A
		MH18	*gs2*	R1	PQ753248	N/A
		MH20	*gs3*	R1	PQ753249	N/A
		MH24	*gs2*	R1	PQ753250	N/A
		MH28	*gs2*	R1	PQ753251	N/A
		MH32	*gs2*	R1	PQ753252	N/A
		MH50	*gs2*	R1	PQ753253	N/A
		MH61	*gs2*	R1	PQ753254	N/A
		MH66	*gs8*	R2	PQ753255	N/A
		TD508	*gs3*	R1	PQ753256	N/A
		TD510	*gs2*	R1	PQ753257	N/A
		TD515	*gs2*	R1	PQ753258	N/A
		TD529	*gs5*	R2	PQ753259	N/A
		TD530	*gs2*	R1	PQ753260	N/A
		TD531	*gs2*	R1	PQ753261	N/A
		TD533	*gs5*	R2	PQ753262	N/A
		TD542	*gs8*	R2	PQ753263	N/A
		TD554	*gs12*	R3	PQ753264	N/A
		TD556	*gs8*	R2	PQ753265	N/A
		UB7	*gs2*	R1	PQ753266	N/A
		UB14	*gs5*	R2	PQ753267	N/A
		UB22	*gs4*	R3	PQ753268	N/A
		UB27	*gs2*	R1	PQ753269	N/A
		UB28	*gs2*	R1	PQ753270	N/A
		UB50	*gs8*	R2	PQ753271	N/A
		UB51	*gs14*	R4	PQ753272	N/A
		UB52	*gs15*	R5	PQ753273	N/A
		UB59	*gs4*	R3	PQ753274	N/A
		T9/94	N/A*	N/A	N/A	PF3D7_0512200

* Sequences excluded from the analysis due to the presence of ambiguous nucleotides.

Table S1 Continue ...

