

Genetic Diversity and Selection at the *Plasmodium vivax* Apical Membrane Antigen-1 (*PvAMA-1*) Locus in a Sri Lankan Population

Anusha M. Gunasekera,¹§ Thilan Wickramarachchi,^{1*}† Daniel E. Neafsey,¹‡ Ishani Ganguli,[§] Lakshman Perera,[†]|| Prasad H. Premaratne,^{*} Daniel Hartl,[¶] Shiroma M. Handunnetti,[†]# Preethi V. Udagama-Randeniya,^{*} and Dyann F. Wirth[§]

*Department of Zoology, University of Colombo, Sri Lanka; †Malaria Research Unit, University of Colombo, Sri Lanka; ‡Broad Research Institute of MIT and Harvard, Cambridge, MA; §Department of Immunology and Infectious Diseases, Harvard School of Public Health, Harvard University; ||Health Center, University of Ruhuna, Meddawaththa, Matara, Sri Lanka; ¶Department of Organismic and Evolutionary Biology, Harvard University; #Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Sri Lanka

Plasmodium vivax apical membrane antigen 1 (*PvAMA-1*) is an important malaria vaccine candidate. We present the first comprehensive analysis of nucleotide diversity across the entire *PvAMA-1* gene using a single population sample from Sri Lanka. In contrast to what has been observed at the *AMA-1* locus of *Plasmodium falciparum*, the signature of diversifying selection is seen most strongly in Domain II of *PvAMA-1*, indicating that the different domains in each species may be subject to varying selective pressures and functional constraints. We also find that recombination plays an important role in generating haplotype diversity at this locus, even in a region of low endemicity such as Sri Lanka. Mapping of diversity and recombination hotspots onto a 3-dimensional structural model of the protein indicates that one surface of the molecule may be particularly likely to bear epitopes for antibody recognition. Regions of this surface that show constrained variability may prove to be promising vaccine targets.

Introduction

Each year, the malarial parasite *Plasmodium vivax* infects 70–80 million people worldwide (Mendis et al. 2001; Sina 2002). In parts of Asia and South America, it is the most prevalent form of the 4 *Plasmodium* species affecting humans. Unlike *Plasmodium falciparum*, infections caused by *P. vivax* are rarely lethal. However, *P. vivax* has a significant impact on the productivity of local populations as the course of infection is usually protracted in nature and development of acquired immunity in holoendemic areas takes several years (Mendis et al. 2001; Sina 2002). The recent emergence of drug resistant strains (Wilairatana et al. 1999; Imwong et al. 2001) further complicates the burden of *P. vivax* malaria and calls for the intensification of research aimed at alternate control methods, such as vaccine development.

Population diversity studies that analyze individual parasite antigens provide useful information on the impact of genetic variation on vaccine efficacy (Cui et al. 2003). Characterizing antigenic polymorphism will also determine whether particular regions of a protein are subject to balancing selection, which, in turn, may uncover targets of naturally acquired immunity in *P. vivax*, as has been demonstrated in *P. falciparum* (Conway et al. 2000). Traditional population genetic analyses predict that patterns of nucleotide polymorphisms within genes (or domains) under selection should statistically depart from those predicted by neutral evolution models (Escalante et al. 2004). Newer approaches that detect variation in both the selection parameter and the recombination rate along a sequence (Wilson and McVean 2006), together with the recent determination of the 3-dimensional crystal structure of various *P. vivax* antigens

(Pizarro et al. 2005; Singh et al. 2006) may help pinpoint exact targets of protective immunity to *P. vivax*, and hence, will be of tremendous relevance to vaccine research.

Effective vaccine targets should show limited sequence diversity because antigenic variation presents a major hurdle in successful vaccine design. Several malarial antigens, including merozoite surface proteins 1 and 2, display high levels of polymorphism, whereas functional constraints limit the degree of variation in others, such as apical membrane antigen-1 (AMA-1) (reviewed in Mahajan et al. 2005). Indeed, AMA-1 is highly immunogenic, protecting against *Plasmodium chabaudi* challenge infection in mice (Crewther et al. 1996) and displaying significant antibody and T cell responses in endemic human populations (Thomas et al. 1994; Lal et al. 1996; Rodrigues et al. 2005; Wickramarachchi et al. 2006). In fact, AMA-1 is currently in clinical trials as a vaccine against *P. falciparum* (Malkin et al. 2005; Saul et al. 2005).

