Published in final edited form as:

Infect Genet Evol. 2009 December ; 9(6): 1295–1300. doi:10.1016/j.meegid.2009.07.005.

Nucleotide sequence polymorphism at the apical membrane antigen-1 locus reveals population history of *Plasmodium vivax* in Thailand

Chaturong Putaporntip¹, Somchai Jongwutiwes^{1,#}, Priscila Grynberg², Liwang Cui³, and Austin L. Hughes^{4,#}

¹Molecular Biology of Malaria and Opportunistic Parasites Research Unit, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand

²Departamento de Bioquímica e Immunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil ³Department of Entomology, The Pennsylvania State University, PA16802, USA ⁴Division of Biological Sciences, University of South Carolina, SC29208, USA.

Abstract

Apical membrane antigen-1 is a candidate for inclusion in a vaccine for the human malaria parasite *Plasmodium vivax*. We collected 231 complete sequences of the gene encoding this antigen (*pvama-1*) from three regions of Thailand, the most extensive collection to date of sequences at this locus. The domain II loop (previously mentioned as a potential vaccine component) was almost completely conserved, with a single amino acid variant (I313R) observed in a single sequence. The 3' portion of the gene (domain II through the stop codon) showed significantly lower nucleotide diversity than the 5' portion (start codon through domain I); and a given domain I sequence might be found in a haplotype with more than one domain II sequence. These results imply a hotspot of recombination between domains I and II. We found significant geographic subdivision among the three regions of Thailand (NW, East, and South) in which collections were made in 2007. Numbers of *P. vivax* infections have experienced overall declines since 1990 in all three regions; but the decline has been most recent in the NW, and there has been a rebound in numbers of infections in the South since 2000. Consistent with population history, amino acid sequence diversity was greatest in the NW. The South, which had by far the lowest sequence diversity of the three regions, showed signs of a population that has expanded from a small number of founders after a bottleneck.

Plasmodium vivax is the most widely distributed geographically of the malaria parasites infecting humans and is responsible for 100-300 million cases annually, especially in South America, Asia and Oceania; yet researchers have comparatively neglected this parasite (Baird 2007). As with other malaria parasites, several of the surface antigens of *P. vivax* have been considered as potential vaccine candidates (Webster and Hill 2003). One of these is apical membrane antigen 1 (AMA-1), which is expressed on the apical surface of the merozoite, the

© 2009 Elsevier B.V. All rights reserved.

#Corresponding authors: A.L. Hughes (phone: 1-803-777-9186; fax: 1-803-777-4002 austin@biol.sc.edu) and S. Jongwutiwes (phone: 662-2564000 ext 3685; fax: 662-2525944; fmedsjw@md.chula.ac.th or jongwutiwes@gmail.com)..

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.

stage that invades host erythrocytes, and is believed to play a role in erythrocyte invasion (Hodder et al. 2001).

A number of studies have addressed polymorphism at the locus encoding AMA-1 of *P. vivax* (the *pvama-1* locus) in various parts of the world (Figtree et al. 2000; Grynberg et al. 2008; Gunasakera et al. 2007; Rajesh et al. 2007). Understanding this polymorphism is a necessary prerequisite for development of an effective vaccine based on AMA-1 (Gunasakera et al. 2007). For example, a highly polymorphic antigen might still be used as the basis for a vaccine if polymorphism in local populations is limited, allowing the development of population-specific vaccines (Tanabe et al. 2004). Alternatively, if certain domains of a polymorphic protein show little polymorphism yet elicit an immune response, such domains might be potential vaccine targets (Gunasakera et al. 2007).

The AMA-1 proteins of both *Plasmodium falciparum* (PfAMA-1) and *P. vivax* (PvAMA-1) include three extracellular domains, designated domains I, II, and III in N- to C-terminal order (Pizarro et al. 2005). In the case of *P. falciparum*, there is strong evidence that amino acid sequence polymorphism in domains I, II, and III is selectively maintained, as indicated by a statistically significant excess of nonsynonymous nucleotide substitutions over synonymous substitutions in the regions of the gene encoding those domains (Verra and Hughes 2000).

