



Spatial Variation in Genetic Diversity and Natural Selection on the Thrombospondin-Related Adhesive Protein Locus of *Plasmodium vivax* (*PvTRAP*)

Rattiporn Kosuwin¹, Chaturong Putaporntip¹, Hiroshi Tachibana², Somchai Jongwutiwes^{1*}

¹ Molecular Biology of Malaria and Opportunistic Parasites Research Unit, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand,

² Department of Infectious Diseases, Tokai University School of Medicine, Kanagawa, Japan

Abstract

Thrombospondin-related adhesive protein (TRAP) of malaria parasites is essential for sporozoite motility and invasions into mosquito's salivary gland and vertebrate's hepatocyte; thereby, it is a promising target for pre-erythrocytic vaccine. TRAP of *Plasmodium vivax* (*PvTRAP*) exhibits sequence heterogeneity among isolates, an issue relevant to vaccine development. To gain insights into variation in the complete *PvTRAP* sequences of parasites in Thailand, 114 *vivax* malaria patients were recruited in 2006–2007 from 4 major endemic provinces bordering Myanmar (Tak in the northwest, $n=30$ and Prachuap Khirikhan in the southwest, $n=25$), Cambodia (Chanthaburi in the east, $n=29$) and Malaysia (Yala and Narathiwat in the south, $n=30$). In total, 26 amino acid substitutions were detected and 9 of which were novel, resulting in 44 distinct haplotypes. Haplotype and nucleotide diversities were lowest in southern *P. vivax* population while higher levels of diversities were observed in other populations. Evidences of positive selection on *PvTRAP* were demonstrated in domains II and IV and purifying selection in domains I, II and VI. Genetic differentiation was significant between each population except that between populations bordering Myanmar where transmigration was common. Regression analysis of pairwise linearized *Fst* and geographic distance suggests that *P. vivax* populations in Thailand have been isolated by distance. Sequence diversity of *PvTRAP* seems to be temporally stable over one decade in Tak province based on comparison of isolates collected in 1996 ($n=36$) and 2006–2007. Besides natural selection, evidences of intragenic recombination have been supported in this study that could maintain and further generate diversity in this locus. It remains to be investigated whether amino acid substitutions in *PvTRAP* could influence host immune responses although several predicted variant T cell epitopes drastically altered the epitope scores. Knowledge on geographic diversity in *PvTRAP* constitutes an important basis for vaccine design provided that vaccination largely confers variant-specific immunity.

Citation: Kosuwin R, Putaporntip C, Tachibana H, Jongwutiwes S (2014) Spatial Variation in Genetic Diversity and Natural Selection on the Thrombospondin-Related Adhesive Protein Locus of *Plasmodium vivax* (*PvTRAP*). PLoS ONE 9(10): e110463. doi:10.1371/journal.pone.0110463

Editor: Georges Snounou, Université Pierre et Marie Curie, France

Received: June 6, 2014; **Accepted:** September 4, 2014; **Published:** October 21, 2014

Copyright: © 2014 Kosuwin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files. All nucleotide sequences are available from the GenBank Database under the accession numbers KJ807186 to KJ807299.

Funding: Ratchadapiseksompotch Fund, Faculty of Medicine, Chulalongkorn University (Grant No. RA56/059) to SJ. Financial support from The Thailand Research Fund through The Royal Golden Jubilee Ph.D. Program (Grant PHD/0278/2552) to RK and SJ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: jongwutiwes@gmail.com

Introduction

In low- and middle-income countries in tropical areas, malaria remains one of the leading ten causes of morbidity and mortality, resulting in an estimated economic loss of nearly 40 million disability-adjusted life-years (DALYs) [1]. Although *Plasmodium falciparum* is the most pernicious and prevalent species, the significance of *P. vivax* should not be underappreciated because it can cause chronic relapsing illness due to reactivation of hypnozoites and it can potentially lead to severe complications similar to those caused by *P. falciparum* [2]. Furthermore, *vivax* malaria occupies a broader geographical range than *falciparum* malaria whilst the emergences of chloroquine- and primaquine-resistant *P. vivax* are of particular concern if they would be widespread as chloroquine-resistant *P. falciparum* [3]. The progress of malaria control and elimination has been hampered by not only the emergence of drug resistant parasites but also insecticide

resistant mosquito vectors; thereby, development of a malaria vaccine offers an important preventive measure [4]. Because *P. vivax* and *P. falciparum* co-circulate in several endemic areas outside Africa where co-infections of both species are not uncommon [5], effective malaria control requires vaccines against both species.

To date, malarial circumsporozoite (CS) protein is a prime candidate for pre-erythrocytic vaccine development [4]. Although CSP-derived immunogens could elicit immunity against sporozoites, the subunit vaccines derived from this molecule such as RTS, S recombinant vaccine against *P. falciparum* has resulted in limited clinical efficacy in field studies [6]. Because vaccines derived from irradiation-attenuated or live sporozoites consistently outperform vaccines incorporating single sporozoite proteins, a more effective pre-erythrocytic stage vaccine may require combination of multiple protective immunogens [7,8]. In a murine model, co-immunization of CSP with thrombospondin-related

adhesive protein (TRAP), a protein mobilized from microneme to the surface of sporozoite [9], has conferred complete protection against parasite challenge whereas vaccination using each of these immunogens could elicit only partial protection [10]. It is important to note that TRAP-specific CD8+ T lymphocytes are prime mediators for protection against sporozoite challenge in mouse vaccination trials, resulting in significant reduction in liver stage parasites [11]. Furthermore, seroepidemiological study has shown that anti-TRAP antibodies were negatively correlated with parasite density among infected individuals in malaria endemic areas [12].

TRAP has been shown to mediate gliding motility and invasion processes of malarial sporozoites into vertebrate's hepatocyte and mosquito's salivary gland [13,14]. TRAP contains a hydrophobic N-terminal peptide (domain I), an integrin-like magnesium binding (or von Willebrand factor) A domain (domain II), thrombospondin type I repeats (domain III), an acidic proline/asparagine-rich region (domain IV), hydrophobic transmembrane domain (domain V) and a cytoplasmic tail (domain VI) [15,16]. The locomotion of sporozoites is mediated by the subpellicular actomyosin system that linked to the cytoplasmic tail of TRAP [17]. Despite functional importance of TRAP in parasite survival, analysis of the TRAP loci of *P. falciparum* (PfTRAP) and of *P. vivax* (PvTRAP) from clinical isolates revealed microheterogeneity of sequence that has been maintained by positive selective pressure [18–20]. Although it remains to be explored whether polymorphism in T cell epitopes of malarial TRAP could alter host cell immune recognition as that observed in CSP of *P. falciparum* [21], a rationale design of a malaria subunit vaccine targeting TRAP requires knowledge on the extent and pattern of sequence diversity in this molecule from diverse endemic areas.

Both *P. vivax* and *P. falciparum* have been circulating in Thailand with almost comparable prevalence since the past 2 decades. However, regional prevalence of *P. vivax* relative to *P. falciparum* seems to differ across major endemic areas of Thailand [22,23]. Our recent studies have shown that *P. vivax* populations in this country exhibited spatial variation in the extent of sequence diversity of major malaria vaccine candidate loci [24,25]. An effective malaria vaccine design undoubtedly requires knowledge on this issue to circumvent possible vaccine escape variants. Herein, we analyzed sequence diversity in the PvTRAP locus of clinical isolates from 4 major malaria endemic areas of Thailand.

Materials and Methods

Human Ethics Statement

The study protocol was approved by the Institutional Review Board on Human Research of Faculty of Medicine, Chulalongkorn University (IRB303/56). Written informed consent was obtained from all participants.

Parasite populations

Blood samples were obtained from 114 *P. vivax*-infected patients in northwestern (Tak province, $n = 30$), southwestern (Prachuap Khirikhan province, $n = 25$), eastern (Chanthaburi province, $n = 29$) and southern (Yala and Narathiwat provinces, $n = 30$) Thailand during 2006 and 2007. Venous blood samples were taken from each patient and were preserved in EDTA anticoagulant. These isolates have been previously determined to contain single clone infection based on sequence analysis of the merozoite surface protein-1 gene (*Pvmsp-1*) as described previously [24,26].

DNA extraction, amplification and sequencing

DNA was prepared from blood samples using QIAamp kit (Qiagen, Hilden, Germany). The complete *PvTRAP* sequence was amplified by nested PCR using a forward outer primer (PvTRAP-F0: 5'-ATGTGTGTACATTTGCGTATG-3', nucleotides -442 to -422 before the start codon of the Salvador I sequence, GenBank accession number XM_001614097) and a reverse outer primer (PvTRAP-R0: 5'-TCATGAAGTGGC-GAAACAAAC-3', nucleotides 1961 to 1968 from the start codon). The inner pairs of primers were PvTRAP-F1 (5'-ATGTCTGTGTGAGCGCGCGGT-3', nucleotides -398 to -378) and PvTRAP-R1 (5'-TCATCAGAAGCAGTTCCAAG-3', nucleotides 1791 to 1810). PCR amplification was performed in a total volume of 30 μ l of the reaction mixture containing template DNA, 2.5 mM MgCl₂, 300 mM each deoxynucleoside triphosphate, 3 μ l of 10x ExTaq PCR buffer, 0.3 μ M of each primer and 1.25 units of ExTaq DNA polymerase (Takara, Seta, Japan). Thermal cycling profile for primary PCR included the preamplification denaturation at 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min, and a final extension at 72°C for 5 min. Nested PCR was done using the same thermal profile except that amplification was done for 25 cycles. DNA amplification was carried out by using a GeneAmp 9700 PCR thermal cycler (Applied Biosystems, Foster City, CA). ExTaq DNA polymerase reportedly possesses efficient 5'→3' exonuclease activity to increase fidelity and no strand displacement (Takara, Japan). The PCR product was examined by electrophoresis in a 1% agarose gel and was purified by using QIAquick PCR purification kit (QIAGEN, Germany). The PvTRAP sequences were determined directly and bi-directionally from PCR-purified templates and were performed on an ABI3100 Genetic Analyzer using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Whenever singleton substitution occurred, sequence was re-determined using PCR products from two independent amplifications from the same DNA template. Sequences in this study have been deposited in the GenBank Database under the accession numbers KJ807186 to KJ807299.