AMA-1 is a microneme-derived surface protein (Waters et al. 1990) that is transported to the parasite exterior just prior to red blood cell invasion. This essential gene (Triglia et al. 2000) is highly conserved among *Plasmodium* species (Cheng and Saul 1994). The disulfide bond pattern between 16 invariant cysteine residues divides the ectoplasmic region of AMA-1 into 3 distinct domains (Hodder et al. 1996; Nair et al. 2002). The recent crystal structure of *P. vivax* apical membrane antigen 1 (*PvAMA-1*) reveals that domains I and II belong to the PAN motif, implicated in receptor binding (Pizarro et al. 2005). Indeed, cell lines expressing recombinant domains of the antigen specifically bound red blood cells (rbc's) (Fraser et al. 2001; Kato et al. 2005). These results, together with the mapping of an epitope of an invasion-inhibitory monoclonal antibody to a putative receptor-binding site within the PAN domain (Pizarro et al. 2005), suggest a critical role for Domain II in rbc invasion.

The immunologic relevance of each of the 3 domains of *P. falciparum* apical membrane antigen 1 (*PfAMA-1*) was specifically evaluated by diversity studies on single endemic

¹ Equal contribution to this work.

Key words: selection, *Plasmodium vivax*, genetic diversity, recombination.

E-mail: dfwirth@hsph.harvard.edu.

Mol. Biol. Evol. 24(4):939–947. 2007

doi:10.1093/molbev/msm013

Advance Access publication January 22, 2007

populations; these analyses found significant evidence for positive diversifying selection on polymorphisms within domain I and III of the molecule, where the rate of nonsynonymous substitution (dN) was found to be much higher than that of synonymous substitution (dS) (Escalante et al. 2001; Polley and Conway 2001; Polley et al. 2003). In comparison with *P. falciparum*, molecular studies on the genetic diversity of *PvAMA-1* are less complete. All analyses to date have focused solely on domain I of the antigen (Figtree et al. 2000; Rodrigues et al. 2005) and have not recovered evidence for diversifying selection in this region (Figtree et al. 2000).

The current investigation undertakes a comprehensive analysis of nucleotide diversity at the entire *PvAMA-1* locus from a single population in Sri Lanka. Using this population sample, we examined variation in the intensity of selection and recombination across the *PvAMA-1* gene. In contrast to studies on *PfAMA-1*, the signature of diversifying selection is seen most strongly in Domain II and to a lesser extent in Domain I of *PvAMA-1*. We also find that recombination plays an important role in generating haplotype diversity across the entire sequence, even in this low endemic setting. Finally, we characterize the specific distribution of selection in the 3-dimensional structure of the protein and evaluate the prospects of *PvAMA-1* as a vaccine target in light of this information.

Materials and Methods

Sample Selection and DNA Extraction

Following review/ethical approval by the Institutional Review Boards of the Faculty of Medicine, University of Colombo, Sri Lanka, and the Harvard School of Public Health, Boston, MA, blood samples were collected with informed consent from *P. vivax*-infected patients (age >15 years) from Kataragama (endemic setting; 6°25'N, 81°20'E; $n = 35$) and from Colombo (nonendemic setting; 7°55'N, 79°50'E; $n = 4$). Patients presenting at hospitals in the latter site were returning from visits to regions with significant *P. vivax* transmission in Sri Lanka. Only blood samples positive for *P. vivax* infection as determined by Giemsa-stained thick smears were considered. Percent parasitemia ranged from 0.0001 to 0.025. The proportion of male to female patients was 34:5.

Genomic parasite DNA was extracted from 5 ml of venous blood as described in (Thomas et al. 2002). Briefly, plasma and white blood cells were removed from whole blood samples following centrifugation and filtration through a CF11 column. Red blood cells were then lysed with 0.015% saponin in NET buffer (150 mM sodium chloride, 10 mM ethylenediaminetetraacetic acid (EDTA), 50 mM Tris, pH 7.5). Parasite pellets were treated with 1% N-lauroylsarcosine (Sarkosyl), RNase A (100 µg/ml) and proteinase K (200 g/ml). Parasite DNA was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 0.3 M sodium acetate and absolute ethanol. DNA pellets were air dried, resuspended in Tris buffer (10 mM Tris and 1 mM EDTA), and stored at -20 °C until further use.