In a sample of 23 *pvama-1* sequences from Sri Lanka, nonsynonymous nucleotide diversity (π_N) was significantly greater than synonymous (π_S) nucleotide diversity in domain II, suggesting that polymorphism in the latter domain is selectively maintained (Gunasakera et al. 2007). On the other hand, these authors found no amino acid sequence polymorphism in an unstructured loop within domain II that is known to be a target for host antibodies (Pizarro et al. 2005), leading to the suggestion that this loop might provide a vaccine component (Gunasakera et al. 2007). Gunasakera et al. (2007) also suggested that there may be certain recombinational hotspots within the *pvama-1* coding region. However, most *pvama-1* sequences that have so far been sampled from natural populations have been partial; thus the action of natural selection and interallelic recombination on different domains of the protein remains poorly understood.

Comparison of a *pvama-1* sequences from domain I of 320 isolates from around the world showed a striking contrast between Old World populations and those from Brazil (Grynberg et al. 2008). In the Brazilian Amazon, where *P. vivax* presumably is of post-Columbian origin, *pvama-1* sequences showed little evidence of population subdivision, consistent with a rapid recent expansion of the parasite (Grynberg et al. 2008). By contrast, there was evidence of substantial subdivision in Asia, with high F_{ST} statistics between *pvama-1* sequences collected from different Southeast Asian countries (Grynberg et al. 2008). However, previous studies have not provided evidence regarding the degree of population subdivision at a finer geographic scale. .

Here we report the largest sample yet analyzed (231 sequences) of complete *pvama-1* sequences from field isolates, isolated from three different geographic regions within a single Southeast Asian country, Thailand (Figure 1). Using comparative sequence analysis, we test for the role of natural selection in maintaining polymorphism at this locus, and the effects of inter-allelic recombination on the extent of polymorphism across the coding region. We analyze differences in allelic frequencies among the three regions in order to test for population subdivision at this locus. Finally, we present data on the cases of *P. vivax* malaria recorded over a thirty-year period in the three provinces from which our sequences were derived, enabling us to look for correlations between the observed pattern of *pvama-1* sequence polymorphism and the population history of the parasite on a short time scale.

Materials and Methods

P. vivax Samples

Pvama-1 sequences were collected from symptomatic malaria patients who acquired the infections from three endemic areas of Thailand, bordering Myanmar, Cambodia and Malaysia, respectively: (1) Tak Province (here designated Northwest or NW); (2) Chantaburi Province (designated East); and Yala and Narathiwat Provinces (designated South; Figure 1). A total of 173 isolates were collected in 2007: 44 from the NW, 56 from the East, and 73 from the South (25 from Yala and 48 from Narathiwat Provinces); and 58 isolates were collected from the NW in 1996. These isolates contained single infections based on genotyping at the highly variable domain of block 6 of the merozoite surface protein 1 locus (Putaporntip et al 2002). We compiled annual totals from 1979-2008 of the numbers of cases of *P. vivax* malaria in the three geographical regions where collections were made, using data from the Division of Vector-Borne Diseases, Ministry of Public Health, Thailand. The ethical aspects of this study have been approved by the Institutional Review Board of Faculty of Medicine, Chulalongkorn University.