Statistical analyses

Sequence alignment with the Clustal X program [27] was done by using the *PvTRAP* sequence of the Salvador I strain as reference. The *PvTRAP* sequences of our previous Thai and Brazilian isolates previously published [19] and those reported by others [28] were included for analysis. For interspecific comparison, all *PvTRAP* sequences were aligned with the TRAP sequences of *P. knowlesi* (*PkTRAP*) and *P. cynomolgi* (*PcyTRAP*) (GenBank accession numbers XM_002259951 and XM_004223600, respectively). Haplotype diversity (h) and its sampling variance were computed according to equations 8.4 and 8.12 but replacing $2n$ by n [29]. Nucleotide diversity (π) was computed from the average number of pairwise sequence differences in the sample [29] using the MEGA version 6 program [30]. Evidence of genetic recombination was determined by 2 different approaches: (i) estimation of the minimum number of recombination events (R_m) that can be parsimoniously inferred from a sample of sequences with Monte Carlo simulations [31] using the DnaSP version 5 software [32] and (ii) searching for recombination breakpoints by phylogenetic approach by the Genetic Algorithm Recombination Detection (GARD) method. Goodness of fit for the GARD method was calculated by Akaike Information Criterion derived from a maximum likelihood model fit to each segment (AICc) using the HyPhy package [33,34].

The rates of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N)

were computed by using Nei and Gojobori's method [35] with Juke and Cantor correction [36] and their standard errors of these parameters were estimated by the bootstrap method with 1,000 pseudoreplicates as implemented in the MEGA 6.0 program [30]. A statistical significance level of differences in d_S and d_N was set at 5%.

Tests for departure from neutral evolution based on intraspecific comparisons were done by Tajima's D , Fu and Li's D^* and F^* statistics. Tajima's D test determines the differences between the average number of nucleotide differences and an estimate of θ from the number of segregating sites where $\theta = 4Ne\mu$, in which Ne and μ are the effective population size and the mutation rate, respectively [37]. Fu and Li's statistics measure the differences between the number of singletons and the total number of mutations (D^* test) and the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences (F^* test) [38]. To examine whether deviation from neutrality occurs in particular fragments of *PvTRAP*, a sliding window analysis of Tajima's D across the coding sequences was done using a window size of 250 bp and a step size of 10 bp [32].

Evidence of departure from neutrality based on interspecific comparison included Fu and Li's D and F tests, and McDonald-Kreitman tests. Fu and Li's D statistics measure the differences between the number of mutations in external branches of the genealogy and the number of mutations while their F statistic compares the differences between the number of mutations in external branches of the genealogy and the average number of nucleotide differences between pairs of sequences [38]. Coalescence simulation with 10,000 pseudoreplicates was done to estimate significance levels of these parameters as implemented in the DnaSP program [32]. The McDonald-Kreitman test contrasts intraspecific polymorphism with interspecific divergence of closely related species. Statistical departure from neutral expectation was computed by Fischer's exact test [39].

Tests of departure from neutrality at specific codons were performed based on estimation of the global ratio of the rate of non-synonymous to synonymous substitutions (dN/dS) across the *PvTRAP* gene. The single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), random effects likelihood (REL) and mixed effects model of evolution (MEME) methods implemented in the HyPhy package [33] were used for analysis. SLAC model is highly conservative based on the maximum likelihood reconstruction of the ancestral sequences and the counts of synonymous and nonsynonymous changes at each codon position in a phylogeny under the assumption of neutral evolution. FEL model compares the ratio of nonsynonymous to synonymous substitution on a site-by-site basis, without assuming an *a priori* distribution of rates across sites whereas REL model first fits a distribution of rates across sites and then infers the substitution rate for individual sites. MEME algorithm detects codons under episodic positive selection unmasked by the abundance of purifying selection along the lineages [40]. Significance level settings for SLAC, FEL and MEME were p values <0.1 and Bayes Factor >1000 for REL followed the default values available on the Datamonkey Web Server [41].

The population genetic structure was analyzed by molecular variance approach (AMOVA) using the Arlequin 3.5 software which is similar to the Weir and Cockerham's method but it takes into account the number of mutations between haplotypes [42]. The implemented algorithm calculates the fixation index F_{ST} identical to the weighted average F -statistic over loci, θ_w [43] and the significance levels of the fixation indices were estimated by a non-parametric permutation [42]. Pairwise Slakin's linearized F_{ST}

was calculated based on $F_{ST}/(1-F_{ST})$ to address the relationship between the magnitude of population differentiation and distance between endemic areas [44].

The genetic structure was also determined by STRUCTURE 2.3.4 program that deploys the Bayesian model-based clustering approach [45]. The most probable number of populations (K) was estimated using an admixture model. All sample data were run for values $K=1-10$, each with a total of 15 iterations. We used 500,000 Markov Chain Monte Carlo generations for each run following a burn-in of 50,000 steps. The most likely number K in the data was estimated by calculating ΔK values [46] and identifying the K value that maximizes the log probability of data, $\ln P(D)$ [45] as implemented in STRUCTURE HARVESTER program [47].

The phylogenetic tree of the *PvTRAP* haplotypes was constructed by using the Maximum Likelihood method based on the model with the lowest Bayesian Information Criterion (BIC) scores [30]. All positions containing gaps and missing data were eliminated. Initial tree(s) for the heuristic search were obtained by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The final tree topology was selected based on superior log likelihood value as implemented in the MEGA version 6 program [30].

Amino acid properties were characterized based on polarity and charge. Polarity of each residue is categorized as polar (S, Y, C, W, H, Q, T and N) or nonpolar (the remainder). Charge property includes positive (H, R and K), negative (D and E) or neutral (the remainder). Prediction of HLA-I binding peptides was based on the method taken into account proteasomal C terminal cleavage and transporter associated with antigen processing (TAP) transport efficiency. The predicted scores were determined by the NetCTL program [48].

Results

Genetic diversity in *PvTRAP*

Of 114 complete *PvTRAP* sequences, 109 isolates contained 1668 bp and 5 isolates from Tak province whose sequences were identical had 1677 bp. Sequence alignment with the Salvador I strain revealed 36 nucleotide substitutions within 33 codons, resulting in 26 nonsynonymous and 7 synonymous changes. Comparing with previous data from Brazilian, Thai [19], South Korean isolates [28], 9 new amino acid substitutions were identified (Table 1). Like our previous findings, the majority of amino acid changes occurred in domains II and IV [19]. Three codons contained more than two substituted nucleotides: codon 206 (after the Salvador 1 sequence) having 4 different substitutions: c.616C>T (P206S), c.617C>G (P206S), c.617C>T (P206L) and c.616_617delinsTT (P206L), codon 268 having double mutations: c.802_803delinsGG (K268G), and codon 424 containing 2 changes: c.1271A>C (N424T) and c.1272C>A (N424K). Among isolates examined herein, no discernible bias in frequency of nucleotide substitutions in each position of the codons occurred, i.e. 11, 11 and 13 substitutions in the first, second and third positions of the codons, respectively. In total, 44 distinct haplotypes were identified in this analysis, one of which had a duplication of 9 nucleotides encoding PDS in domain IV as previously noted [19] and responsible for size variation (Tables 1 and S1).

The distribution of *PvTRAP* haplotypes in southern Thailand (Yala and Narathiwat provinces) was rather skewed towards few haplotypes as shown by a relatively lower value of haplotype diversity ($h = 0.653 \pm 0.063$) whereas parasite populations from

Table 1. Distribution of PvTRAP haplotypes from 4 major malaria endemic areas of Thailand.

Haplotype	Amino acid residue																Prachuap Khirikhan	Chanthaburi	Yala & Narathiwat	Total					
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1									
#1	S	S	E	.	.	I	T	K	.	I	.	G	.	R	.	R	.	D	.	3	4	-	-	-	7
#2	S	S	.	.	I	T	K	.	I	.	K	G	.	R	.	R	.	D	.	3	-	-	-	-	3
#3	S	S	E	.	I	T	K	.	I	.	K	G	.	R	.	R	.	.	G	.	2	-	-	-	2
#4	S	S	E	.	I	T	K	.	I	.	G	.	R	.	R	.	K	T	.	A	1	-	-	-	1
#5	S	S	E	.	I	T	K	.	I	.	G	.	R	.	P	D	S	K	.	D	.	5	-	-	5
#6	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	.	G	.	3	1	-	-	-	4
#7	S	S	E	.	I	T	K	.	I	.	G	A	.	R	.	R	.	G	.	1	1	-	-	-	2
#8	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	K	.	D	.	1	-	-	-	1
#9	S	S	.	.	I	T	K	.	I	.	S	G	.	E	R	.	R	.	D	.	2	-	-	-	2
#10	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	.	.	1	1	-	-	2
#11	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	.	.	1	-	-	-	1
#12	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	.	.	1	-	-	-	1
#13	S	S	E	.	I	T	K	.	I	.	K	G	.	R	.	R	.	D	.	1	4	-	-	-	5
#14	S	S	E	.	I	T	K	.	I	.	K	G	.	R	.	R	.	K	.	A	2	-	-	-	2
#15	S	S	E	.	I	T	K	.	I	.	K	G	.	R	.	R	.	.	.	1	1	-	-	-	2
#16	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	D	.	1	1	-	-	2
#17	S	S	E	.	I	T	K	.	I	.	G	.	R	.	R	.	K	.	A	2	-	-	-	3	
#18	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	K	D	.	1	-	-	1
#19	S	S	.	.	I	T	K	.	I	.	R	.	G	.	R	.	R	.	D	.	1	-	-	-	1
#20	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	K	T	.	A	-	1	-	-	1
#21	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	T	D	.	1	-	-	1
#22	S	S	.	.	R	.	I	T	.	I	K	G	.	R	.	R	.	K	.	D	.	1	-	-	1
#23	S	S	.	.	I	T	K	.	I	.	L	.	G	.	R	.	R	.	.	.	1	-	-	-	1
#24	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	T	.	.	.	2	-	-	-	2
#25	S	S	.	.	I	T	K	.	I	.	K	G	.	R	.	R	.	K	.	.	1	-	-	-	1
#26	S	S	.	.	R	.	I	T	.	I	K	G	.	R	.	R	.	R	.	D	.	1	-	-	1
#27	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	D	.	A	1	-	-	1
#28	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	G	.	10	-	-	-	10
#29	S	S	.	.	I	T	K	.	I	.	G	.	R	.	R	.	R	.	.	G	.	1	-	-	1