Parasite Genotyping

Parasite stages that infect the human host are haploid and thus, each contains only 1 allele of a given locus from

the *vivax* genome. A patient blood sample may, however, be infected with multiple strains of *P. vivax*. In order to exclude multiclonal infections from the present study, each of the 39 DNA samples were genotyped at the polymorphic MSP-3 α locus using a combined nested polymerase chain reaction (PCR) and restriction fragment length polymorphism strategy described by (Bruce et al. 1999). Thirty-four single clone infections were chosen for further study.

PvAMA1 PCR Amplification and Sequencing

PvAMA-1 is encoded by a single locus (~1.7 kb). The entire *PvAMA-1* gene, including a portion of the upstream untranslated region, the prodomain, ectodomain, and C-terminal tail, was PCR amplified and sequenced. Four overlapping templates (see supplementary material I, Supplementary Material online) were amplified separately using the following primer pairs: Uf, 5'-CACTGGGC-GCTTAAAAAGATTATGTAG - 3'; Ur, 5'-CTTTTTTT-GCTCCCTTTTTTGCTCC-3'; 1f, 5'-CCTACCAGCG-GTTACTTCCA-3'; 1r, 5'-CCCATATTTTCTGCGCT-GATAAATAC-3'; 2f, 5'-GGCAAGGTATAAAGACAA-TGTAGAG-3'; 2r, 5'-CAGATTCATGTTCTCGATTG-TTTC-3'; 3f, 5'-GAGGATTTAACTGGGCAAATTTTCG-3'; 3r, 5'-TCAGTAGTACGGCTTCTCCAT-3'.

PCR conditions were optimal when DNA templates were incubated with 0.2 mM dNTP, 1X PCR buffer (Promega), 2 mM MgCl₂, 0.5 Units of Taq polymerase and 0.2 µM of each primer pair, and taken through a denaturation step at 94 °C; 39 cycles of amplification at 94 °C for 1 min, 57 °C for 1 min, and 72 °C for 1 min; and a final extension step at 72 °C for 10 min. The resulting PCR products (~700–800 bp) were resolved on agarose gels and purified by the PCR Purification Column kit (Qiagen, Crawley, United Kingdom), by the Gel Extraction spin kit (Qiagen, Crawley, United Kingdom), or by ExoSAP-IT treatment (USB, Ohio) as per the manufacturer's protocol for preparation of templates for direct sequencing. Three separate PCR samples for each region of the gene being amplified were sequenced in both the forward and reverse directions using Big Dye termination chemistry (ABI prism automated sequencer). Hence, all single nucleotide polymorphisms uncovered in the current study were confirmed by achieving 6-fold sequence coverage at each of the samples analyzed.

Sequence Alignment and Analysis

A single continuous sequence of 1,504 nucleotides at 6-fold coverage was derived from 23 patient isolates and deposited in GenBank (accession number EF218679–EF218701). Raw sequence data were checked and aligned using Seqman II and the Clustal algorithm in MegAlign (DNASar, Madison, WI). DNASar was also used to translate the sequence and identify synonymous and nonsynonymous amino acid substitutions between all 23 samples. Statistical analyses were carried out using DnaSP version 4.0 (Rozas J and Rozas R 1999), and phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.1 (Kumar et al. 2004; <http://www.megasoftware.net/>).

Tests for DNA Diversity

Several measures of genetic diversity were employed in this study. DnaSP was used to calculate 1) π : an estimate of the average number of substitutions between any 2 sequences, assuming the sample is random (all sites—synonymous and replacement substitutions—were considered in this calculation); 2) S : the number of polymorphic sites in the sample, which is dependant on the sample size and length of sequence; and 3) H : the number of haplotypes. Because transitions are generally more frequent than transversions, we also used the Tamura 3 parameter model to estimate d (MEGA 3.1). This statistic is comparable to π , but corrects for biases in GC content and the transition/transversion ratio. We obtained similar values for both d and π in all 3 ectodomains (data not shown).

