Published in final edited form as:

Infect Genet Evol. 2009 September ; 9(5): 821–826. doi:10.1016/j.meegid.2009.04.017.

Limited Global Diversity of the *Plasmodium vivax* Merozoite Surface Protein 4 Gene

Chaturong Putaporntip^a, Somchai Jongwutiwes^{a,*}, Marcelo U. Ferreira^b, Hiroji Kanbara^c, Rachanee Udomsangpetch^d, and Liwang Cui^{e,*}

^a Molecular Biology of Malaria and Opportunistic Parasites Research Unit, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok, 10330 Thailand ^b Departamento de Parasitologia, Instituto de Ciencias Biomedicas da Universidade de sao Paulo, Brazil ^c Department of Protozoology, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan ^d Department of Pathobiology, Faculty of Science, Mahidol University, Bangkok, Thailand ^e Department of Entomology, The Pennsylvania State University, 501 ASI Building, University Park, PA 16802, USA

Abstract

Merozoite surface proteins (MSPs) of the malaria parasites are major candidates for vaccine development targeting asexual blood stages. However, the diverse antigenic repertoire of these antigens that induce strain-specific protective immunity in human is a major challenge for vaccine design and often determines the efficacy of a vaccine. Here we further assessed the genetic diversity of *Plasmodium vivax* MSP4 (PvMSP4) protein using 195 parasite samples collected mostly from Thailand, Indonesia and Brazil. Overall, PvMSP4 is highly conserved with only eight amino acid substitutions. The majority of the haplotype diversity was restricted to the two short tetrapeptide repeat arrays in exon 1 and 2, respectively. Selection and neutrality tests indicated that exon 1 and the entire coding region of PvMSP4 were under purifying selection. Despite the limited nucleotide polymorphism of PvMSP4, significant genetic differentiation among the three major parasite populations was detected. Moreover, microgeographical heterogeneity was also evident in the parasite populations from different endemic areas of Thailand.

Introduction

Plasmodium vivax malaria is considered as the most neglected, highly prevalent, and potentially dangerous disease (Baird, 2007; Price et al., 2007). Currently, it is estimated that 2.6 billion people live at the risk of vivax malaria and there are 130–145 million vivax malaria cases per year (Hay et al., 2004; Guerra et al., 2006). Although rarely fatal, a *P. vivax* infection is by no means clinically benign, and the episode of debilitating intensity can be repeated several times due to relapses (Price et al., 2007). Recently, several prospective studies have reported that clinical and severe vivax malaria, manifested by severe anaemia, respiratory

*Corresponding authors: Somchai Jongwutiwes, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok, 10330 Thailand, Tel: (+66) 2-25264000 ext 3685, Fax: (+66) 2-2525944, jongwutiwes@gmail.com, Liwang Cui Department of Entomology, The Pennsylvania State University, 501 ASI Building, University Park, PA 16802, USA, Tel: 1-814-863-7663, Fax: 1-814-865-3048, luc2@psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

distress and impaired consciousness, was common in hyperendemic areas of Asia and South America (Barcus et al., 2007; Rodriguez-Morales et al., 2007; Genton et al., 2008; Tjitra et al., 2008). Furthermore, the first-line therapies comprised of chloroquine and primaquine remain unchanged for the past 50 years, and the appearance of resistant parasites to these drugs are of great concern for vivax malaria control. As malaria eradication is once again on the agenda of international malaria control communities (Roberts and Enserink, 2007), these efforts must also include elimination of vivax malaria.

Effective control of malaria requires comprehensive approaches integrating multiple strategies such as chemotherapy, vaccination, and vector control. For vaccine development, antigens located on the surface or apical organelles of merozoites are prime candidates, given their potentials of eliciting immunity that blocks merozoite invasion of host erythrocytes (Richie and Saul, 2002). The effect of vertebrate host immunity on shaping the parasite population structure is well known and considered mostly responsible for the diverse antigenic repertoire of the parasite. There is strong evidence showing the induction of strain-specific protective immunity in human (e.g., Hodder et al. 2001; Cortes et al. 2005). Therefore, the significant levels of polymorphism present in these antigens create a major challenge for the development of a successful vaccine. Since the antigenic repertoire of a specific antigen is shaped by multiple factors such as the selection strength of host immune responses, functional constraints, and frequency of genetic recombination, detailed studies of antigenic diversity of a vaccine candidate in different malaria-endemic areas must be undertaken (Cui et al., 2003).

At least ten distinct *P. vivax* merozoite surface proteins (PvMSPs) have been identified experimentally or by *in silico* analysis of the *P. vivax* genome (Carlton et al., 2008). Among them, PvMSP4 shows significant homology to PfMSP4 in terms of sequence identity and gene structure (Black et al., 2002). In *P. falciparum* and *P. vivax*, MSP4 and MSP5 have similar gene organizations and are located in tandem on chromosome 2 and 4, respectively, suggesting that these two genes might have originated from an ancient gene duplication event. In rodent malaria parasites, a single gene MSP4/5 is found with similarity to both MSP4 and 5. The PvMSP4 gene has a two-exon structure, with a 300 bp exon 1 and 387 bp exon 2 separated by a 156 bp intron for the Thai-NYU strain. Like PfMSP4, PvMSP4 is a GPI-anchored surface protein with an epidermal growth factor (EGF)-like domain encoded in the second exon (Black et al., 2002).