Tak, Prachuap Khirikhan and Chanthaburi provinces had significantly higher haplotype diversity values, suggesting a more even distribution of haplotype frequencies in these latter endemic regions (Table 2). It is noteworthy that the predominant haplotypes differed between endemic areas. For examples, haplotypes #39 (n = 15) and #42 (n = 10) were exclusively found in Yala and Narathiwat provinces whereas haplotypes #28 (n = 10) and #30 (n = 8) were unique for Chanthaburi province. In contrast, several haplotypes were shared between isolates from Tak and Prachuap Khirikhan provinces (Table 1). The nucleotide diversity (π) of *PvTRAP* varied from 0.00117 ± 0.00014 (isolates from Yala and Narathiwat provinces) to 0.00272 ± 0.00032 (isolates from Prachuap Khirikhan province). The magnitudes of nucleotide diversity in this locus did not significantly differ among isolates from Tak, Prachuap Khirikhan and Chanthaburi provinces but these were significantly greater than that from Yala and Narathiwat provinces ($p < 10^{-4}$) (Table 2).

Population differentiation

We did not observe population differentiation between isolates from Tak and Prachuap Khirikhan provinces in which the F_{ST} values did not deviate significantly from zero, implying that they shared a common origin or a high genetic admixture. However, *P. vivax* populations from Tak and Prachuap Khirikhan provinces displayed the F_{ST} values that were significantly greater than zero when compared with populations from Chanthaburi province ($p < 10^{-5}$) as well as that from Yala and Narathiwat provinces ($p < 10^{-5}$). A significant departure from neutrality was also noted when isolates from Yala and Narathiwat provinces were compared with those from Chanthaburi province ($p < 10^{-5}$), implying genetic

differentiation between populations with minimum or absence of gene flow between these areas. Furthermore, the F_{ST} values between Tak isolates in the present study (collected in 2006–2007) and those collected in 1996 [19] did not differ significantly ($p = 0.23$), suggesting genetic stability in *P. vivax* population in this area (Table 3). On the other hand, *P. vivax* populations from Brazil, South Korea and Thailand were significantly different from one another, indicating strong differentiation and reproductive isolation among populations. These results were in line with analysis using STRUCTURE version 2.3.4 software in which the subpopulation likelihood reached a plateau from $K = 3$ to $K = 6$. Consistently, a greatest delta K value was also obtained at $K = 3$ (Figure 1). The genetic structure at $K = 3$ could differentiate Brazilian isolates from Thai and South Korean parasite populations while a separation of South Korean and southern Thai *P. vivax* populations were not perceived. Similar findings were observed at $K = 4$. A more concordant results between those obtained by STRUCTURE program and the F_{ST} parameters occurred when $K = 5$ in which isolates from Tak collected during different periods and from Prachuap Khirikhan were indistinguishable whereas South Korean and southern Thai populations were distinct (Figure 1). Regression analysis has shown positive correlation between pairwise linearized F_{ST} and the natural logarithm of linear distance between endemic areas (Mantel test, $r = 0.4619$, $p = 0.024$). Likewise, positive correlation was found when geographic linear distance between endemic areas was considered (Mantel test, $r = 0.4503$, $p = 0.028$), implying that the majority of *P. vivax* populations in Thailand were genetically isolated by distance (Figure 2).

Table 2. Variation in the PvTRAP sequences of Thai isolates from diverse endemic areas.

	Tak	Prachuap Khirikhan	Chanthaburi	Yala and Narathiwat	All
N	30	25	29	30	114
M	14	24	18	7	32
S	14	23	18	7	31
H	17	18	12	6	43
$h \pm S.D.$	0.956 ± 0.018	0.957 ± 0.027	0.818 ± 0.053	0.653 ± 0.063	0.954 ± 0.008
$\pi \pm S.D.$	0.00220 ± 0.00017	0.00272 ± 0.00032	0.00212 ± 0.00032	$0.00117 \pm 0.00014###$	0.00272 ± 0.00009
$d_s \pm S.E. (all)$	0.00018 ± 0.00017	0.00074 ± 0.00043	0.00069 ± 0.00052	0.00201 ± 0.00142	0.00118 ± 0.00081
Domain I	0.00000 ± 0.00000	0.00397 ± 0.00467	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00087 ± 0.00097
Domain II	0.00000 ± 0.00000	0.00029 ± 0.00027	0.00048 ± 0.00046	0.00176 ± 0.00130	0.00067 ± 0.00038
Domain III	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
Domain IV	0.00000 ± 0.00000	0.00002 ± 0.00002	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
Domain V	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
Domain VI	0.00250 ± 0.00281	0.00575 ± 0.00593	0.00721 ± 0.00781	0.01906 ± 0.00000	0.01245 ± 0.01187
$d_N \pm S.E. (all)$	$0.00278 \pm 0.00090^{**}$	$0.00327 \pm 0.00086^{**}$	$0.00252 \pm 0.00077^*$	0.00091 ± 0.00052	0.00314 ± 0.00090
Domain I	0.00000 ± 0.00000	0.00233 ± 0.00232	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00053 ± 0.00055
Domain II	$0.00264 \pm 0.00134^*$	$0.00428 \pm 0.00161^*$	$0.00406 \pm 0.00161^*$	0.00014 ± 0.00014	0.00396 ± 0.00173
Domain III	0.00423 ± 0.00426	0.00442 ± 0.00439	0.00067 ± 0.00071	0.00000 ± 0.00000	0.00261 ± 0.00266
Domain IV	$0.00381 \pm 0.00162^*$	$0.00314 \pm 0.00132^*$	$0.00248 \pm 0.00118^*$	0.00227 ± 0.00141	$0.00373 \pm 0.00157^*$
Domain V	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
Domain VI	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000

N, number of sequences; M, number of mutations; S, number of segregating sites; H, number of haplotypes.

Z-tests of the hypothesis that π equals the corresponding value for each population: ### $p < 0.0001$.

Z-tests of the hypothesis that mean d_s equals that of mean d_N : * $p < 0.05$; ** $p < 0.01$.

doi:10.1371/journal.pone.0110463.t002

Table 3. Genetic differentiation (F_{ST} indices) of *P. vivax* populations based on the PvTRAP locus.

	Tak (2006–7)	Prachuap Khirikhan	Chanthaburi	Yala and Narathiwat	Tak (1996)	Brazil
Prachuap Khirikhan	0.0001					
Chanthaburi	0.2324***	0.2319***				
Yala and Narathiwat	0.5001***	0.4657***	0.3615***			
Tak (1996)	0.0098	0.0003	0.2147***	0.3867***		
Brazil	0.5072***	0.4708***	0.4380***	0.4603***	0.4325***	
Korea	0.7057***	0.6907***	0.5514***	0.7305***	0.5742***	0.5630***

*** $p < 0.00001$.

doi:10.1371/journal.pone.0110463.t003

Tests for departure from neutral evolution

Analysis of patterns of nucleotide substitutions in the *PvTRAP* sequences of isolates in this study reveals that the rate of nonsynonymous substitutions per nonsynonymous sites or d_N significantly exceeds that of synonymous substitutions per synonymous sites or d_S in all parasite populations ($p < 0.05$ or $p < 0.01$)

except *P. vivax* isolates from Yala and Narathiwat provinces (Table 2). In contrast, d_S was greater than d_N for isolates from Yala and Narathiwat provinces although the difference was not statistically meaningful ($p = 0.467$). Closer looks into the rate of nucleotide substitutions in each domain of *PvTRAP* has revealed that d_N significantly outnumbered d_S in domains II and IV of