Amplification of the *PvAMA-1* gene

DNA of these malaria blood samples were extracted by using the QIAGEN DNA minikit (Hilden, Germany) following the manufacturer's protocol. After the purification procedure, DNA samples were stored at -30°C until use. The complete coding region of *PvAMA-1* was amplified by polymerase chain reaction (PCR) using forward and reverse primers, *PvAMA1-F0*: 5'-GCAGAGAGAGCAAACCAAATCG-3' and *PvAMA1-R0*: 5'-GGCAAGCGAGTTGGCCAAGCAAA-3', respectively. Thermal cycling profiles contained a preamplification denaturation at 94°C , 1 min; 35 cycles of denaturation at 94°C , 30 s; annealing at 60°C , 30 s; extension at 72°C , 2 min, and post amplification extension at 72°C , 5 min. DNA amplification was performed by using a GeneAmp 9700 PCR thermal cycler (Applied Biosystems, Foster City, CA). To minimize the error introduced in the sequences during PCR amplification, we used ExTaq DNA polymerase because it possesses efficient 5' $\rightarrow$ 3' exonuclease activity to increase fidelity and no strand displacement (Takara, Japan). The size of PCR product was examined by electrophoresis in a 1% agarose gel and visualized under a UV transilluminator (BioRad Molecular Imager ChemiDoc™ XRS, USA).

DNA sequencing

DNA sequences were determined directly and from both directions for each template using the Big Dye Terminator v3.1 Cycle Sequencing Kit on an ABI3100 Genetic Analyzer (Applied Biosystems, USA). Overlapping sequences were obtained by using sequencing primers. When singleton substitution was detected, sequence was re-determined using PCR products from two independent amplifications using the same DNA template. DNA sequences were submitted to the Genbank database under the accession numbers FJH784891-FJ785121.

Statistical Methods

Coding sequences were aligned using the CLUSTAL W program (Thompson et al. 1997). We used Nei and Gojobori's (1986) method to estimate d_N and the number of synonymous substitutions per synonymous site (d_S). In preliminary analyses, the methods of Li (1993) and Yang and Nielsen (2000) yielded essentially identical results, as expected because the number of substitutions per site was low in this case (Nei and Kumar 2000). We computed the mean of all pairwise d_S values, designated the synonymous nucleotide diversity (π_S); and the mean of all pairwise d_N values, designated the nonsynonymous nucleotide diversity (π_N). Standard errors of π_S and π_N were estimated by the bootstrap method (Nei and Kumar 2000); 1000 bootstrap samples were used. We estimated π_S and π_N separately for the following regions

(Gunasakera et al. 2007; Pizarro et al. 2005): (1) the 5' region (including the signal peptide) from the start codon to the start of domain I (codons 1-42); (2) domain I (codons 43-248); domain II (codons 249-385), including the domain II loop (codons 293-334); domain III (codons 386-487); the 3' region from the end of domain III to the stop codon, including the transmembrane and cytoplasmic regions (codons 488-563).

Pairwise F_{ST} among populations was estimated using the Arlequin 3.11 software (Excoffier et al. 2005). The gene diversity (heterozygosity) of amino acid sequence haplotypes in domains I and II was estimated by the formula:

$$1 - \sum_{i=1}^n x_i^2$$

where n is the number of alleles and x_i is the population frequency of the i^{th} allele (Nei 1987, p. 177) expressed as a percentage. Tajima's (1989) D test was used to test for balancing selection and/or other effects of population history within samples.

Because pairwise sequence comparisons are not independent, the significance of the correlation between the pairwise proportion of amino acid difference in domain I and that in domain II was tested by a randomization test. In the randomization test, 1000 pseudo data sets were created by sampling from the data with replacement; and the observed correlation coefficient was compared with the distribution of coefficients computed for the pseudo data sets.

Results

P. vivax Malaria 1979-2008

In all three of the regions where we obtained *pvama-1* sequences (Figure 1), there was an overall decline in the number of reported cases of *P. vivax* malaria over the 30-year interval 1979-2008 (Figure 2). However, the patterns differed substantially among geographic regions. Around 1991, a precipitous decline in cases occurred in the East and South, but not in the NW (Figure 2). In the NW, 58.5% of the cases in the data set occurred after 1990, while in the East only 35.7% of cases occurred after 1990 and in the South only 28.3% of cases occurred after 1990 (Figure 2). This difference in proportions $\chi^2 = 16471.6$, 1 d.f. $P < 0.001$) and in the comparison between the NW and the South, ($\chi^2 = 33388.5$, 1 d.f. $P < 0.001$).