Continent	Country	Isolate	Haplotype	Indel	Accession ID	Accession ID
					NCBI	PlasmoDB
Asia	Laos	Dd2	N/A*	N/A	DS016656	N/A
		Dd2-1	N/A*	N/A	N/A	PF3D7_0512200
		Dd2-2	N/A*	N/A	N/A	PF3D7_0512201
Asia	Vietnam	V1_S	N/A*	N/A	N/A	PF3D7_0512200
		Vietnam Oaknoll FVO	*gs2*	R1	AOPP01000952	N/A
		UGT.5	*gs17*	R1	KE124450	PF3D7_0512200
		NF135/5.C10	*gs19*	R1	KI926026	N/A
Asia	Malaysia	CAMP/Malaysia	*gs16*	R1	KI927480	PF3D7_0512200
Asia	India	ICH-CR14	*gs18*	R4	GG664981	N/A
Africa	The Gambia	BM_0008	N/A*	N/A	N/A	PF3D7_0512200
		BM_0009	N/A*	N/A	N/A	PF3D7_0512200
		RV_3600 :	*gs16*	R1	N/A	PF3D7_0512200
		RV_3606 :	N/A*	N/A	N/A	PF3D7_0512200
		RV_3610	*gs1*	R1	N/A	PF3D7_0512200
		RV_3611	*gs1*	R1	N/A	PF3D7_0512200
		RV_3614	*gs1*	R1	N/A	PF3D7_0512200
		RV_3630	*gs23*	R1	N/A	PF3D7_0512200
		RV_3635	*gs27*	R1	N/A	PF3D7_0512200
		RV_3637	*gs31*	R1	N/A	PF3D7_0512200
		RV_3642	*gs29*	R1	N/A	PF3D7_0512200
		RV_3650	*gs3*	R1	N/A	PF3D7_0512200
		RV_3655	*gs3*	R1	N/A	PF3D7_0512200
		RV_3671	*gs23*	R1	N/A	PF3D7_0512200
		RV_3672	*gs30*	R1	N/A	PF3D7_0512200
		RV_3673	*gs25*	R1	N/A	PF3D7_0512200
		RV_3675	*gs28*	R1	N/A	PF3D7_0512200
		RV_3687	*gs22*	R1	N/A	PF3D7_0512200
		RV_3695	*gs3*	R1	N/A	PF3D7_0512200
		RV_3696	*gs22*	R1	N/A	PF3D7_0512200
		RV_3701	*gs24*	R1	N/A	PF3D7_0512200
		RV_3702	*gs26*	R1	N/A	PF3D7_0512200
		RV_3703	*gs2*	R1	N/A	PF3D7_0512200
		RV_3708	*gs2*	R1	N/A	PF3D7_0512200
		RV_3714	*gs10*	R1	N/A	PF3D7_0512200
		RV_3717	*gs27*	R1	N/A	PF3D7_0512200
		RV_3721	*gs1*	R1	N/A	PF3D7_0512200
		RV_3729	*gs32*	R1	N/A	PF3D7_0512200
		RV_3730	*gs20*	R1	N/A	PF3D7_0512200
		RV_3731	N/A*	N/A	N/A	PF3D7_0512200
		RV_3735	*gs21*	R1	N/A	PF3D7_0512200
		RV_3736	*gs21*	R1	N/A	PF3D7_0512200
		RV_3737	*gs1*	R1	N/A	PF3D7_0512200
		RV_3739	*gs16*	R1	N/A	PF3D7_0512200
		RV_3740	*gs19*	R1	N/A	PF3D7_0512200
		RV_3741	N/A*	N/A	N/A	PF3D7_0512200
		RV_3764	N/A*	N/A	N/A	PF3D7_0512200
		RV_3766	N/A*	N/A	N/A	PF3D7_0512200
		RV_3769	*gs23*	R1	N/A	PF3D7_0512200
		TRIPS_301	*gs22*	R1	N/A	PF3D7_0512200
		TRIPS_303	*gs10*	R1	N/A	PF3D7_0512200
		TRIPS_331	N/A*	N/A	N/A	PF3D7_0512200
		TRIPS_355	N/A*	N/A	N/A	PF3D7_0512200
		TRIPS_364	N/A*	N/A	N/A	PF3D7_0512200
		TRIPS_373	*gs33*	R1	N/A	PF3D7_0512200
		TRIPS_410	*gs1*	R1	N/A	PF3D7_0512200
		TRIPS_433	*gs16*	R1	N/A	PF3D7_0512200
		TRIPS_437	*gs36*	R1	N/A	PF3D7_0512200
		TRIPS_440	*gs3*	R1	N/A	PF3D7_0512200
		TRIPS_456	*gs2*	R1	N/A	PF3D7_0512200
		TRIPS_461	N/A*	N/A	N/A	PF3D7_0512200
		TRIPS_467	*gs20*	R1	N/A	PF3D7_0512200

* Sequences excluded from the analysis due to the presence of ambiguous nucleotides.

Table S1 Continue ...