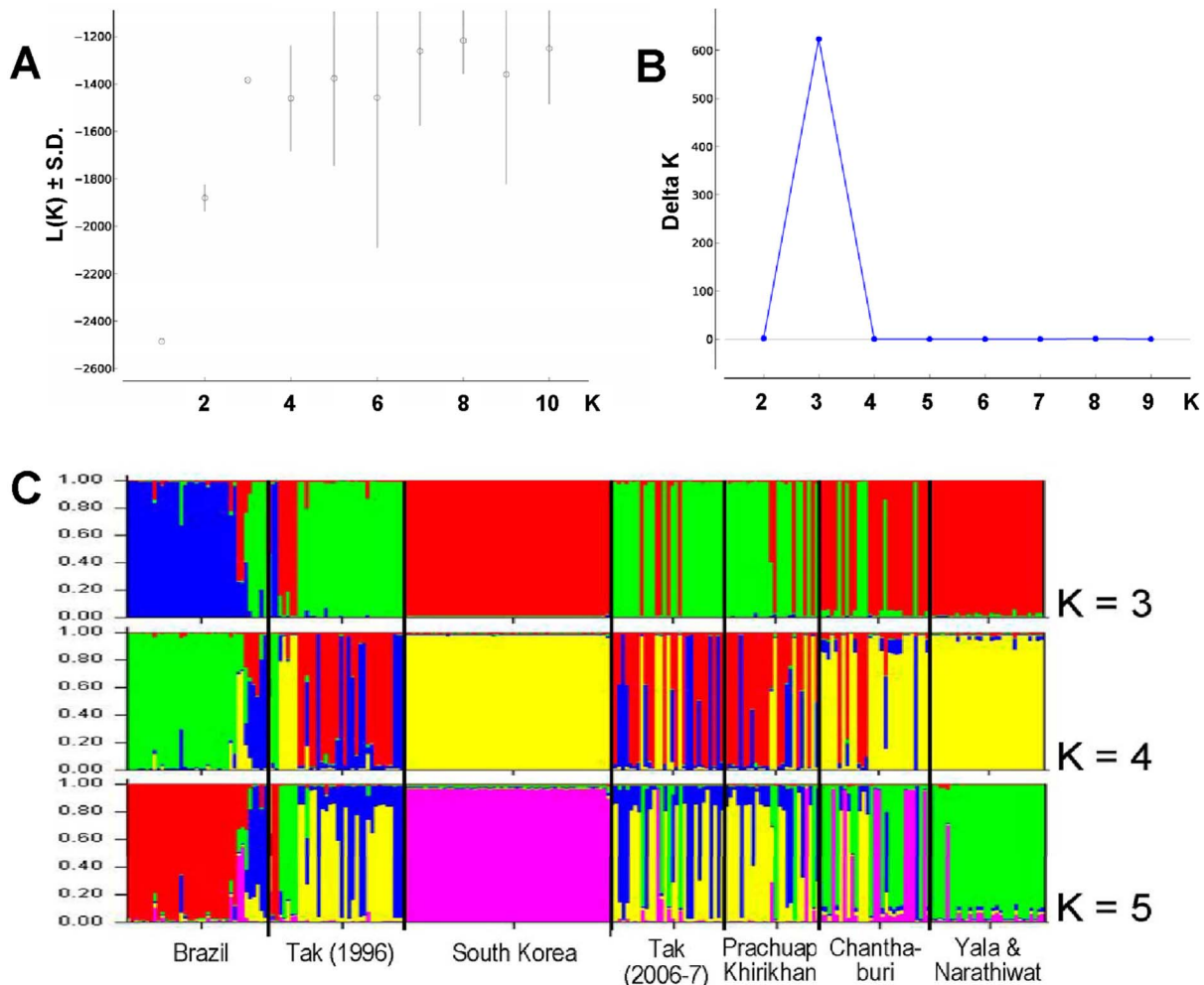


Figure 1. Determination of the most likely number of clusters of *PvTRAP* haplotypes by *ad hoc* methods: (A) relationship between K and mean of estimated mean log likelihood of K and its standard deviation or $L(K) \pm S.D.$ [45], (B) relationship between K and delta K [46]. (C) Graphical representation of the genetic structure of populations from Brazil, South Korea and each endemic area of Thailand. Results are given for K3, K4 and K5.

doi:10.1371/journal.pone.0110463.g001

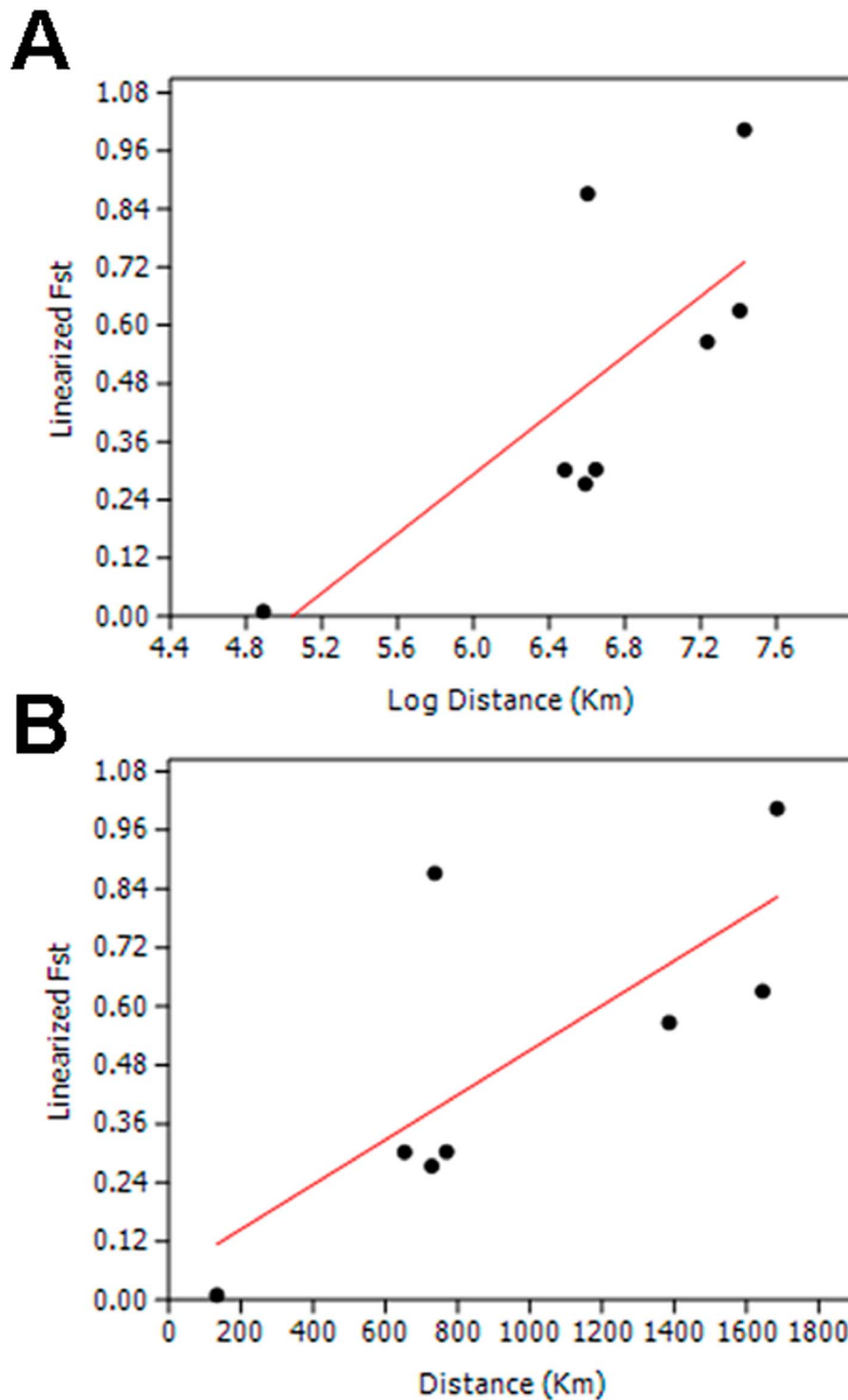


Figure 2. Regression analysis of pairwise linearized F_{st} regressed on natural logarithm geographic distances between *P. vivax* populations in Thailand (A) and pairwise linearized F_{st} regressed on geographic distances between populations (B).
doi:10.1371/journal.pone.0110463.g002

isolates from Tak, Prachuap Khirikhan and Chanthaburi provinces. Although no significant difference in d_S and d_N for the entire coding region of the *PvTRAP* gene was observed when all parasite populations were considered, d_N in domain IV was obviously greater than d_S (Table 2). Meanwhile, statistics based on Tajima's D and Fu & Li's D^* and F^* did not show evidences of departure

from neutral expectation (Table 4). However, sliding window analysis reveals significant negative Tajima's D in domain II of *P. vivax* populations from Prachuap Khirikhan and Chanthaburi provinces, suggesting purifying selection at certain residues in this gene. In contrast, evidence of balancing selection was detected in domain IV of Yala and Narathiwat isolates as shown by significant

Table 4. Intraspecific and interspecific neutrality tests for the PvTRAP locus from each endemic area of Thailand.

Province/Endemic area	Tajima's <i>D</i>	Fu & Li's <i>D</i> *	Fu & Li's <i>F</i> *	Out group: <i>P. cynomolgi</i>		Out group: <i>P. knowlesi</i>	
				Fu & Li's <i>D</i>	Fu & Li's <i>F</i>	Fu & Li's <i>D</i>	Fu & Li's <i>F</i>
Tak (n = 30)	0.097	-0.143	-0.080	0.431	0.350	0.987	1.067
Prachuap Khirikhan (n = 25)	-0.967	-0.686	-0.905	-0.086	-0.401	0.284	0.025
Chanthaburi (n = 29)	-0.818	-1.429	-1.451	-0.490	-0.798	-0.080	-0.383
Yala and Narathiwat (n = 30)	0.278	-0.120	0.000	0.010	0.227	0.427	0.591
All (n = 114)	-0.698	-1.714	-1.572	-1.055	-1.120	-0.534	-0.553

All values did not have significance departure from zero.
doi:10.1371/journal.pone.0110463.t004

positive *D* values. Meanwhile, no departure from neutrality was detected in malaria population from Tak province (Figure 3).

Results from interspecific comparison between *PvTRAP* and *PcyTRAP* using the MacDonald-Kreitman test has shown that each parasite population except that from Yala and Narathiwat provinces had significant deviation in positive direction from neutral expectation ($p < 0.01$), reflecting a signature of positive or balancing selection. Consistent results were obtained when the TRAP gene of *P. knowlesi* was used as the outgroup sequence (Table 5). In contrast, application of the Fu & Li's *D* and *F* tests using either *PcyTRAP* or *PkTRAP* as outgroup sequences did not yield significant departure from neutrality. Meanwhile, positively selected sites were identified at 4 codons in domain II and 3 codons in domain IV by one or more of the methods implemented in the HyPhy package (Table 6). Likewise, negatively selected sites were detected in residue 21 in domain I, residues 30, 100, 133 in domain II and residues 519 and 539 in domain VI, suggesting functional constraint occurring at certain residues in this protein.