Beginning in about the year 2000, a decline in numbers of cases began in the NW (Figure 2). For the nine years 2000-2008, the number of cases in the NW was strongly negatively correlated with the year ($r = -0.862$; $P = 0.003$). Over the same time period, the number of cases in the East was also negatively correlated with the year ($r = -0.856$; $P = 0.003$). In the South, on the other hand, the number of cases, which had fallen since about 1991, began to increase (Figure 2). In fact, over the years 2000-2008, the number of cases in the South was significantly positively correlated with the year ($r = 0.777$; $P = 0.014$).

Synonymous and Nonsynonymous Nucleotide Diversity

For 231 *pvama-1* sequences from Thailand, we estimated synonymous (π_S) and nonsynonymous (π_N) nucleotide diversity separately for the three extracellular domains and for the coding regions located 5' and 3' to these domains (Table 1). In the case of domain II, we analyzed the loop region and the remainder of the domain separately (Table 1). None of the regions showed π_S significantly different from π_N (Table 1). In the present data, π_N was slightly greater than π_S in domain II as a whole and in the portion of the domain excluding the loop, but the difference was not significant in either case (Table 1). For the entire coding region,

π_S was slightly but not significantly greater than π_N (Table 2). Similar patterns were seen when the data for the four collections (NW 1996, NW 2007, East 2007, and South 2007) were analyzed separately (not shown).

As reported by Gunasakera et al. (2007), the domain II loop was highly conserved at the amino acid sequence level, with π_N about 1 in 10,000 (Table 1). However, unlike the data of Gunasakera et al. (2007), we did find one polymorphic amino acid site in the domain II loop. At amino acid position 313, one sequence collected in the NW in 1996 showed Ile instead of Arg. The near-absence of nonsynonymous polymorphism supported the hypothesis that this loop is subject to strong purifying selection. This hypothesis was further supported by the fact that π_N in the loop (0.0001 ± 0.0001 ; Table 1) was significantly less than that in the remainder of the gene (0.0076 ± 0.0016 ; Z-test; $P < 0.001$). Likewise, although there was no synonymous polymorphism in the loop to serve as a basis for comparison, π_N in the loop was significantly less than π_S in the whole gene (0.0091 ± 0.0027 ; Table 1; Z-test; $P < 0.001$).

Both π_S and π_N values were significantly greater in domain I than in the domain II loop, in domain III, or in the 3' remainder (Table 1). In fact, no synonymous differences were observed in any of the latter regions, and the level of nonsynonymous diversity was likewise low (Table 1). This pattern suggested that interallelic recombination may have homogenized the 3' portion of the gene, starting with domain II, leading to reduced diversity in comparison to the 5' portion of the gene. We tested this hypothesis by examining synonymous and nonsynonymous nucleotide diversity in each of these portions of the *pvama-1* coding region separately for each of the four collections (Table 2). In every collection except South 2007, π_N in the 3' portion was significantly lower than that in the 5' portion (Table 2). Furthermore, in both NW collections, π_S in the 3' portion was significantly lower than that in the 5' portion; and the overall mean values of π_S and π_N in the 3' portion for all four collections were both significantly less than the corresponding values for the 5' portion (Table 2). These results suggest that within a given *pvama-1* haplotype, the 5' and 3' portions (and thus domain I and domain II) may have different evolutionary histories a result of recombination.