Continent	Country	Isolate	Haplotype	Indel	Accession ID	Accession ID
					NCBI	PlasmoDB
		TRIPS_470	N/A*	N/A	N/A	PF3D7_0512200
		TRIPS_474	*gs1*	R1	N/A	PF3D7_0512200
		TRIPS_480	*gs34*	R1	N/A	PF3D7_0512200
		TRIPS_482	*gs22*	R1	N/A	PF3D7_0512200
		TRIPS_487	*gs16*	R1	N/A	PF3D7_0512200
		TRIPS_490	*gs29*	R1	N/A	PF3D7_0512200
		TRIPS_499	*gs35*	R1	N/A	PF3D7_0512200
		TRIPS_501	N/A*	N/A	N/A	PF3D7_0512200
		TRIPS_504	*gs3*	R1	N/A	PF3D7_0512200
		TRIPS_700	*gs1*	R1	N/A	PF3D7_0512200
		TRIPS_704	*gs22*	R1	N/A	PF3D7_0512200
		TRIPS_708	*gs16*	R1	N/A	PF3D7_0512200
		TRIPS_759	*gs21*	R1	N/A	PF3D7_0512200
Africa	Senegal	SenT001.07	N/A*	N/A	N/A	PF3D7_0512200
		SenT001.08	*gs45*	R1	N/A	PF3D7_0512200
		SenT002.07	*gs10*	R1	N/A	PF3D7_0512200
		SenT002.09	*gs1*	R1	N/A	PF3D7_0512200
		SenT015.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT015.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT018.09	*gs1*	R1	N/A	PF3D7_0512200
		SenT021.09	*gs10*	R1	N/A	PF3D7_0512200
		SenT022.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT024.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT029.09	*gs1*	R1	N/A	PF3D7_0512200
		SenT032.09	*gs47*	R1	N/A	PF3D7_0512200
		SenT033.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT033.09	*gs3*	R1	N/A	PF3D7_0512200
		SenT037.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT042.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT044.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT044.10	*gs3*	R1	N/A	PF3D7_0512200
		SenT046.10	*gs2*	R1	N/A	PF3D7_0512200
		SenT047.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT061.09	*gs48*	R5	N/A	PF3D7_0512200
		SenT063.07	N/A*	N/A	N/A	PF3D7_0512200
		SenT064.10	N/A*	N/A	N/A	PF3D7_0512200
		SenT066.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT067.09	*gs37*	R1	N/A	PF3D7_0512200
		SenT075.10	N/A*	N/A	N/A	PF3D7_0512200
		SenT077.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT084.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT087.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT090.08	*gs43*	R1	N/A	PF3D7_0512200
		SenT090.09	*gs16*	R1	N/A	PF3D7_0512200
		SenT092.08	*gs10*	R1	N/A	PF3D7_0512200
		SenT093.09	*gs37*	R1	N/A	PF3D7_0512200
		SenT094.09	*gs35*	R1	N/A	PF3D7_0512200
		SenT097.09	*gs3*	R1	N/A	PF3D7_0512200
		SenT101.09	*gs16*	R1	N/A	PF3D7_0512200
		SenT102.08	*gs16*	R1	N/A	PF3D7_0512200
		SenT104.07	*gs1*	R1	N/A	PF3D7_0512200
		SenT106.08	*gs1*	R1	N/A	PF3D7_0512200
		SenT110.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT111.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT112.09	N/A*	N/A	N/A	PF3D7_0512200
		SenT113.09	*gs10*	R1	N/A	PF3D7_0512200
		SenT119.09	*gs49*	R5	N/A	PF3D7_0512200
		SenT123.09	*gs3*	R1	N/A	PF3D7_0512200
		SenT127.09	*gs1*	R1	N/A	PF3D7_0512200
		SenT128.08	*gs1*	R1	N/A	PF3D7_0512200
		SenT135.09	*gs3*	R1	N/A	PF3D7_0512200
		SenT137.08	*gs1*	R1	N/A	PF3D7_0512200
		SenT137.09	*gs3*	R1	N/A	PF3D7_0512200

* Sequences excluded from the analysis due to the presence of ambiguous nucleotides.

Table S1 Continue ...