PfMSP4 is being considered as a component of a subunit vaccine (Doolan and Stewart, 2007). Immunization of mice with *Plasmodium yoelii* MSP4/5 is able to confer significant protection against lethal challenges with blood stage *P. yoelii* (Goschnick et al., 2004). Recombinant PfMSP4 protein is immunogenic in laboratory animals and recognized by antibodies from humans living in malaria endemic areas (Wang et al., 1999; 2001). Limited population studies indicate that PfMSP4 has low sequence diversity (Wang et al., 2002; Benet et al., 2004; Polson et al., 2005), and analysis of PvMSP4 sequences from 30 Colombian samples revealed a similar result of limited polymorphisms (Martinez et al., 2005). Yet, it is not clear whether this reflects a recent expansion of similar strains in this region or global conservation of the *PvMSP4* locus. To determine whether this observation can be extended to other vivax endemic regions, we have sequenced 195 samples from several vivax endemic countries. Our analysis supports the finding that PvMSP4 has limited genetic diversity.

1. Materials and Methods

2.1. Source of *P. vivax* samples

We have procured 195 *P. vivax* finger-prick samples on filter papers from Thailand (n=130), Indonesia (n=34), Brazil (n=24), India (n=1), Papua New Guinea (n=1), Solomon islands (n=1), China (n=2) and Vietnam (n=2). This study was approved by the institutional review board at

Chulalongkorn University. Parasite genomic DNA was extracted by using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). The DNA purification procedure was essentially as described in the manufacturer's instruction manual with minor modifications. The purified DNA was dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and stored at -20°C until use. All samples were genotyped at the highly polymorphic *PvMSP1* locus to confirm that each sample was from a single *P. vivax* isolate (data not show). Further, samples showing superimposed electropherogram signals during *PvMSP4* sequencing were also excluded from analysis.

2.2. Polymerase chain reaction (PCR) and sequencing of PvMSP4

The DNA fragment spanning the entire coding region of the *PvMSP4* gene was amplified using primers based on the *PvMSP4* sequence of the Thai-NYU strain: 5'-ATGAAGGTGGCCTACTTTTGTG-3' and 5'-AGGTGTGTAGGGAGTAGGGA-3'. PCR conditions were denaturation at 94°C for 30 sec, annealing at 60°C for 30 sec and extension at 72°C for 1 min for 35 cycles. To minimize errors introduced in the sequences during PCR amplification, we used ExTaq DNA polymerase (Takara, Shiga, Japan) that has efficient 5' $\rightarrow$ 3' exonuclease activity to increase fidelity and shows no strand displacement. The PCR products were examined by electrophoresis in a 1% agarose gel and visualized with an UV transilluminator.

PCR products were purified using a PCR Purification kit (Qiagen), quantified, and used for direct sequencing. Sequencing was performed in both directions for each template using the BigDye Terminator version 3.1 Cycle Sequencing Kit on an ABI3100 Genetic Analyzer (Applied Biosystems). Overlapping sequences were obtained by using sequencing primers. Whenever singleton occurred, the sequence was re-determined using DNA templates from two independent amplifications from the same DNA samples.

2.3. Data analysis

Sequences were aligned by the Clustal X program with minor manual adjustment made by visual inspection. Haplotype diversity, nucleotide diversity, Tajima's D, Fu & Li's D* and F*, Fu's Fs and minimum number of recombination events (Rm) were estimated using the DnaSP 4.1 software (Rozas et al., 2003). The number of nonsynonymous substitutions per nonsynonymous site (*dN*) and the number of synonymous substitutions per synonymous site (*dS*) were computed using the Nei-Gojobori method as implemented in the MEGA 4 program (Tamura et al., 2007). Interpopulation variance indices (*Fst*) and mismatch distribution were analyzed by the Arlequin software version 3.11 (Excoffier et al., 2005). Prediction of HLA-I binding peptides followed the method of Larsen et al. (2005) at <http://www.cbs.dtu.dk/services/NetCTL>. The *PvMSP4* sequences of the Sal-1 and Thai-NYU strains were retrieved from the GenBank under the accession numbers XM001612944 and AF403475 and used for comparison.