Domain I and II Haplotypes

In order to assess the potential impact of recombination on the antigenic properties of domains I and II, we examined the patterns of amino acid sequence difference in domains I and II. When we plotted the proportion of amino acid difference (p) in domain II versus that in domain I for all pairwise comparisons among the 231 sequences, there was a weak but significant positive correlation between p in domain II and that in domain I ($r = 0.298$; $P < 0.001$; randomization test; Figure 3). However, there were certain comparisons with very high p in domain I but with few or no differences in domain II (Figure 3). Conversely, there were certain comparisons with very high p in domain II but with few or no differences in domain I (Figure 3). There were 823 comparisons with no amino acid differences in domain II yet at least one amino acid difference in domain I. Of these, 69 (8.4%) comparisons showed p in domain I greater than 0.04; and 514 (62.5%) showed p in domain I greater than 0.02. Likewise, there were 468 comparisons with no amino acid differences in domain I yet at least one amino acid difference in domain II. Of these, 4 (0.9%) comparisons showed p in domain II greater than 0.04; and 123 (26.3%) showed p in domain II greater than 0.02.

Table 3 summarizes the numbers of distinct amino acid haplotypes observed in domain I and in domain II in each of the collections. The gene diversity ("heterozygosity") of haplotypes in both domains was very high (over 85%) in all collections except South 2007 (Table 3). Although more sequences (73) were obtained in the South 2007 collection than in any of the other collections, in that collection only 4 haplotypes were observed in domain I and only 4 haplotypes in domain II (Table 3). Moreover, in the South 2007 collection, each domain I haplotype was observed in combination with just one domain II haplotype; as a result there

were just 4 haplotypes for domains I and II combined (Table 3). By contrast, in all of the other collections, there were more domain I haplotypes than domain II haplotypes, resulting in a slightly higher gene diversity of domain I haplotypes than of domain II haplotypes (Table 3). Moreover, in these collections, certain domain I haplotypes were seen in combination with more than one domain II haplotype; as a result, the total number of haplotypes in both regions combined substantially exceeded that of domain I haplotypes alone (Table 3). Thus, recombination events between domain I and domain II played a role in increasing overall diversity across the two domains.

Population Differences

Tajima's D computed for the 1996 NW sample was 0.106, whereas that for the 1996 SE sample was 1.069. D for the 2007 East sample was 0.481; and that for the 2007 South sample was 0.462. None of these three values was significantly different from zero. Likewise, no significant D values were obtained when this statistic was calculated separately for the different domains (not shown). We estimated pairwise F_{ST} between each of the three regions in the 2007 collections (Table 4). The results showed substantial subdivision, as indicated by highly significant F_{ST} values (Table 4). The values of F_{ST} were similar whether based on the 5' portion of the gene (the start of the coding region through domain I) alone, or the 3' portion of the gene (domain II through the end of the coding region; Table 4).

Discussion

In the most extensive collection to date of complete sequences of the *pvama-1* gene of *Plasmodium vivax*, overall nucleotide diversity at both synonymous (π_S) and nonsynonymous (π_N) sites was less than 1% (Table 1). The overall nonsynonymous nucleotide diversity (0.0071) was over half that observed on the *ama-1* gene of *Plasmodium falciparum* (Verra and Hughes 2000; Escalante et al. 2001). In our data for *P. vivax*, the highest π_N values were seen in domain I (0.0127) and in the non-loop portion of domain II (0.010%); but these values were much lower than the over 3 % value of π_N observed in *P. falciparum* (Verra and Hughes 2000; Escalante et al. 2001).

Given the available data, it does not seem possible to decide definitively whether or not polymorphism at the *pvama-1* locus is selectively maintained. Indeed, obtaining an unambiguous signal regarding natural selection is difficult when the level of nucleotide diversity is as low as in the case of *pvama-1*. Gunasakera et al. (2007) found that π_N was significantly greater than π_S in domain II of *pvama-1* in a sample of 23 sequences from Sri Lanka. In our tenfold larger sample, π_N was slightly greater than π_S in domain II; but the difference was not significant (Table 1). Likewise, Tajima's D did not show a significantly positive value, which would be suggestive of balancing selection, in any of our collections. Thus, we found no clear indication that the polymorphism at the *pvama-1* locus is maintained by balancing selection, at least in Thailand. It is possible of course that the selective environment is different in Sri Lanka. In any event, our results agree with those of Gunasakera et al. (2006) in revealing a substantially lower level of nonsynonymous polymorphism at this locus in *P. vivax* than in *P. falciparum*, implying that natural selection acts differently at this locus in the two species.