Continent	Country	Isolate	Haplotype	Indel	Accession ID	Accession ID
					NCBI	PlasmoDB
		SenT139.08	*gs1*	R1	N/A	PF3D7_0512200
		SenT140.08	*gs16*	R1	N/A	PF3D7_0512200
		SenT142.09	*gs44*	R1	N/A	PF3D7_0512200
		SenT144.08	*gs46*	R1	N/A	PF3D7_0512200
		SenT145.08	*gs1*	R1	N/A	PF3D7_0512200
		SenT149.09	*gs26*	R1	N/A	PF3D7_0512200
		SenT150.09	*gs10*	R1	N/A	PF3D7_0512200
		SenT151.09	*gs3*	R1	N/A	PF3D7_0512200
		SenT170.08	*gs43*	R1	N/A	PF3D7_0512200
		SenT175.08	*gs10*	R1	N/A	PF3D7_0512200
		SenT179.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT180.08	*gs47*	R1	N/A	PF3D7_0512200
		SenT190.08	*gs22*	R1	N/A	PF3D7_0512200
		SenT197.08	*gs16*	R1	N/A	PF3D7_0512200
		SenT224.08	*gs1*	R1	N/A	PF3D7_0512200
		SenT227.08	*gs3*	R1	N/A	PF3D7_0512200
		SenT230.08	N/A*	N/A	N/A	PF3D7_0512200
		SenT235.08	*gs43*	R1	N/A	PF3D7_0512200
		SenT238.08	*gs3*	R1	N/A	PF3D7_0512200
Africa	Mali	303.1	*gs1*	R1	N/A	PF3D7_0512200
		309.1	*gs37*	R1	N/A	PF3D7_0512200
		318.1	*gs37*	R1	N/A	PF3D7_0512200
		326.1	*gs26*	R1	N/A	PF3D7_0512200
		327.1	*gs1*	R1	N/A	PF3D7_0512200
		365.1	N/A*	N/A	N/A	PF3D7_0512200
		366.1	*gs3*	R1	N/A	PF3D7_0512200
		377.1	*gs16*	R1	N/A	PF3D7_0512200
		383.1	*gs38*	R1	N/A	PF3D7_0512200
		397.1	*gs1*	R1	N/A	PF3D7_0512200
		398.1	*gs1*	R1	N/A	PF3D7_0512200
		58.1	*gs39*	R1	N/A	PF3D7_0512200
		M113-A	*gs3*	R1	N/A	PF3D7_0512200
		M219-D	*gs22*	R1	N/A	PF3D7_0512200
		ML01	N/A*	N/A	N/A	PF3D7_0512200
		PS149	*gs41*	R1	N/A	PF3D7_0512200
		PS170	*gs16*	R1	N/A	PF3D7_0512200
		PS183	*gs1*	R1	N/A	PF3D7_0512200
		PS186	*gs3*	R1	N/A	PF3D7_0512200
		PS189	*gs1*	R1	N/A	PF3D7_0512200
		PS206_E11	*gs2*	R1	N/A	PF3D7_0512200
		PS250	*gs42*	R1	N/A	PF3D7_0512200
		SEV2_H5	*gs53*	R1	N/A	PF3D7_0512200
Africa	Uganda	7G8_2	N/A*	N/A	N/A	PF3D7_0512200
		UGK_396.1	*gs22*	R1	N/A	PF3D7_0512200
		UGK_408.2	*gs50*	R1	N/A	PF3D7_0512200
		UGK_432.4	*gs16*	R1	N/A	PF3D7_0512200
		UGK_443.2	*gs51*	R1	N/A	PF3D7_0512200
		UGK_659.1	N/A*	N/A	N/A	PF3D7_0512200
		UGK_661.1	N/A*	N/A	N/A	PF3D7_0512200
		UGK_674.4	N/A*	N/A	N/A	PF3D7_0512200
		UGK_707.3	N/A*	N/A	N/A	PF3D7_0512200
		UGK_730.2	*gs1*	R1	N/A	PF3D7_0512200
		UGK_815.1	N/A*	N/A	N/A	PF3D7_0512200
		Palo Alto	*gs16*	R1	KI927385	N/A
Africa	Ghana	GB4	*gs52*	R1	N/A	PF3D7_0512200
Africa	Tanzania	Tanzania 2000708	*gs54*	R1	KI926318	N/A
South America	Brazil	7G8	*gs55*	R1	KE123592	PF3D7_0512200
		CS2	*gs2*	R1	N/A	PF3D7_0512200
		IT	*gs2*	R1	PFIT_0512300	PF3D7_0512200
		SantaLucia_SalvadorI	gs2	R1	ABHA02000770	PF3D7_0512200

* Sequences excluded from the analysis due to the presence of ambiguous nucleotides.

Table S1 Continue ...