Amino acid substitutions and predicted HLA binding

The amino acid substitutions in PvTRAP of Thai isolates had slightly higher percentage of conservative changes with respect to polarity and charge property as 67.7% and 58.1% of these substitutions, respectively, were unchanged. Similar percentages of these changes were observed among worldwide isolates, being 64.6% and 56.9%, respectively (Table S2). Closer look into amino acid substitutions and potential HLA-binding peptides as predicted by the high scores for the C-terminal cleavage and the transporter associated with antigen processing efficiency [48] have shown that a number of substituted residues have remarkably reduced predicted scores; thereby, the property of these epitopes could be altered. It is noteworthy that substituted residues in several of these potential epitopes occurred in domains II and IV of PvTRAP (Table 6).

Intragenic recombination

Evidence of recombination in the *PvTRAP* locus analyzed by using the HyPhy package has identified evidence of 1 recombination breakpoint with significant topological incongruence between AICc score of the best fitting GARD model ($p < 0.05$) for *P. vivax* populations in Thailand except those from Yala and Narathiwat provinces ($0.05 < p < 0.1$). Although these recombination breakpoints occur at different positions, i.e. between nucleotides 1008 and 1009, 1170 and 1171, and 1173 and 1174 for Chanthaburi, Prachuap Khirikhan and Tak provinces, respectively, it is noteworthy that all were located in domain IV of the *PvTRAP* gene. Meanwhile, estimation of the parameter *Rm* reveals 3 and 5 recombination sites in parasite populations from Tak and Prachuap Khirikhan provinces whereas Chanthaburi isolates and southern Thai *P. vivax* populations had equal number of potential recombination sites (*Rm* = 2). Locations of potential recombination sites for each population are listed in Table S3, most of which span domain IV. These results suggest that intragenic recombination could shape diversity at the *PvTRAP* locus.

Phylogenetic relationship

Model test reveals that the Tamura 3-parameter with the rate variation model that allowed for some sites to be evolutionarily invariable gave the lowest BIC value. The tree with the highest log likelihood (-3285.15) reveals that the majority of Brazilian isolates were placed in a separate lineage while the majority of Tak isolates collected in 1996, in 2006–2007 and most Prachuap Khirikhan isolates were clustered together. Several isolates from Chantha-

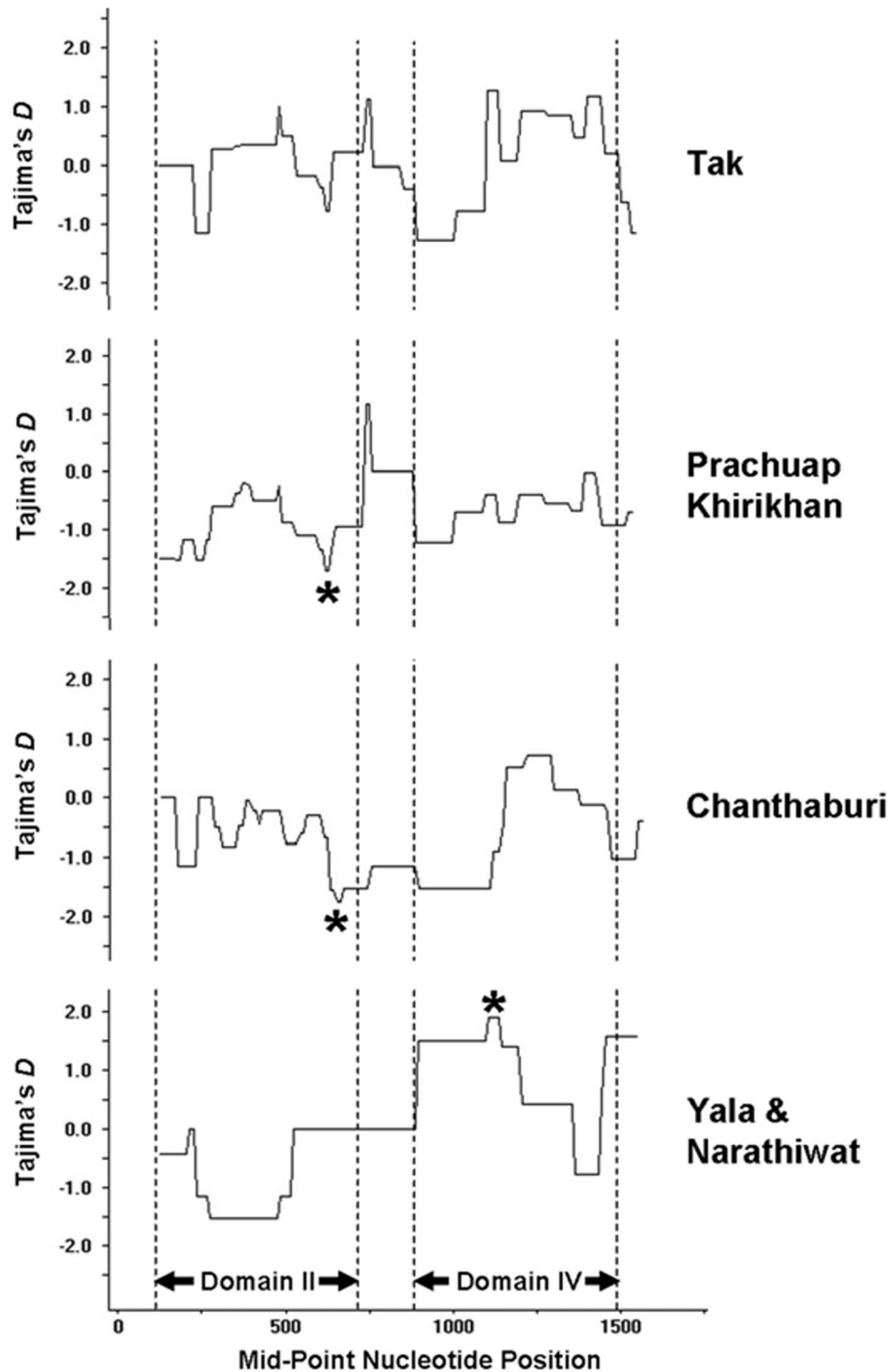


Figure 3. Sliding window plots of Tajima's D across *PvTRAP* sequences from each population in Thailand. Asterisks denote values with $p < 0.05$. Plots are based on a window size of 250 bp and a step size of 10 bp.
doi:10.1371/journal.pone.0110463.g003

Chanthaburi, Yala and Narathiwat provinces seem to be located in separate branches from the remaining Thai isolates. However, there was no clear lineage for each parasite population because several isolates from different endemic areas shared or was placed in closely related branches. More importantly, all branches in this phylogenetic inference received bootstrap values less than 80%,

suggesting no distinct lineage of the *PvTRAP* haplotypes from worldwide origins (Figure 4).

Discussion

Host adhesion and motility are fundamental properties of malarial sporozoites to establish infection that involves sporozoite-specific

Table 5. McDonald–Kreitman tests on TRAP of *Plasmodium vivax* from diverse geographic origins with *P. cynomolgi* orthologue as outgroup species.

Population	Polymorphic changes within <i>P. vivax</i>		Fixed differences between species		Neutrality index	ρ value
	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous		
Tak	1	13	116	140	10.77	0.0045
Prachuap Khirikhan	3	21	117	138	5.93	0.0011
Chanthaburi	2	16	117	140	6.69	0.0053
Yala and Narathiwat	3	4	115	145	1.06	1.0000
All	6	23	117	136	3.298	0.0096

Note: Repeats are excluded from all analyses.
doi:10.1371/journal.pone.0110463.t005

proteins, one of which is TRAP. It has been shown that TRAP-deficient sporozoites were severely impaired in their host cell adherence property and gliding motility; thereby, invasion of mosquito's salivary glands and hepatocytes was interrupted [14,49]. Two extracellular portions of malarial TRAP, the von Willebrand A-domain and the thrombospondin repeats located in domains II and IV, respectively, are crucial for initial host cell adherence and stabilization of adhesion/deadhesion during gliding mobility of sporozoites [49,50]. Crystal structure analyses spanning the von Willebrand A-domain and the thrombospondin repeats of PvTRAP and PfTRAP reveal two conformational states, open and closed structures. Such structural phase transition is possibly responsible for 'stick-and-slip' or gliding motility of sporozoites through the actomyosin motility apparatus [51,52]. The sequence motif of metal-ion-dependent adhesion sites (MIDAS) in von Willebrand A-domain and the extensible β ribbon that implicates in ligand binding during the open high-affinity state [51] were perfectly conserved in all PvTRAP haplotypes as shown in this study. Although amino acid substitutions were observed in the β 1 and α 3 domains, charge and polarity of these substituted residues were retained, suggesting structural constraint in these portions of the protein (Table 1). Experimental studies have shown that artificially engineered mutations of the MIDAS-coordinating amino acid residues that were conserved across malaria species could impair gliding motility and sporozoite infectivity to both mosquito and mammalian hosts [49,50]. On the other hand, alteration in both charge and polarity profiles of some substituted amino acids (residues 143, 166, 172 and 176, Tables 1 and S2) were identified in α 4 and α 5 domains, implying that changes in these residues may not drastically affect functional conformation of PvTRAP. Nevertheless, further experimental studies are required to address this issue.