Confirming the important result of Gunasakera et al. (2007), we found a high degree of amino acid sequence conservation in the domain II loop. Unlike Gunasakera et al. (2007), we did find one polymorphic amino acid position in the domain II loop. However, the polymorphic variant (I313R) at this site was evidently quite rare, being found in just one of 231 sequences. The fact that the domain II loop shows a much lower level of nonsynonymous nucleotide diversity than either the remainder of domain II or the remainder of the gene supports the hypothesis that the amino acid sequence of this loop is subject to strong purifying selection. The existence of one

or a few very rare variants would not necessarily argue against Gunasakera et al.'s (2007) suggestion that the domain II loop of AMA-1 might be used as a component of a vaccine for *P. vivax*.

Our results suggested that recombination, occurring between domains I and II, has been a recurrent feature of the evolution of *pvama-1*. As a result of this recombination, and subsequent events of either genetic drift or directional selection, the 3' portion of the gene (from domain II through the end of the coding region) has been homogenized relative to the 5' portion of the gene (from the start of the coding region through domain I). Thus, both synonymous and nonsynonymous nucleotide diversity in the 3' portion of the gene were found to be significantly lower in the 3' portion of the gene than in the 5' portion of the gene. An additional effect of this recombination has to create new combinations of domain I and domain II amino acid sequence motifs. However, the number of such combinations was still limited in comparison to the theoretical maximum. Overall, we observed 90 different amino acid sequence haplotypes in domains I and II combined; this amounts to only 4.1% of the possible combinations of the 56 domain I haplotypes and the 39 domain II haplotypes (Table 3).

Pairwise F_{ST} estimates among geographic regions (NW, East, and South) provided significant evidence of population subdivision. This degree of population subdivision is in marked contrast to the Brazilian Amazon, where there was no population subdivision at the *pvama-1* locus between populations as geographically distant as out Thai populations or even more so (Grynberg et al. 2008). These results thus support the hypothesis that *P. vivax* has infected human populations of Southeast Asia for a long time, in contrast to its post-Columbian introduction and recent spread in Brazil (Grynberg et al. 2008). The long-term history of this parasite in human populations in Southeast Asia has evidently allowed for genetic differentiation of local populations, contrary to what is seen in South America.

Using amino acid haplotypes in domains I and II as markers, the most diverse of the populations sampled in the present study was the NW (Table 3). The East was somewhat less diverse than the NW, and the South substantially less diverse than either NW or East (Table 3). The higher domain I and II haplotype diversity in the NW than in the East is consistent with the more recent decline in *P. vivax* infection in the NW than in the East, resulting in higher mean levels of human infection in the NW than in the East in both 1996 and 2007 (Figure 1). The fact that haplotype diversities in the NW did not differ substantially between 1996 and 2007 may seem surprising given the sharp decline in infections in the NW over that period (Figure 1). However, the absence of an observable effect of the population decline in the NW may be explained by the fact that the decline has been very rapid, and a prolonged bottleneck is required to produce a substantial reduction in gene diversity (Nei et al. 1975). *P. falciparum* infections also declined in the NW over the same time span, with little change in diversity at the *pfmsp-2* and *pfmsp-1* antigen loci, although the former showed substantial allele frequency differences (Putaporntip et al. 2008,2009).

The greatly reduced domain I and II haplotype diversity seen in the South may at least in part reflect a population history including a sharp decline in the early 1990's followed by a rebound after 2000 (Figure 2). The limited number of *pvama-1* allelic variants found in the South, in spite of the fact that more sequences (73) were obtained there in 2007 than in either of the other regions (Table 3), suggests a population expanding rapidly from a small number of founders. It would be of interest to test whether other polymorphic markers besides *pvama-1* likewise show indications of rapid expansion from a small number of founders, in order to obtain further evidence regarding the factors responsible for the unusual pattern seen at *pvama-1* in this region.