Continent	Country	Isolate	Haplotype	Indel	Accession ID	Accession ID
					NCBI	PlasmoDB
South America	French Guiana	G224	*gs55*	R1	N/A	PF3D7_0512200
		H209	*gs55*	R1	N/A	PF3D7_0512200
		M271-D	*gs1*	R1	N/A	PF3D7_0512200
		M312-D	*gs1*	R1	N/A	PF3D7_0512200
		N011-A	*gs55*	R1	N/A	PF3D7_0512200
		N023-A	*gs3*	R1	N/A	PF3D7_0512200
		N071-I	*gs3*	R1	N/A	PF3D7_0512200
		N164-A	*gs55*	R1	N/A	PF3D7_0512200
		N497-C	gs3	R1	N/A	PF3D7_0512200
		N579-A	*gs1*	R1	N/A	PF3D7_0512200
		O079-B	*gs22*	R1	N/A	PF3D7_0512200
		O141-A	*gs55*	R1	N/A	PF3D7_0512200
		O222-A	*gs1*	R1	N/A	PF3D7_0512200
		O306-A	*gs55*	R1	N/A	PF3D7_0512200
		O314	*gs1*	R1	N/A	PF3D7_0512200
		P164-C	*gs55*	R1	N/A	PF3D7_0512200
		P167-B	*gs55*	R1	N/A	PF3D7_0512200
		P237-C	*gs55*	R1	N/A	PF3D7_0512200
		P241-D	*gs55*	R1	N/A	PF3D7_0512200
		PS097	*gs1*	R1	N/A	PF3D7_0512200
		PS103	*gs40*	R1	N/A	PF3D7_0512200
		PS122_G11	*gs37*	R1	N/A	PF3D7_0512200
South America	Honduras	HB3	*gs3*	R1	CH671920	PF3D7_0512200
South America	El Salvador	P196J3-C	*gs1*	R1	N/A	PF3D7_0512200
Unknown		3D7	*gs1*	R1	XM_001351643	PF3D7_0512200
		708A	N/A*	N/A	N/A	PF3D7_0512200

* Sequences excluded from the analysis due to the presence of ambiguous nucleotides.

Table S2 Nucleotide sequences of primers for PCR amplification and DNA sequencing of the *gs* gene. Nucleotide positions corresponded to positions in the *gs* gene of *P. falciparum* strain 3D7 [16].

Primer name	Primer sequences (5′–)	Estimated Tm (°C)
GS1F	ATGGAAAGAAAGGTAGATGAGTTTTA	61
GS1941R	CCCCTCAATGTTCAGTTAAAAAAAAAGAATCC	70
GS712F	CCATATATAGGAAATGAAGAGGAACA	62
GS1316F	CTTTACCTTATCAACTGTTGGTTCG	64
GS1560R	CCATGTAAATTATTTTTCCCTCCTTCT	64
GS1074R	CCAGGTTTGAATATCTCATTATAAGGA	63
GS516R	CGATTAGTCATATAATCAGATCTTCCT	60

Table S3 Polymorphic nucleotide positions in the *gs* gene of *P. falciparum*. The first codon is the sequence of the reference strain 3D7. A region at the nucleotides positions 412 and 480 contains size polymorphism in nucleotide repeats contributing to 5 sequence repeat patterns (see Fig. 1).

Nucleotide	Polymorphism		Amino acid	Substitution
position	nucleotide	amino acid substitution	position	frequency (%)
18	GAT/C	D/D	6	80.27/19.73
21	GAG/A	E/E	7	80.27/19.73
31*	G/TTT	V/F	11	99.6/0.4
41	AA/GA	K/R	14	99.6/0.4
101	AG/AG	R/K	34	96.86/3.14
143	AC/AT	T/N	48	99.6/0.4
207	AAC/G	N/K	69	99.1/0.9
239	CA/CT	H/P	80	99.6/0.4
256	C/TTA	L/L	86	99.1/0.9
331	C/GTA	L/V	111	99.6/0.4
354	ATT/A	I/I	118	95.07/4.93
359	GA/TA	E/V	120	96.41/3.59
402	AAG/T	K/N	134	99.6/0.4
432*	GAT/C	D/D	144	98.65/1.35
439*	G/AAC/T	D/N	147	98.65/1.35
441*				98.65/1.35
472	A/TGT	S/C	158	99.6/0.4
502	A/GTT	I/V	168	99.6/0.4
513	GTT/C	V/V	171	99.6/0.4
558	ATA/C	I/I	186	99.6/0.4
565	A/GAG	K/E	189	99.6/0.4
599	AA/CA	K/T	200	99.6/0.4
762	GAT/A	D/E	254	99.6/0.4
811	A/GAA	K/E	271	97.76/2.24
888*	CAT/C	H/H	296	99.6/0.4
981	ATT/G	I/M	327	99.1/0.9
1024	G/AAT	D/N	342	99.6/0.4
1086	GAG/C	E/D	362	97.76/2.24
1105	C/ATT	L/I	369	99.6/0.4
1126*	C/ATA	L/I	376	65.92/34.08
1138	T/CTA	L/L	380	98.65/1.35
1233	GCA/T	A/A	411	98.65/1.35
1278	AGA/G	R/R	426	99.6/0.4
1426	G/AAA	E/K	476	99.6/0.4
1438*	A/GAA	K/E	480	99.6/0.4
1448	AC/TG	T/M	483	99.6/0.4
1470	GCC/T	A/A	490	99.6/0.4
1545	TTA/G	L/L	515	98.65/1.35
1611	AAG/T	K/N	537	99.6/0.4
1908*	AAG/T	K/N	636	37.22/62.78
1956*	TTA/T	L/F	652	98.65/1.35

* Indicates the polymorphic sites detected in *gs* gene of *P. falciparum* in Thailand.