3. Results

3.1. Genetic diversity of the PvMSP4

To assess the global genetic diversity of *P. vivax* parasites at the *PvMSP4* locus, we determined the *PvMSP4* sequence from 195 parasite clinical samples from various vivax malaria-endemic areas. Together with two *PvMSP4* sequences from the Sal-1 and Thai-NYU strain, 53 haplotypes were identified (Table 1). Geographically, the 130 Thai, 34 Indonesian, and 24 Brazilian samples contained 35, 15, and 12 *PvMSP4* haplotypes, respectively. Haplotype 1, 3, and 5 were the most abundant and they accounted for 52% of the *P. vivax* samples from these three countries. There were also shared haplotypes among the regions: 7 (#1, 3, 4, 5, 6, 23, and 26) shared by Thai and Indonesian populations, 6 (#1, 3, 5, 23, 43, and 46) by Indonesian and

Brazilian populations, and 5 (#1, 3, 5, 11, and 23) by Thai and Brazilian populations. Previously, a tetrapeptide repeat motif was found in the second exon preceding the EGF domain, which varies in the number of repeats (Martinez et al., 2005). In addition to this repeat region, we have identified another polymorphic repeat array in exon 1 corresponding to amino acids 47–74 of the Sal I strain (Fig. 1). This repeat array consists of six tetrapeptide motifs (SGGA, SGGS, SGDS, SGGL, SGDA and IGDS), which are repeated 5–9 times in different isolates (Table 1). The second repeat array in exon 2, with the majority of the tetrapeptide motifs as DDHG, varies in copy number of 1–10. The variation of sequence and number of repeat array could be generated from slipped-strand mispairing mechanism, resulting in duplication, deletion or mutation of certain repeat units (such as 'SGGS' and 'SGGL'). The polymorphisms in the two repeat arrays have contributed to most of the haplotype diversity and protein length polymorphism of PvMSP4. If the repeat haplotypes were not taken into account, only 9 haplotypes could be assigned to the 197 PvMSP4 sequences. Outside of the repeats, the sequences are highly conserved with only ten single nucleotide polymorphisms (SNPs) (Supplemental Fig. 1).

Nucleotide diversity (π) was limited in both exons of *PvMSP4* in all three *P. vivax* populations (<0.00063). The Indonesian parasite population had the highest nucleotide diversity whereas the Brazilian population had no segregation sites (Table 2). Two silent SNPs and four amino acid substitutions were observed in exon 1, while four nonsynonymous substitutions were found in exon 2. One amino acid substitution (E/H) occurred in the EGF-like motif, akin to what was observed in the EGF-domain of PvMSP1. Intriguingly, none of the substitutions in the non-repeat regions was shared between Colombian isolates (Martinez et al., 2005) and the isolates analyzed in this study (Supplemental Table 1). The intron was conserved both in size and sequence with only one SNP found in samples from Thailand and Indonesia.

3.2. Selective pressures on PvMSP4

Statistical tests were performed to detect departure from neutrality in exon 1, 2, and entire *PvMSP4* (Table 2). The Tajima's D values for either exon 1, 2, or the entire coding region were negative, while significance was detected for exon 1 or the entire coding regions of *PvMSP4* in the Thai *P. vivax* population. This implies that the *P. vivax* population could have undergone population expansion or have been under purifying selection. A similar trend was observed when the data were analyzed by the Fu and Li's D* and F* tests. However, these tests cannot be implemented with the Brazilian population due to lack of segregating sites. Mismatch analysis using parameters estimated under the sudden population expansion model has shown that the sum of square deviation and Harpending's raggedness index for each population shows a significant departure from such an assumption ($p < 0.000$). Fu's Fs test for selective neutrality did not generate significant values for any populations. Taken together, results from these tests indicate that exon 1 and the entire coding region of *PvMSP4* were under purifying selection.

We have calculated the number of nonsynonymous substitutions per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) in each domain of the *PvMSP4* gene of isolates from Thailand and Indonesia (Table 3). In the entire coding region, the difference between dN and dS did not significantly depart from zero (Table 3). Despite this, in the region spanning nucleotides 598 to 702 in exon 2 (positions referred to the Thai NYU sequence) where a predicted epitope (YDDAEDDDL) was located, dN significantly outnumbered dS ($p < 0.05$), suggesting micro-scale signature of positive selection similar to what was found in PfMSP4 (Jongwutiwes et al., 2002). Closer look into the substituted residues in exons 1 and 2 of *PvMSP4* has identified a number of potential HLA class I-binding epitopes as predicted by the high scores for the C-terminal cleavage and the transporter associated with antigen processing efficiency (Table 4). These epitopes are either clustered upstream of repeat 1 or immediately following repeat 2. It is noteworthy that amino acid substitutions in some of

these peptides abolished the predicted property of the epitopes. However, there was no statistical difference between dN and dS among predicted epitopes in exon 1 (data not shown).

3.3. Genetic differentiation

The extent of geographic differentiation among the Thai, Indonesian, and Brazilian *P. vivax* populations was estimated by using Wright's F -statistics, which revealed genetic differentiation among these three parasite populations (Table 5). The average fixation index of each population was similarly low (0.133–0.144), whereas interpopulation F_{st} differed significantly. The two Southeast Asian populations ($F_{st}=0.042$) were more closely related than they were to the Brazilian population ($F_{st}=0.198$ and 0.300, respectively). Despite the small values of the interpopulation F_{st} indices, they were all significantly different by the permutation test ($p = 0.000 - 0.036$), indicating that genetic differentiation exists among the *P. vivax* populations from these countries.

3.4. Recombination

The minimum number of recombination events in the genealogy of the *PvMSP4* locus of each population was zero, suggesting that intragenic recombination did not contribute to the microheterogeneity of this gene. Linkage disequilibrium analysis implemented in the DnaSP program did not support recombination either.