Acknowledgments

We are grateful to all patients who donated their blood samples for this study and to Thongchai Hongsriruang, Sunee Seethamchai, Pannadhat Areekul and the staff of the Bureau of Vector Borne Disease, Department of Disease Control, Ministry of Public Health, Thailand, for assistance in field work. This research was supported by grants from the National Research Council of Thailand and the Thai Government Research Budget to S.J and C.P.; The Thailand Research Fund (RMU5080002) to C.P.; grants from the National Institutes of Health GM43940 to A.L.H and D43TW006571 to C.P. and L.C.

References

- Baird JK. Neglect of *Plasmodium vivax* malaria. Trends Parasitol 2007;23:533–539.
- Escalante A.A, Grebert HM, Chaiyaroj SC, Magris M, Biswas S, Nahlen BL, Lal AA. Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. X. Asembo Bay Cohort Project. Mol. Biochem. Parasitol 2001;113:279–287. [PubMed: 11295182]
- Excoffier L, Laval G, Schneider S. Arlequin version 3.0: an integrated software package for population genetics data analysis. Evol. Bioinform. Online 2005;1:47–50.
- 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]
- 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]
- Grynberg P, Fontes CJ, Hughes AL, Braga ÉM. Polymorphism at the apical membrane antigen I locus reflects world population history of *Plasmodium vivax*. BMC Evol. Biol 2008;8:123. [PubMed: 18445274]2008
- 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]
- Li W-H. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol 1993;36:96–99. [PubMed: 8433381]
- Nei, M. Molecular Evolutionary Genetics. Columbia University Press; New York: 1987.
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol 1986;3:418–426. [PubMed: 3444411]
- Nei, M.; Kumar, S. Molecular evolution and phylogenetics. Oxford University Press; New York: 2000.
- Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations. Genet. Res 1975;35:65–83. [PubMed: 7450499]
- Pizarro JC, Vulliez-Le Normand B, Chesne-Seck ML, Collins CR, Withers-Martinez C, Hackett F, Blackman MJ, Faber BW, Remarque EJ, Kocken CH, Thomas AW, Bentley GA. Crystal structure of the malaria vaccine candidate apical membrane antigen 1. Science 2005;308:408–411. [PubMed: 15731407]
- 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, Hughes AL. Differential selective pressures on the merozoite surface protein 2 locus of *Plasmodium falciparum* in a low endemic area. Gene 2008;427:51–57. [PubMed: 18840512]
- Putaporntip C, Jongwutiwes S, Hughes AL. Natural selection maintains a stable polymorphism at the circumsporozoite protein locus of *Plasmodium falciparum* in a low endemic area. Infect. Genetic. Evol. 2009(in press)
- Rajesh V, Elamaram M, Vidya S, Gowrishankar M, Kochar D, Das A. *Plasmodium vivax*: Genetic diversity of the apical membrane antigen-1 (AMA-1) in isolates from India. Exp. Parasitol 2007;116:252–256. [PubMed: 17336299]

- Tanabe K, Sakihama N, Kaneko A. Stable SNPs in malaria antigen genes in isolated populations. *Science* 2004;303:493. [PubMed: 14739451]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Diggins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;123:585–595. [PubMed: 2513255]
- Verra F, Hughes AL. Evidence for ancient balanced polymorphism at the apical membrane antifen-1 (AMA-1) locus of *Plasmodium falciparum*. *Mol. Biochem. Parasitol* 2000;105:149–153. [PubMed: 10613707]
- Webster D, Hill AV. Progress with new malaria vaccines. *Bull. World Health Org* 2003;81:902–909. [PubMed: 14997243]
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol* 2000;17:32–43. [PubMed: 10666704]

Figure 1.
Map of Thailand, showing the three collection areas (shaded).