Table S4 Frequency and polymorphic SNPs in the 55 haplotypes in the *gs* gene of *P. falciparum*.

Haplotype		Nucleotide position																																									
Name	N	18	21	31	41	101	143	207	239	256	331	354	359	402	432	439	441	R	472	502	513	558	565	599	762	811	888	981	1024	1086	1105	1126	1138	1233	1278	1426	1438	1448	1470	1545	1611	1908	1956
gs1	34	GAT	GAG	GTT	AAA	AGG	ACT	AAC	CAT	CTA	CTA	ATT	GAG	AAG	GAT	GAC		R1	AGT	ATT	GTT	ATA	AAG	AAA	GAT	AAA	CAT	ATT	GAT	GAG	CTT	CTA	TTA	GCA	AGA	GAA	AAA	ACG	GCC	TTA	AAG	AAG	TTA
*gs2**	32																															ATA										AAT	
*gs3**	24																																									AAT	
*gs4**	4																	R3														ATA										AAT	
*gs5**	8																	R2																								AAT	
*gs6**	1																	R3																								AAT	TTT
*gs7**	1																	R2																								AAT	TTT
*gs8**	13																	R2														ATA										AAT	
*gs9**	1																																									AAT	TTT
*gs10**	9																															ATA											
*gs11**	1																															ATA					GAA					AAT	
*gs12**	2																	R3																								AAT	
*gs13**	1			TTT														R3																								AAT	
*gs14**	1																	R4									CAC					ATA										AAT	
*gs15**	1														GAC	AAT		R5														ATA										AAT	
gs16	15	GAC	GAA																																								
gs17	1				AGA																											ATA										AAT	
gs18	1												GTG					R4												GAC													
gs19	2	GAC	GAA																													ATA										AAT	
gs20	2	GAC	GAA					AAG																																		AAT	
gs21	3																									GAA																	
gs22	9	GAC	GAA																																							AAT	
gs23	3	GAC	GAA																															GCT								AAT	
gs24	1																														ATT												
gs25	1	GAC	GAA																											GAC													
gs26	3	GAC	GAA																													ATA											
gs27	2																																CTA									AAT	
gs28	1																													GAC													
gs29	2																									GAA																AAT	
gs30	1																											ATG															
gs31	1																											ATG				ATA										AAT	
gs32	1																								GAA								CTA								AAT	AAT	
gs33	1																																						GCT			AAT	
gs34	1																				GTC											ATA											
gs35	2									TTA																																	
gs36	1	GAC	GAA																											GAC	Ÿ	ATA										AAT	
gs37	5					AAG							GTG																													AAT	
gs38	1																					ATC										ATA											
gs39	1																							ACA																		AAT	
gs40	1					AAG							GTG																														
gs41	1																																		AGG							AAT	
gs42	1																																			AAA							
gs43	3	GAC	GAA																																					TTG			
gs44	1	GAC	GAA										GTG																														
gs45	1																													GAC												AAT	
gs46	1						AAT																																			AAT	
gs47	2														GAC																	ATA										AAT	
gs48	1					AAG									GAC	AAT		R5																									
gs49	1	GAC	GAA													AAT		R5														ATA										AAT	
gs50	1	GAC	GAA											AAT																		ATA											
gs51	1	GAC	GAA								GTA									GTT												ATA										AAT	
gs52	1								CCT																				AAT														
gs53	1																						GAG																				
gs54	1	GAC	GAA																TGT																							AAT	
gs55	11											ATA																														AAT	

_ Indicate identical nucleotides to *gs1* haplotype.
* Indicate haplotype of gs detected in *P. falciparum* population in Thailand.