4. Discussion

In this study, we have characterized the genetic diversity of *PvMSP4* gene using 195 field samples representing diverse endemic areas of *P. vivax* parasite. Our results confirmed the previous finding about the conservation of this gene. Most of the observed haplotype diversity of *PvMSP4* is due to polymorphism in repeat copy number of two repeat arrays. The overall nucleotide diversity of *PvMSP4* is much lower than what has been determined for other *P. vivax* antigens such as *MSP1* (Putaporntip et al., 2002), *MSP3 α* (Rayner et al., 2002; Mascorro et al., 2005), *MSP3 β* (Rayner et al., 2004), *AMA1* (Figtree et al., 2000; Gunasekera et al., 2007; Grynberg et al., 2008), or even *TRAP* (Putaporntip et al., 2001). Moreover, the genetic diversity of *PvMSP4* was almost ten times lower than that for *PfMSP4* (Jongwutiwes et al., 2002; Benet et al., 2004; Polson et al., 2005). Our finding from this study is consistent with the notion that haplotype diversity is positively correlated with the level of malaria endemicity. Analyses using both microsatellite and antigenic markers have detected much higher genetic diversity of *P. vivax* populations in Asia than in South America, which might reflect the population history of the parasite (Gunasekera et al., 2007; Ferreira et al., 2007; Imwong et al., 2007; Karunaweera et al., 2008). Our results are in stark contrast to those from an earlier study of Colombian *P. vivax* samples although that study also revealed limited polymorphisms within the *PvMSP4* gene (Martinez et al., 2005). Not only did the *P. vivax* population from Colombia have much higher nucleotide diversity than this work, but none of the SNPs were found in the parasite samples in this study. We don't know whether this discrepancy truly reflects geographical differentiation of the parasite populations in South American regions or is due to sequencing errors.

Different tests for selection suggest the occurrence of opposing forces of selection on the *PvMSP4* locus. As seen in the Duffy-binding protein, this is often imposed by the host immune pressure and functional constraint of the protein. Unlike most malaria parasite surface antigens that are under balancing selection, multiple neutrality tests strongly suggest that *PvMSP4* has been under purifying selection. Interestingly, several amino acid substitutions in both exons might be major HLA epitopes, suggesting that the polymorphisms observed were likely selected by the host immune systems. The presence of micro-scale signature of positive selection at some clustered sites in exon 2 where a predicted T cell epitope was located and

the lack of such signature in T cell epitopes in exon 1 could imply that the former epitope might be under more efficient immune selection than the latter. Nevertheless, further experimental immunological correlates are undoubtedly required to address this issue. The lack of balancing selection on this merozoite surface protein might be due to functional constraints, since PfMSP4 is essential for *P. falciparum* invasion of red blood cells and genetic knockout of this gene cannot be obtained (Cowman and Crabb, 2006). The limited polymorphism in PvMSP4 in global samples is in sharp contrast to its highly polymorphic paralogue PvMSP5 (Gomez et al., 2005). This could be due to their functional distinctions and/or different locations in the merozoite, which impose different levels of exposure of the molecules to the host immunity (Black et al., 2002).

F statistics detected different levels of significant genetic differentiation among the Thai, Indonesian, and Brazilian parasite populations, which is also reflected in the distribution patterns of individual PvMSP4 haplotypes in different parasite populations. While large geographical (continental) population differentiation is apparent, the populations appear to also have significant differences even within each continent. In particular, significant genetic differentiation exists between the two Asian populations, and between the Brazilian and Colombian populations. Of a similar notice is the different genetic diversity in our Thai samples, where 53 isolates were obtained from Southern Thailand. These samples had very limited haplotype diversity, and most of them were PvMSP4 haplotypes 5 (n=40) and 1 (n=11). This genetic structure resembles clonal expansion of the parasite, which is in striking contrast to the persistent and highly heterogeneous populations we have observed in western Thailand (Putaporntip et al., 2002; Mascorro et al., 2005). Both the macro-and micro-geographical genetic differences may be related to the demographic histories of the parasite populations. In Brazil, the lack of segregating sites at the PvMSP4 locus suggests recent spread of the parasite, a phenomenon also supported by the analysis of the PvAMA1 gene (Grynberg et al., 2008). In southern Thailand, the area has experienced a recent increase in *P. vivax* prevalence, which might explain the observed extreme paucity of PvMSP4 haplotypes. Taken together, the evolution of parasite antigens as candidates of vaccines needs to be examined on a detailed epidemiological scale.

Since malaria vaccines usually have to take the allele-specific immunity into account, the conservation in PvMSP4 in global parasite populations makes it easier to design a vaccine to cover the antigenic repertoire of PvMSP4. Functional characterization of the immunogenicity of PvMSP4 would allow further evaluation of its vaccine potential. Besides, its similarity to PfMSP4 especially in the EGF-like domain (67% identity) suggests that cross-species immunity to MSP4 may be elicited from natural exposure to either parasite species (Woodberry et al., 2008). Such vaccines may prove to be useful to control malaria in areas where both parasite species are endemic.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to all patients who donated their blood samples for this analysis. This work was supported by research grants from The National Research Council of Thailand and The Thai Government Research Budget to S.J. and C.P., a grant from The Thailand Research Fund (RMU5080002) to C.P. and a grant (D43-TW006571) from Fogarty International Center, National Institutes of Health to L.C.