Figure 2.
Reported numbers of cases of *P. vivax* malaria in the three collection areas from 1979-2008
(source: Division of Vector-Borne Diseases, Ministry of Public Health, Thailand).

Figure 3. Plot of the proportion of amino acid difference (p) in domain II vs. that in domain I for all pairwise comparisons among 231 *pvama-1* sequences ($r = 0.298$; $P < 0.001$; randomization test).

Table 1

Synonymous (π_S) and nonsynonymous (π_N) nucleotide diversity ($\pm$ S.E.) in *pvama-1* sequences (N = 231) from Thailand.

Region	No. codons	π_S	π_N
5' remainder	42	0.0046 ± 0.0048	0.0050 ± 0.0035
Domain I	206	0.0193 ± 0.0071	0.0127 ± 0.0034
Domain II			
Loop	42	0.0000 ± 0.0000^a	0.0001 ± 0.0001^b
Remainder	95	0.0100 ± 0.0061	0.0105 ± 0.0042
All	137	0.0071 ± 0.0044	0.0072 ± 0.0028
Domain III	102	0.0000 ± 0.0000^a	0.0016 ± 0.0010^b
3' remainder	75	0.0000 ± 0.0000^a	0.0000 ± 0.0000^b
All	562	0.0091 ± 0.0027	0.0071 ± 0.0015

Z-tests of the hypothesis that π_S or π_N equals the corresponding value in Domain I:

^a $P < 0.01$

^b $p < 0.001$.

Table 2

Synonymous (π_S) and nonsynonymous (π_N) nucleotide diversity ($\pm$ S.E.) in 5' and 3' portions of *pvama-1* from different collections.

Collection	5' portion		3' portion	
	π_S	π_N	π_S	π_N
NW 1996	0.0148 $\pm$ 0.0057	0.0121 $\pm$ 0.0033	0.0025 $\pm$ 0.0018 ^a	0.0014 $\pm$ 0.0014 ^a
NW 2007	0.0156 $\pm$ 0.0057	0.0124 $\pm$ 0.0033	0.0036 $\pm$ 0.0021 ^a	0.0042 $\pm$ 0.0016 ^a
East 2007	0.0193 $\pm$ 0.0066	0.0120 $\pm$ 0.0032	0.0058 $\pm$ 0.0031	0.0035 $\pm$ 0.0013 ^a
South 2007	0.0080 $\pm$ 0.0041	0.0057 $\pm$ 0.0021	0.0000 $\pm$ 0.0000	0.0018 $\pm$ 0.0009
Mean	0.0148 $\pm$ 0.0057	0.0105 $\pm$ 0.0016	0.0030 $\pm$ 0.0012 ^b	0.0034 $\pm$ 0.0006 ^b

^aZ-tests of the hypothesis that π_S or π_N in the 3' portion equals the corresponding value in the 5' region: $P < 0.05$.

^bPaired t-tests of the hypothesis than mean π_S or π_N in the 3' portion equals the corresponding value in the 5' region: $p < 0.001$.

Table 3

Diversity of amino acid sequence haplotypes in Domain I and Domain II.

Collection	No. sequences	Domain I		Domain II		Domains I & II	
		No. haplotypes	Gene diversity	No. haplotypes	Gene diversity	No. haplotypes	Gene diversity
NW 1996	58	40	93.7%	29	90.7%	52	96.1%
NW 2007	44	25	96.6%	19	94.6%	31	97.7%
East 2007	56	18	86.4%	17	85.1%	24	87.8%
South 2007	73	4	43.6%	4	43.6%	4	43.6%
All	231	56	90.6%	39	89.3%	90	92.4%

Table 4Pairwise F_{ST} estimates between populations in 2007 collections.

	NW vs. East	NW vs. South	East vs. South
5' portion	0.059***	0.291***	0.344***
3' portion	0.076***	0.327***	0.315***

P < 0.001.