The extent of sequence diversity of the *PvTRAP* locus among parasite populations in Thailand seems to be comparable among endemic areas except that from southern region (Table 2). The results from this study are in line with our recent analyses of other major vaccine antigens, i.e. apical membrane antigen-1 (Pvama-1), merozoite surface protein-1 (Pvmsp-1), merozoite surface protein-4 (Pvmsp-4) and merozoite surface protein-5 (Pvmsp-5) of *P. vivax* showing that southern parasite population possesses significantly lower levels of nucleotide diversity, number of haplotypes and haplotype diversity than those of northwestern parasite population [24,25,53]. Heterogeneity in the *PvTRAP* sequences could have been arisen from selective pressures and intragenic recombination. Recombination confers evolutionary advantages because creation of adaptive traits or removal of deleterious mutants is enabled. Importantly, most recombination sites were identified within or spanning domain IV of the *PvTRAP* locus where signature of

positive selection was identified. The relatively higher levels of nucleotide diversity of *PvTRAP* from Thai-Myanmar border (Tak and Prachuap Khirikhan isolates) than that from Thai-Cambodia border (Chanthaburi isolates) seems to be in line with R_m in each population. It is likely that transmigration of people could maintain and spread of *P. vivax* harboring variant *PvTRAP* haplotypes while recombination could further generate novel alleles. Intriguingly, it seems that reduction in genetic diversity of southern parasite population may not be simply explained by clonal population structure of *P. vivax* in this region because recombination event could be traced despite the paucity of potential recombination sites observed ($R_m = 2$). In contrast, several decades of extensive malaria control in this region and a lack of remarkable re-introduction of malaria cases from Malaysia could have resulted in population bottleneck as previously suggested [24]. Therefore, after bottleneck effect and relaxation of malaria control in southern Thailand because of difficulty in implementation due to local political unrest, it could be possible that recombination between different *PvTRAP* haplotypes may gradually increase the level of genetic diversity in this population.

Our analyses have shown that domains II and IV of the *PvTRAP* locus had significant difference in the rate of nonsynonymous substitutions than that of synonymous substitutions. Evidence of positive or balancing selection was also reaffirmed by the McDonald-Kreitman test. Although departure from neutrality in the entire coding PvTRAP sequence was not apparent by using Tajima's D and related statistics, sliding window analysis has identified significant negative Tajima's D values in domains II of *P. vivax* population from Prachuap Khirikhan and Chanthaburi provinces while significant positive values occurred in domain IV among isolates from Yala and Narathiwat provinces. Differential patterns of departure from neutrality among populations could reflect different evolutionary forces exerted on PvTRAP of each population. Alternatively, intrinsic statistical property of Tajima's D statistics can be confounded by demographic processes [54]. Therefore, significant positive Tajima's D observed in southern Thai isolates could imply balancing selection or a recent bottleneck because an excess of intermediate frequency alleles as rare alleles are lost from the population due to genetic drift can be found immediately after the bottleneck. Meanwhile, significant negative Tajima's D value in domains II could imply purifying selection and selective sweep while migration and population growth may be the confounders [55]. Nevertheless, positively selected codons were found based on analyses using SLAC, FEL, REL and MEME methods while negatively selected residues were detected in domains I, II and VI (Table 6). Taken together, it seems that co-existence of both

Table 6. Codon-based analysis of departure from neutrality in the PvTRAP locus of Thai isolates.

Domain	Codon	SLAC		FEL		REL		MEME		Consensus*		
		$d_N - d_S$	p value	$d_N - d_S$	p value	$d_N - d_S$	p value	Bayes Factor	$\omega+$	p value	SLAC	FEL
II	120	93.598	0.213	2821.14	0.091	73.567	9897.75	> 100	0.115		+	+
II	134	74.642	0.496	2675.58	0.208	73.381	195736	> 100	0.230			+
II	206	84.790	0.280	3760.95	0.105	73.323	702283	> 100	0.051			+
II	255	149.777	0.112	3884.15	0.063	73.501	2.27475×10^{10}	> 100	0.084		+	+
IV	410	95.051	0.401	3087.63	0.178	73.377	8780900	> 100	0.203			+
IV	421	55.136	0.594	1821.01	0.252	73.389	2484.53	> 100	0.267			+
IV	439	83.134	0.304	2728.63	0.104	73.541	267258	> 100	0.129			+
I	21	-102.062	0.151	-2637.21	0.063	-38.418	35812.5	< 100	> 0.100		-	-
II	30	-204.123	0.023	-5228.32	0.011	-38.410	29828.7	< 100	> 0.100		-	-
II	100	-46.299	0.333	-1399.13	0.131	-38.407	29051.4	< 100	> 0.100		-	-
II	133	-31.970	0.483	-1019.35	0.210	-38.100	8558.41	< 100	> 0.100		-	-
VI	519	-92.599	0.111	-2873.15	0.032	-38.407	29054.6	< 100	> 0.100		-	-
VI	539	-83.719	0.199	-3571.07	0.051	-38.391	27525.8	< 100	> 0.100		-	-

*positively selected site, +; negatively selected site, -.

Default significance levels in Datamonkey program: SLAC, $p < 0.1$; FEL, $p < 0.1$; REL, Bayes Factor > 1000 ; MEME, $p < 0.1$. doi:10.1371/journal.pone.0110463.t006

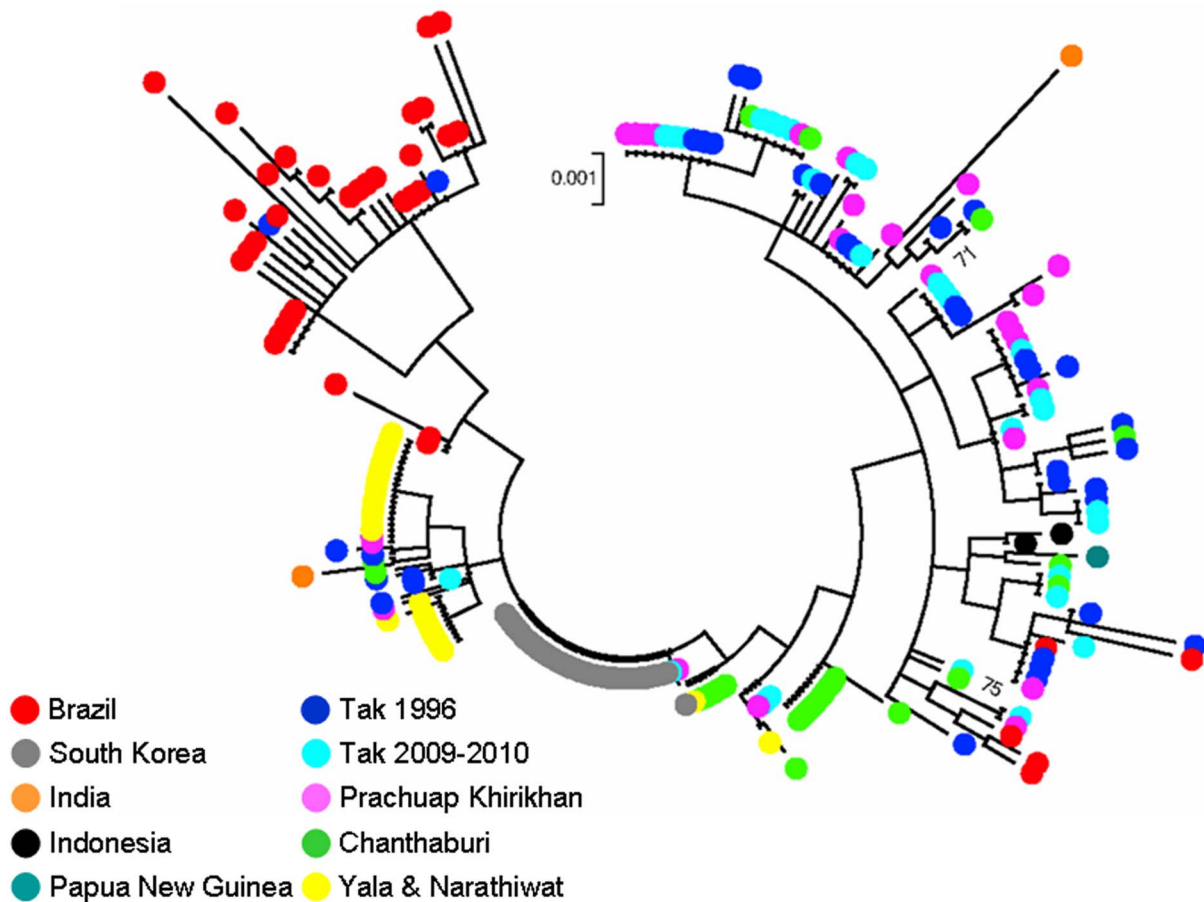


Figure 4. Phylogenetic analysis inferred from a total of 246 PvTRAP sequences from this study and those previously reported [19,28] using the Maximum Likelihood method based on the Tamura 3-parameter model. Gaps and missing data are excluded from analysis. Bootstrap values more than 70% are shown. Dots with different colors represent isolates and their geographic origins as listed on the lower left. doi:10.1371/journal.pone.0110463.g004

negatively selected amino acids and residues under positive selection could stem from functional constraint of this molecule and probably driven by host immune pressure, respectively. Meanwhile, it seems that intragenic recombination was pronounced in populations from Tak and Prachuap Khirikhan provinces while relatively few recombinations occurred in parasites from Chanthaburi province and southern isolates. These findings are in accord with our previous analysis on *PvMSP4* and *PvMSP5* in which frequency of recombination seems to correlate with endemicity of each area [25,53].