References

- Baird JK. Neglect of *Plasmodium vivax* malaria. Trends Parasitol 2007;23:533–539. [PubMed: 17933585]
- Barcus MJ, Basri H, Picarima H, Manyakori C, Sekartuti, Elyazar I, Bangs MJ, Maguire JD, Baird JK. Demographic risk factors for severe and fatal vivax and falciparum malaria among hospital admissions in northeastern Indonesian Papua. Am J Trop Med Hyg 2007;77:984–991. [PubMed: 17984364]
- Benet A, Livingstone T, Reeder JC, Cortes A. Diversity of the *Plasmodium falciparum* vaccine candidate merozoite surface protein 4 (MSP4) in a natural population. Mol Biochem Parasitol 2004;134:275–280. [PubMed: 15003847]
- Black CG, Barnwell JW, Huber CS, Galinski MR, Coppel RL. The *Plasmodium vivax* homologues of merozoite surface proteins 4 and 5 from *Plasmodium falciparum* are expressed at different locations in the merozoite. Mol Biochem Parasitol 2002;120:215–224. [PubMed: 11897127]
- Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, Del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczyk M, Meyer EV, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, Salzberg L, Stoeckert CJ, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 2008;455:757–763. [PubMed: 18843361]
- Cortes A, Mellombo M, Masciantonio R, Murphy VJ, Reeder JC, Anders RF. Allele specificity of naturally acquired antibody responses against *Plasmodium falciparum* apical membrane antigen 1. Infect Immun 2005;73:422–430. [PubMed: 15618180]
- Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. Cell 2006;124:755–766. [PubMed: 16497586]
- Cui L, Escalante AA, Imwong M, Snounou G. The genetic diversity of *Plasmodium vivax* populations. Trends Parasitol 2003;19:220–226. [PubMed: 12763428]
- Dooland DL, Steward VN. Status of malaria vaccine R&D in 2007. Expert Rev Vaccines 2007;6:903–905. [PubMed: 18377353]
- Excoffier L, Laval G, Schneider S. Arlequin ver.3.0: An integrated software package for population genetics data analysis. Evol Bioinform Online 2005;1:47–50.
- Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL. Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. J Infect Dis 2007;195:1218–1226. [PubMed: 17357061]
- Figtree M, Pasay CJ, Slade R, Cheng Q, Cloonan N, Walker J, Saul A. *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. Mol Biochem Parasitol 2000;108:53–66. [PubMed: 10802318]
- Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, Alpers MP, Müller I. *Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea. PLoS Med 2008;5:e127. [PubMed: 18563961]
- Gomez A, Suarez CF, Martinex P, Saravia C, Patarroyo MA. High polymorphisms in *Plasmodium vivax* merozoite surface protein-5 (MSP5). Parasitology 2006;133:661–672. [PubMed: 16978450]
- Goschnick MW, Black CG, Kedzierski L, Holder AA, Coppel RL. Merozoite surface protein 4/5 provides protection against lethal challenge with a heterologous malaria parasite strain. Infect Immun 2004;72:5840–5849. [PubMed: 15385485]
- Grynberg P, Fontes CJ, Hughes AL, Braga EM. Polymorphism at the apical membrane antigen 1 locus reflects the world population history of *Plasmodium vivax*. BMC Evol Biol 2008;8:123. [PubMed: 18445274]
- Guerra CA, Snow RW, Hay SI. Mapping the global extent of malaria in 2005. Trends Parasitol 2006;22:353–358. [PubMed: 16798089]
- Gunasekera AM, Wickramarachchi T, Neafsey DE, Ganguli I, Perera L, Premaratne PH, Hartl D, Handunnetti SM, Udagama-Randeniya PV, Wirth DF. Genetic diversity and selection at the *Plasmodium vivax* apical membrane antigen-1 (PvAMA-1) locus in a Sri Lankan population. Mol Biol Evol 2007;24:939–947. [PubMed: 17244598]

- Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis* 2004;4:327–336. [PubMed: 15172341]
- Hodder AN, Crewther PE, Anders RF. Specificity of the protective antibody response to apical membrane antigen 1. *Infect Immun* 2001;69:3286–3294. [PubMed: 11292751]
- Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, Mayxay M, Newton PN, Kim JR, Nandy A, Osorio L, Carlton JM, White NJ, Day NP, Anderson TJ. Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. *Int J Parasitol* 2007;37:1013–1022. [PubMed: 17442318]
- Jongwutiwes S, Putaporntip C, Friedman R, Hughes AL. The extent of nucleotide polymorphism is highly variable across a 3-kb region on *Plasmodium falciparum* chromosome 2. *Mol Biol Evol* 2002;19:1585–1590. [PubMed: 12200485]
- Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL, Kawamoto F, Hartl DL, Wirth DF. Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. *Gene* 2008;410:105–112. [PubMed: 18226474]
- Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M. 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* 2005;35:2295–2303. [PubMed: 15997466]
- Martinez P, Suarez CF, Gomez A, Cardenas PP, Guerrero JE, Patarroyo MA. High level of conservation in *Plasmodium vivax* merozoite surface protein 4 (PvMSP4). *Infect Genet Evol* 2005;5:354–361. [PubMed: 16168942]
- Mascorro CN, Zhao K, Khuntirat B, Sattabongkot J, Yan G, Escalante AA, Cui L. Molecular evolution and intragenic recombination of the merozoite surface protein MSP-3_ from the malaria parasite *Plasmodium vivax* in Thailand. *Parasitology* 2005;131:25–35. [PubMed: 16038393]
- Mendis K, Sina BJ, Marchesini P, Carter R. The neglected burden of *Plasmodium vivax* malaria. *Am J Trop Med Hyg* 2001;64:97–106. [PubMed: 11425182]
- Polson HE, Conway DJ, Fandeur T, Mercereau-Puijalon O, Longacre S. Gene polymorphism of *Plasmodium falciparum* merozoite surface proteins 4 and 5. *Mol Biochem Parasitol* 2005;142:110–115. [PubMed: 15878789]
- Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 2007;77(6 Suppl):79–87. [PubMed: 18165478]
- Putaporntip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho WG, Kaneko A, Kanbara H, Hattori T, Tanabe K. Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. *Proc Natl Acad Sci USA* 2002;99:16348–16353. [PubMed: 12466500]
- Putaporntip C, Jongwutiwes S, Tia T, Ferreira MU, Kanbara H, Tanabe K. Diversity in the thrombospondin-related adhesive protein gene (TRAP) of *Plasmodium vivax*. *Gene* 2001;268:97–104. [PubMed: 11368905]
- Rayner JC, Corredor V, Feldman D, Ingravallo P, Iderabdullah F, Galinski MR, Barnwell JW. Extensive polymorphism in the *Plasmodium vivax* merozoite surface coat protein MSP-3 α is limited to specific domains. *Parasitology* 2002;125:393–405. [PubMed: 12458823]
- Rayner JC, Huber CS, Feldman D, Ingravallo P, Galinski MR, Barnwell JW. *Plasmodium vivax* merozoite surface protein PvMSP-3 beta is radically polymorphic through mutation and large insertions and deletions. *Infect Genet Evol* 2004;4:309–319. [PubMed: 15374528]
- Richie TL, Saul A. Progress and challenges for malaria vaccines. *Nature* 2002;415:694–701. [PubMed: 11832958]
- Roberts L, Enserink M. Malaria. Did they really say eradication? *Science* 2007;318:1544–1545. [PubMed: 18063766]
- Rodriguez-Morales AJ, Benítez JA, Arria M. Malaria mortality in Venezuela: focus on deaths due to *Plasmodium vivax* in children. *J Trop Pediatr* 2008;54:94–101. [PubMed: 17906318]
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003;19:2496–2497. [PubMed: 14668244]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599. [PubMed: 17488738]

- Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M, Lampah DA, Price RN. Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. PLoS Med 2008;5:e128. [PubMed: 18563962]
- Wang L, Black CG, Marshall VM, Coppel RL. Structural and antigenic properties of merozoite surface protein 4 of *Plasmodium falciparum*. Infect Immun 1999;67:2193–2200. [PubMed: 10225874]
- Wang L, Marshall VM, Coppel RL. Limited polymorphism of the vaccine candidate merozoite surface protein 4 of *Plasmodium falciparum*. Mol Biochem Parasitol 2002;120:301–303. [PubMed: 11897136]
- Wang L, Richie TL, Stowers A, Nhan DH, Coppel RL. Naturally acquired antibody responses to *Plasmodium falciparum* merozoite surface protein 4 in a population living in an area of endemicity in Vietnam. Infect Immun 2001;69:4390–4397. [PubMed: 11401978]
- Woodberry T, Minigo G, Piera KA, Hanley JC, de Silva HD, Salwati E, Kenangalem E, Tjitra E, Coppel RL, Price RN, Anstey NM, Plebanski M. Antibodies to *Plasmodium falciparum* and *Plasmodium vivax* merozoite surface protein 5 in Indonesia: species-specific and cross-reactive responses. J Infect Dis 2008;198:134–142. [PubMed: 18471084]

Fig. 1. Genomic organization of the *PvMSP4* gene showing the two-exon structure, the positions of the two repeat arrays (Rep), the EGF-like domain, and GPI-anchor sequence. Numbers are based on the *P. vivax* Sal I strain.