Despite low level of sequence diversity, genetic distances inferred from the *PvTRAP* locus among *P. vivax* populations could be deployed for analysis of population structure. Pairwise comparison of populations revealed that South Korean isolates displayed high and significant *Fst* indices when compared with malaria populations in other endemic areas. Likewise, slightly lower but significant *Fst* indices were observed when isolates from Brazil were compared with populations elsewhere (Table 3). Therefore, gene flow between *P. vivax* populations in Thailand, Brazil and Korea may not occur because geographic distance has precluded intra- and intercontinental spread of parasites. Although malaria can be transferred between remote endemic areas through infected travelers, this situation could be inefficient due to rarity in numbers of such cases to contribute to genetic admixture of parasites in these three countries. However, closer look into genetic differentiation between 4 major malaria endemic areas of

Thailand, significant genetic differentiation of *P. vivax* population between endemic areas was observed except parasites from Tak province collected over a decade apart (1996 and 2006–2007) having low and non-significant *Fst* indices. Therefore, diversity in the *PvTRAP* locus of parasites in Thailand seems to exhibit spatial but not temporal variation. It is noteworthy that regression analysis of linear geographic as well as normal logarithm of geographic distances and pairwise linearized *Fst* values yielded significant correlation (Mantel test, $p = 0.024$ and 0.028 , respectively), indicating that *P. vivax* populations in these 4 endemic areas have been isolated by distance (Figure 2). Therefore, it seems likely that the flight range of mosquito vectors was the major determinant for magnitude of genetic differentiation of *P. vivax* populations between endemic areas. However, there remain some other factors that could influence and shape genetic structure of malaria in Thailand. Although geographic distance between Tak and Prachuap Khirikhan provinces is about 160 kilometers longer than that between Tak and Chanthaburi provinces, significant genetic differentiation occurred in the latter but not the former. It should be noted that transmigration of people along Thai-Myanmar border where both Tak and Prachuap Khirikhan provinces are located is intense during the past decades whereas transmigration of gem miners who, not uncommonly, carried malaria parasites in their circulation between Tak and Chanthaburi provinces had ceased for over 3 decades ago [56]. Furthermore, the abundance of malaria vectors along Thai-Myanmar

Table 7. Amino acid substitutions in potential T cell epitopes of PvTRAP and predicted scores*.

HLA	Domain	Predicted epitopes	Predicted score	%Reduction	No. isolates
A1	I	YLLV <u>F</u> LLY	2.1645		112
		YLLV <u>V</u> FLLY	1.6240	24.97	2
A1	II	VCNESVD <u>L</u> Y	0.9376		113
		VCNESVD <u>L</u> I	0.7571	19.25	1
A2	I	F <u>L</u> LYVSIFA	1.1395		112
		F <u>F</u> LYVSIFA	0.5416	52.47	2
A2	II	NMTAAL <u>E</u> EV	1.1851		47
		NMTAAL <u>D</u> EV	1.1104	6.30	65
		NMTAAL <u>D</u> DV	0.7819	34.02	2
A3	II	Y <u>T</u> ALEVAKK	0.8093		63
		Y <u>R</u> ALEVAKK	0.2512	68.96	51
A3	II	KVTEL <u>R</u> KSY	0.7907		102
		KVTEL <u>R</u> KTY	0.5974	24.45	12
A24	I	V <u>F</u> LLYSIF	1.7214		112
		V <u>F</u> LYVSIF	1.4858	13.69	2
A26	II	ELR <u>K</u> TYSPI	2.0070		12
		ELR <u>K</u> SYSPI	1.9132	4.67	102
		ESVD <u>L</u> YLLV	1.3830	31.09	1
		ESVD <u>I</u> YLLV	1.3656	31.96	113
B7	II	KLK <u>Q</u> RNV <u>S</u> L	0.8786		112
		KLK <u>K</u> RNV <u>S</u> L	0.8400	4.39	1
		KLK <u>Q</u> RNV <u>T</u> L	0.7263	17.33	1
B7	II	K <u>Q</u> RNV <u>T</u> LAV	0.9688		1
		K <u>K</u> RNV <u>S</u> LAV	0.8745	9.73	1
		K <u>Q</u> RNV <u>S</u> LAV	0.7908	18.37	112
B7	IV	LPVPAP <u>L</u> PA	1.3094		2
		LPVPAP <u>L</u> PT	1.1464	12.45	112
B8	II	RPR <u>E</u> RNCKF	1.9566		1
		RPR <u>L</u> RNCKF	1.3892	29.00	2
		RPR <u>S</u> RNCKF	1.0950	44.04	2
		RPR <u>E</u> PNCKF	1.0440	46.64	109
B8	II	ELR <u>K</u> TYSPI	0.9655		12
		ELR <u>K</u> SYSPI	0.8494	12.02	102
B27	II	<u>K</u> RNV <u>S</u> LAVI	1.2095		1
		<u>Q</u> RNV <u>S</u> LAVI	0.9591	20.70	112
		<u>Q</u> RNV <u>T</u> LAVI	0.9591	20.70	1
B27	II	K <u>Q</u> RNV <u>S</u> LAV	0.7781		112
		K <u>Q</u> RNV <u>T</u> LAV	0.7781	0	1
		K <u>K</u> RNV <u>S</u> LAV	0.5384	30.81	1
B27	III	G <u>R</u> GTHSRSR	1.2775		96
		G <u>K</u> GTHSRSR	0.4904	61.61	18
B39	I	SYLLV <u>V</u> F <u>L</u> L	0.7915		112
		SYLLV <u>V</u> F <u>F</u> L	0.6755	14.66	2
B44	IV	NEKVIP <u>N</u> PL	1.4650		110
		NEKVIP <u>T</u> PL	1.4469	1.24	3
		NEKVIP <u>K</u> PL	1.2157	17.02	1
B58	I	KSYLLV <u>V</u> F <u>F</u>	1.4989		2
		KSYLLV <u>V</u> F <u>L</u>	0.8347	44.31	112
B62	II	K <u>Q</u> RNV <u>T</u> LAV	1.1504		1
		K <u>Q</u> RNV <u>S</u> LAV	1.1443	0.53	112

Table 7. Cont.

HLA	Domain	Predicted epitopes	Predicted score	%Reduction	No. isolates
		KKRNVS L AV	0.2828	75.42	1

*Based on the C-terminal cleavage and the transporter associated with antigen processing efficiency [48].
doi:10.1371/journal.pone.0110463.t007

border could facilitate gene flow of *P. vivax* in these areas. In contrast, ecological niches of *Anopheles* vectors between Tak and Chanthaburi provinces are largely interrupted by focal deforestation and urbanization, precluding natural spread of malaria between these areas.

Analysis of hierarchical splitting of gene pools by STRUCTURE has shown that $K = 3$ is an optimum value. Intriguingly, $K = 3$ could not subdivide South Korean and southern Thai isolates that in fact differed biologically and genetically. A very long incubation period of vivax malaria from 230 to 300 days has been recognized among South Korean isolates [28] but, to our knowledge, has never been recorded in Thai strains. Furthermore, phylogenetic analysis has placed all South Korean isolates to a separate cluster from southern Thai isolates, *albeit* without significant bootstrap support (Figure 3). In contrast, the partition at $K = 5$ is more consistent with results from analysis using *Fst* indices in which distinct subpopulation structure was perceived between isolates from southern Thailand and those from South Korea whilst at $K < 5$, no discernible difference was observed between these populations. Meanwhile, phylogenetic inference does not support distinct clusters of isolates relating to their geographic origins because no branch received remarkably high bootstrap values, consistent with our previous study [19]. Therefore, genetic diversity of the *PvTRAP* locus displays tendency towards geographic variation (Figure 4).

Protective immunity against malarial TRAP involves both humoral and cell mediated immune responses. Antibodies against TRAP targeting domains involved in gliding motility and hepatocyte binding ligands could prevent sporozoite invasion into hepatocytes [13] while TRAP-specific cytotoxic T cells could destroy sporozoite-infected hepatocytes [57]; thereby, exoerythrocytic development does not ensue. The roles of anti-TRAP antibodies in partial protection from or significantly reduced risk of falciparum malaria have been demonstrated in African endemic areas [12,58,59]. With analogy to the thrombospondin-related motif (TRM) of PfTRAP that binds heparin sulfate proteoglycan of hepatocyte, the TRM in domain III of PvTRAP is relatively conserved with two amino acid substitutions, i.e. WTACSVTC-GR(or K)GTH (or Q) SRSR, while the MIDAS motif is perfectly conserved among clinical isolates. Therefore, limited diversity at the functional domains in PvTRAP has encouraged vaccine incorporation. Meanwhile, several CD4+ and CD8+ T cell epitopes have been mapped across PfTRAP, some of which have been associated with protection [57,60,61]. Importantly, T cell responses to PfTRAP were commonly allele-specific although certain epitopes

could induce or serve as targets of cross-reactive immunity [60]. Recent studies have highlighted the importance of protective roles of CD8+ T cell and memory T cell responses to PfTRAP from clinical malaria [61,62]. The significance of diversity in these T cell epitopes among natural malaria population remains to be elucidated in term of vaccine efficacy. To date, little is known about immunological responses to PvTRAP although vaccination study using synthetic peptide encompassing hepatocyte-binding ligand of the protein has conferred protection against parasite challenge in *Aotus* monkeys [63]. Our analysis suggests that polymorphisms in PvTRAP could affect T cell recognition because amino acid substitutions in several predicted HLA-binding peptides have remarkably altered the predicted epitope scores (Table 7).

In conclusion, our analysis has shown that genetic diversity in the *PvTRAP* locus of Thai *P. vivax* isolates exhibits geographic variation and population structure. Positive selection and intragenic recombination have shaped diversity at the *PvTRAP* locus. The low level of sequence diversity in *PvTRAP* among clinical isolates as shown in this study may not drastically compromise vaccine design.

Supporting Information

Table S1 Haplotypes of PvTRAP from Thai isolates.
(DOC)

Table S2 Characteristics of amino acid substitutions in PvTRAP of worldwide and Thai isolates.
(DOC)

Table S3 Minimum number of recombination events (Rm) and recombination sites in the PvTRAP gene of each parasite population in Thailand.
(DOC)

Acknowledgments

We are grateful to all patients who donated their blood samples for this study. We thank the staff of the Bureau of Vector Borne Disease, Department of Disease Control, Ministry of Public Health, Thailand, for assistance in field work.

Author Contributions

Conceived and designed the experiments: SJ CP RK. Performed the experiments: RK CP SJ. Analyzed the data: SJ CP RK. Contributed reagents/materials/analysis tools: SJ CP HT. Contributed to the writing of the manuscript: SJ CP. Reviewed the paper: SJ CP RK HT.