Table 1
Summary of PvMSP4 haplotypes (H) and their geographical distribution among the Thai, Indonesian (Ind.) and Brazilian (Br.) *P. vivax* populations

H	Exon-1					Intron SNP (A/G)	Exon-2				Distribution (n)						
	19	23	44	48	87		91	Repeats*	118	168	171	202	Repeats*	Thai (130)	Ind. (34)	Br. (24)	Total (188)
1	Y	R	G	G	D	S	122342	A	Y	D	D	N	7777	17	1	1	19
2	Y	R	G	G	D	S	122342	A	Y	N	D	N	777777	2	0	0	2
3	Y	R	G	G	D	S	122342	A	Y	D	D	N	778	9	5	2	16
4	Y	R	G	G	D	S	122342	A	Y	D	D	N	77	3	4	0	7
5	Y	R	G	G	D	S	122342	A	Y	D	D	N	7778	48	11	4	63
6	Y	R	G	G	D	S	122342	A	Y	D	D	N	777	7	1	0	8
7	Y	R	G	G	D	S	1223342	A	Y	D	D	N	7777778	1	0	0	1
8	Y	R	G	G	D	S	12222342	A	Y	D	D	N	77777	4	0	0	4
9	Y	R	G	G	Y	S	1222342	A	Y	D	D	N	777777	1	0	0	1
10	Y	R	G	G	D	S	1223442	A	Y	D	D	N	7777	1	0	0	1
11	Y	R	G	G	D	S	1223442	A	Y	D	D	N	77777	2	0	2	4
12	Y	R	G	G	D	S	122342	A	Y	D	D	N	7777778	2	0	0	2
13	Y	R	G	G	D	S	122342	A	Y	D	D	N	77778	2	0	0	2
14	Y	R	G	G	D	S	1222342	A	Y	D	D	N	7778	1	0	0	1
15	Y	R	G	G	D	S	122342	A	Y	D	D	N	777777	2	0	0	2
16	Y	R	G	g	D	S	22342	A	Y	D	D	N	7777778	2	0	0	2
17	Y	R	G	G	D	S	122344242	A	Y	D	D	N	7	1	0	0	1
18	Y	R	G	G	D	S	1223442	A	Y	D	D	N	77778	1	0	0	1
19	Y	R	G	G	D	S	1222342	A	Y	D	D	N	777	7	0	0	7
20	Y	R	G	G	D	S	122342	A	Y	N	D	N	7777	1	0	0	1
21	Y	R	G	G	D	S	122342	A	Y	N	D	N	77777	1	0	0	1
22	Y	R	G	G	D	S	1223442	A	Y	D	D	N	7	1	0	0	1
23	Y	R	G	G	D	S	122342	A	Y	D	D	N	77777	1	2	2	5
24	Y	R	G	G	D	S	12234442	A	Y	N	D	N	7777777	1	0	0	1
25	Y	R	G	G	D	S	1222342	A	Y	D	D	N	77	2	0	0	2
26	Y	R	G	G	D	S	122342	G	Y	D	D	N	7778	1	1	0	2
27	Y	R	G	G	D	S	12222342	A	Y	D	D	N	78	1	0	0	1

H	Exon-1						Intron SNP (A/G)	Exon-2				Distribution (n)					
	19	23	44	48	87	91		Repeats*	118	168	171	202	Repeats*	Thai (130)	Ind. (34)	Br. (24)	Total (188)
28	Y	K	G	G	D	S	122342	A	Y	D	D	N	777	1	0	0	1
29	Y	R	G	G	D	S	122232342	A	Y	D	D	N	77777	1	0	0	1
30	Y	R	G	G	D	S	122342	A	Y	D	D	H	7777	1	0	0	1
31	Y	R	G	G	D	S	1222342	A	Y	N	D	N	7777777	1	0	0	1
32	F	R	G	G	D	S	122342	A	Y	D	D	N	77777778	1	0	0	1
33	Y	R	G	G	D	S	122342	A	Y	D	D	N	7777	1	0	0	1
34	Y	R	G	G	D	S	122342	A	F	D	D	N	7777	1	0	0	1
35	Y	R	G	G	D	S	1122342	A	Y	D	D	N	778	1	0	0	1
36	Y	R	G	G	D	S	12342	A	Y	D	D	N	7777778	0	0	0	0
37	Y	R	G	G	D	S	12342	A	Y	D	D	N	7778	0	0	0	0
38	Y	R	G	G	D	S	12342	A	Y	D	D	N	77777	0	0	0	0
39	Y	R	G	G	D	S	122342	A	Y	D	D	N	7	0	1	0	1
40	Y	R	g	G	D	S	122342	A	Y	D	D	N	7778	0	1	0	1
41	Y	R	G	G	D	L	122342	A	Y	D	D	N	777	0	1	0	1
42	Y	R	G	G	D	L	122342	A	Y	D	D	N	778	0	2	0	2
43	Y	R	G	G	D	S	122342	A	Y	D	D	N	7777777	0	1	2	3
44	Y	R	G	G	D	S	122542	A	Y	D	D	N	777	0	1	0	1
45	Y	R	G	G	D	S	122342	A	Y	D	N	N	7778	0	1	0	1
46	Y	R	G	G	D	S	122342	A	Y	D	D	N	777778	0	1	3	4
47	Y	R	G	G	D	S	122342	A	Y	D	N	N	77778	0	0	0	0
48	Y	R	G	G	D	S	12342	A	Y	D	D	N	7777778	0	0	2	2
49	Y	R	G	G	D	S	122342	A	Y	D	D	N	777777778	0	0	1	1
50	Y	R	G	G	D	S	12342	A	Y	D	D	N	7777777	0	0	1	1
51	Y	R	G	G	D	S	12342	A	Y	D	D	N	77778	0	0	1	1
52	Y	R	G	G	D	S	122342	A	Y	D	D	N	7777778	0	0	3	3
53	Y	R	G	G	D	S	122342	A	Y	D	D	N	777777777	0	0	0	0

Substitutions are highlighted. Lower case letters denote synonymous substitutions. Codon number is based on the PnMSP4 of Sal I strain.