References

- Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ (2006) Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 367: 1747–1757.
- Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, et al. (2007) Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 77(Suppl.): 79–87.
- Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW (2004) The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis* 4: 327–336.
- Crompton PD, Pierce SK, Miller LH (2010) Advances and challenges in malaria vaccine development. *J Clin Invest* 120: 4168–4178.
- Mayxay M, Pukrittayakamee S, Newton PN, White NJ (2004) Mixed-species malaria infections in humans. *Trends Parasitol* 20: 233–240.
- Casares S, Brumeanu TD, Richie TL (2010) The RTS,S malaria vaccine. *Vaccine* 28: 4880–4894.
- Hoffman SL, Goh LM, Luke TC, Schneider I, Le TP, et al. (2002) Protection of humans against malaria by immunization with radiation-attenuated *Plasmodium falciparum* sporozoites. *J Infect Dis* 185: 1155–1164.

8. Roestenberg M, McCall M, Hopman J, Wiersma J, Luty AJ, van Gemert GJ, et al. (2009) Protection against a malaria challenge by sporozoite inoculation. *N Engl J Med* 361: 468–477.
9. Rogers WO, Malik A, Mellouk S, Nakamura K, Rogers MD, et al. (1992) Characterization of *Plasmodium falciparum* sporozoite surface protein 2. *Proc Natl Acad Sci USA* 89: 9176–9180.
10. Khumsmith S, Charoenvit Y, Kumar S, Sedegah M, Beaudoin RL, et al. (1991) Protection against malaria by vaccination with sporozoite surface protein 2 plus CS protein. *Science* 252: 745–718.
11. Hafalla JCR, Bauza K, Friesen J, Gonzalez-Aseguinolaza G, Hill AVS, et al. (2013) Identification of targets of CD8+ T cell responses to malaria liver stages by genome-wide epitope profiling. *PLoS Pathog* 9: e1003303.
12. Scarselli E, Tolle R, Koita O, Diallo M, Müller HM, et al. (1993) Analysis of the human antibody response to thrombospondin-related anonymous protein of *Plasmodium falciparum*. *Infect Immun* 61: 3490–3495.
13. Müller HM, Reckman I, Hollingdale MR, Bujard H, Robson KJH, et al. (1993) Thrombospondin related anonymous protein (TRAP) of *Plasmodium falciparum* binds specifically to sulfated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. *EMBO J* 12: 2881–2889.
14. Sultan AA, Thathy V, Frevert U, Robson KJH, Crisanti A, et al. (1997) TRAP is necessary for gliding motility and infectivity of *Plasmodium* sporozoites. *Cell* 90: 511–522.
15. Robson KJ, Hall JR, Jennings MW, Harris TJ, Marsh K, et al. (1988) A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* 335: 79–82.
16. Templeton TJ, Kaslow DJ. (1997) Cloning and cross-species comparison of the thrombospondin-related anonymous protein (TRAP) gene from *Plasmodium knowlesi*, *Plasmodium vivax* and *Plasmodium gallinaceum*. *Mol Biochem Parasitol* 84: 13–24.
17. Kappe SH, Buscaglia CA, Bergman LW, Coppens I, Nussenzweig V (2004) Apicomplexan gliding motility and host cell invasion: overhauling the motor model. *Trends Parasitol* 20: 13–16.
18. Jongwutiwes S, Putaporntip C, Kanbara H, Tanabe K (1998) Variation in the thrombospondin-related adhesive protein (TRAP) gene of *Plasmodium falciparum* from Thai field isolates. *Mol Biochem Parasitol* 92: 349–353.
19. Putaporntip C, Jongwutiwes S, Tia T, Ferreira MU, Kanbara H, et al. (2001) Diversity in the thrombospondin-related adhesive protein gene (TRAP) of *Plasmodium vivax*. *Gene* 268: 97–104.
20. Ohashi J, Suzuki Y, Naka I, Hananantachai H, Patarapotikul J (2014) Diversifying selection on the thrombospondin-related adhesive protein (TRAP) gene of *Plasmodium falciparum* in Thailand. *PLoS One* 9: e90522.
21. Gilbert SC, Plebanski M, Gupta S, Morris J, Cox M, et al. (1998) Association of malaria parasite population structure, HLA, and immunological antagonism. *Science* 279: 1173–1177.
22. Putaporntip C, Hongsrimuang T, Seethamchai S, Kobasa T, Limkittikul K, et al. (2009) Differential prevalence of *Plasmodium* infections and cryptic *Plasmodium knowlesi* malaria in humans in Thailand. *J Infect Dis* 199: 1143–1150.
23. Jongwutiwes S, Buppan P, Kosuvin R, Seethamchai S, Pattanawong U, et al. (2011) *Plasmodium knowlesi* malaria in humans and macaques, Thailand. *Emerg Infect Dis* 17: 1799–1806.
24. Jongwutiwes S, Putaporntip C, Hughes AL (2010) Bottleneck effects on vaccine-candidate antigen diversity of malaria parasites in Thailand. *Vaccine* 28: 3112–3117.
25. Putaporntip C, Udomsangpetch R, Pattanawong U, Cui L, Jongwutiwes S. (2010) Genetic diversity of the *Plasmodium vivax* merozoite surface protein-5 locus from diverse geographic origins. *Gene* 456: 24–35.
26. Putaporntip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho W-G, et al. (2002) Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. *Proc Natl Acad Sci USA* 99: 16348–16353.
27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
28. Nam MH, Jang JW, Kim H, Han ET, Lee WJ, et al. (2011) Conserved sequences of thrombospondin-related adhesive protein gene of *Plasmodium vivax* in clinical isolates from Korea. *Trop Med Int Health* 16: 923–928.
29. Nei M (1987) Molecular evolutionary genetics. New York: Columbia University Press.
30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729.
31. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164.
32. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
33. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies *Bioinformatics* 21: 676–679.
34. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23: 1891–1901.
35. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.
36. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In Munro HN editor. *Mammalian Protein Metabolism* 21–132 Academic Press New York.
37. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.
38. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
39. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351: 652–654.
40. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8: e1002764.
41. Kosakovsky Pond SL, Frost SDW (2005) Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21: 2531–2533.
42. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
43. Weir BS, Cockerham CC (1984) Estimating *F* statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
44. Slatkin M (1993) Isolation by distance in equilibrium and non equilibrium populations. *Evolution* 47: 264–279.
45. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
46. Evanno G, Regnaut S, Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
47. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour* 4: 359–361.
48. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, et al. (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 35: 2295–2303.
49. Wengelnik K, Spaccapelo R, Naitza S, Robson KJH, Janse CJ, et al. (1999) The A-domain and the thrombospondin-related motif of *Plasmodium falciparum* TRAP are implicated in the invasion process of mosquito salivary glands. *EMBO J* 18: 5195–5204.
50. Matuschewski K, Nunes AC, Nussenzweig V, Menard R (2002) *Plasmodium* sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *EMBO J* 21: 1597–1606.
51. Song G, Koksall AC, Lu C, Springer TA (2012) Shape change in the receptor for gliding motility in *Plasmodium* sporozoites. *Proc Natl Acad Sci USA* 109: 21420–21425.
52. Pihlajamaa T, Kajander T, Knuuti J, Horkka K, Sharma A, et al. (2013) Structure of *Plasmodium falciparum* TRAP (thrombospondin-related anonymous protein) A domain highlights distinct features in apicomplexan von Willebrand factor A homologues. *Biochem J* 450: 469–476.
53. Putaporntip C, Jongwutiwes S, Ferreira MU, Kanbara H, Udomsangpetch R, et al. (2009) Limited global diversity of the *Plasmodium vivax* merozoite surface protein 4 gene. *Infect Genet Evol* 9: 821–826.
54. Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol* 13: 494–504.
55. Maruyama T, Fuerst PA (1984) Population bottlenecks and non-equilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. *Genetics* 108: 745–763.
56. Thimasam K, Jatapadma S, Vijaykadga S, Sirichaisinthop J, Wongsrichanalai C (1995) Epidemiology of malaria in Thailand. *J Travel Med* 2: 59–65.
57. Schneider J, Gilbert SC, Blanchard TJ, Hanke T, Robson KJH, et al. (1998) Enhanced immunogenicity for CD8+ T cell induction and complete protective efficacy of malaria DNA vaccination by boosting with modified vaccinia virus Ankara. *Nat Med* 4: 397–402.
58. John CC, Zickafoose JS, Sumba PO, King CL, Kazura JW (2003) Antibodies to the *Plasmodium falciparum* antigens circumsporozoite protein, thrombospondin-related adhesive protein, and liver-stage antigen 1 vary by ages of subjects and by season in a highland area of Kenya. *Infect Immun* 71: 4320–4325.
59. John CC, Tande AJ, Moormann AM, Sumba PO, Lanar DE, et al. (2008) Antibodies to pre-erythrocytic *Plasmodium falciparum* antigens and risk of clinical malaria in Kenyan children. *J Infect Dis* 197: 519–526.
60. Flanagan KL, Plebanski M, Odhiambo K, Sheu E, Mwangi T, et al. (2006) Cellular reactivity to the p. *Falciparum* protein trap in adult kenyans: novel epitopes, complex cytokine patterns, and the impact of natural antigenic variation. *Am J Trop Med Hyg* 74: 367–375.
61. Ewer KJ, O'Hara GA, Duncan CJ, Collins KA, Sheehy SH, et al. (2013) Protective CD8+ T-cell immunity to human malaria induced by chimpanzee adenovirus-MVA immunisation. *Nat Commun* 4: 2836.
62. Todryk SM, Bejon P, Mwangi T, Plebanski M, Urban B, et al. (2008) Correlation of memory T cell responses against TRAP with protection from clinical malaria, and CD4+ CD25 high T Cells with susceptibility in Kenyans. *PLoS ONE* 3: e2027.
63. Castellanos A, Arévalo-Herrera M, Restrepo N, Gulloso L, Corradin G, et al. (2007) *Plasmodium vivax* thrombospondin related adhesion protein: Immunogenicity and protective efficacy in rodents and Aotus monkeys. *Mem Inst Oswaldo Cruz* 102: 411–416.