* Repeat barcodes: 1 = SGGA; 2 = SGG; 3 = SGDS; 4 = SGGL; 5 = IGDS; 6 = SGDA; 7 = DDHG and 8 = DDQG.

Table 2 Polymorphism and neutrality tests for *PvMSP4* gene among Thai, Indonesian and Brazilian isolates

Population	Domain	N	S	K	H±SD	π±SD	Tajima's D	Fu & Li's D*	Fu & Li's F*
Thailand	Exon 1	130	4	5	0.119±0.038	0.00048±0.00016	-1.58661	-2.80103*	-2.83676*
	Exon 2	130	3	4	0.118±0.038	0.00034±0.00011	-1.34288	-2.09312	-2.18058
	All	130	7	9	0.216±0.048	0.00037±0.00009	-1.83279*	-3.35248**	-3.35929**
Indonesia	Exon 1	34	2	3	0.116±0.074	0.00045±0.00029	-1.49966	-2.34612	-2.43332
	Exon 2	34	2	3	0.219±0.090	0.00063±0.00027	-1.06791	-0.77789	-0.99545
	All	34	4	5	0.321±0.100	0.00052±0.00018	-1.61558	-2.06368	-2.24505
Brazil	All	24	0	1	0.000±0.000	0.00000±0.00000			

N, number of isolate; S, number of polymorphic sites; K, number of haplotype; H, haplotype diversity; π, nucleotide diversity; SD, standard deviation.

* and ** indicate statistical significance at 0.05 and 0.02 level.

Table 3
Mean and standard error values of synonymous (dS) and nonsynonymous (dN) substitutions within PvMSP4*

Domain	Population	H	d _S ± S.E.	d _N ± S.E.
Exon 1	Thailand	15	0.0049 ± 0.0047	0.0018 ± 0.0010
	Indonesia	4	0.0068 ± 0.0069	0.0044 ± 0.0031
	All	18	0.0058 ± 0.0044	0.0026 ± 0.0011
Exon 2	Thailand	20	0 ± 0	0.0021 ± 0.0015
	Indonesia	10	0 ± 0	0.0007 ± 0.0007
	All	23	0 ± 0	0.0022 ± 0.0013
Nucleotides 598–702**	Thailand	3	0 ± 0	0.0156 ± 0.0098
	Indonesia	2	0 ± 0	0.0117 ± 0.0110
	All	4	0 ± 0	0.0176 ± 0.0089***
Exon 1 & 2	Thailand	35	0.0014 ± 0.0014	0.0011 ± 0.0005
	Indonesia	15	0.0009 ± 0.0009	0.0010 ± 0.0006
	All	43	0.0015 ± 0.0013	0.0013 ± 0.0005

* d_S and d_N are shown as mean ± standard error (S.E.).

** Positions are after *PvMSP4* of the Thai NYU strain. Tests of the hypothesis that mean d_S equals that of d_N:

*** $p < 0.05$.

Table 4

Amino acid substitutions that influence predicted HLA-bound peptides in PvMSP4.

Domain	Predicted epitope	HLA Supertype	Predicted score	Isolates
Exon 1	DLLIIFSL <u>Y</u>	A1	1.0	Most common
	DLLIIFSL <u>F</u>	A1	0.5	TSY769 (Thailand)
	DLLIIFSL <u>Y</u>	A26	1.2	Most common
	DLLIIFSL <u>F</u>	A26	0.6	TSY769 (Thailand)
	GRRSAFAGIA	B27	1.2	Most common
	G <u>K</u> RSAFAGIA	B27	0.4	TSY525 (Thailand)
	<u>Y</u> FDGRRSAF	B62	0.9	Most common
	<u>F</u> FDGRRSAF	B62	0.7	TSY769 (Thailand)
Exon 2	YDD <u>A</u> EDDDL	B39	1.0	Most common
	YD <u>N</u> AEDDDL	B39	0.6	T064, T667, T752, T36 and TSY649 (Thailand) PBR20 (Indonesia) AD93 (Brazil)

* Threshold for epitope identification was set at ≥ 0.75 using weight on C terminal cleavage at ≥ 0.15 and weight on transporter associated with antigen processing transport efficiency at ≥ 0.05 .

Table 5Interpopulation variance indices of *P. vivax* populations from Thailand, Indonesia and Brazil*

	Thailand	Indonesia	Brazil
Thailand	0.138	0.036	0.000
Indonesia	0.042	0.144	0.000
Brazil	0.198	0.300	0.133

* Pairwise *F*_{st} values (lower diagonal) between *P. vivax* populations and their permuted *p* values (upper diagonal). Interdiagonal represents population specific *F*_{st} indices.