Tests of Neutrality

The ratio of nonsynonymous to synonymous substitutions (dN/dS) is widely used as an indicator of the action of natural selection in gene sequences. An excess of nonsynonymous relative to synonymous mutations is a clear signal of positive selection, whereas a lack of nonsynonymous relative to synonymous polymorphisms suggests negative or purifying selection imposed by functional constraint. These rates of substitution within species were estimated using the method of Nei and Gojobori (Nei and Gojobori 1986) with the Jukes and Cantor correction (Jukes and Cantor 1969) as implemented in the MEGA 3.1 software. Standard error was determined by 1,000 bootstrap replications and the 2 rates were compared with the Z-test of selection (MEGA 3.1).

The McDonald–Kreitman test was applied as a second test of natural selection. Here, the ratio of nonsynonymous to synonymous changes within and between species is compared. The *Plasmodium cynomolgi* (Dutta et al. 1995) AMA1 sequence was used for interspecific comparison with *P. vivax* (88% sequence identity). Under neutrality these ratios will be similar, whereas an excess of intraspecific nonsynonymous polymorphisms is suggestive of diversifying positive selection or relaxed purifying selection. Fisher's exact test was applied to the data to test for significant nonrandomness ($P < 0.05$), and the skew from randomness was calculated as the Neutrality Index in DnaSP.

Analysis of Recombination and Linkage Disequilibrium

Analysis of recombination in *PvAMA-1* was performed on the alignment of the 23 sequences by DnaSP in order to calculate the minimum number of recombination events that have occurred along the sequence and to estimate the recombination parameter, C . This statistic incorporates the effective population size and probability of recombination between adjacent nucleotides per generation. The D' and R^2 indices of linkage disequilibrium (LD) were also estimated across all polymorphic sites (DnaSP), and the relationship between LD and distance between nucleotide sites of each pairwise comparison was plotted (see supplementary material II, Supplementary Material online).

D' has a potential range from -1 to 1 and the magnitude of disequilibrium is indicated by its departure from zero in either direction. The statistical significance of each

Table 1
Prior Distributions

	Objective	Subjective
μ	Improper inverse (0.1) ^a	Exponential mean 0.015
κ	Improper inverse (3.0) ^a	Exponential mean 2.86
ϕ	Improper inverse (0.1) ^a	Exponential mean 0.1
ω	Inverse (0.01,100) ^b	Exponential mean 1.0
ρ	Inverse (0.01,100) ^b	Exponential mean 0.1

^a starting value.

^b range (min, max).

pairwise test of LD on these haploid data was evaluated by χ^2 test. All pairwise analyses are not fully independent because of LD itself; hence, the proportion of significant ($P < 0.05$) pairwise tests was compared for pairs of sites separated by different molecular map distances (< 300 bp or > 300 bp). Comparison of these ratios over the 2 separate distance ranges was carried out by χ^2 analysis.

Characterizing Selection in the Presence of Recombination

Variation in selection (ω) and recombination rate (ρ) across the *PvAMA-1* locus was analyzed using the omega-Map software package (Wilson and McVean 2006). This method employs a Bayesian approach to parameter estimation that is independent of phylogeny, and therefore, less likely to falsely identify sites subject to diversifying selection in sequences that display clear evidence of recombination. Positional variation in ω and ρ was investigated using a sliding window of 30 codons. Analyses were conducted using both an objective and a subjective set of priors (table 1) to test the sensitivity of the outcome to prior distribution. Means for subjective priors for μ (mutation rate), κ (transition/transversion ratio), and ρ were chosen using the posterior distributions generated with the objective prior set. A mean value of 1 was used for the ω subjective prior based on a null hypothesis of selective neutrality. For each set of priors, 2 independent MCMC chains were run for 500,000 iterations, with a 20,000 iteration burn-in. Paired chains were found to have converged in less than 250,000 iterations and were therefore merged to infer posterior distributions over ω and ρ . The posterior probability of selection at each codon location was mapped onto a 3-dimensional structure file of *PvAMA-1* (PDB file: 1W8K; Pizarro et al. 2005) to visualize the spatial distribution of selected sites. Protein structure was visualized using Protein Explorer (www.proteinexplorer.org; Martz 2002).

Results

Sequence Polymorphism across *PvAMA-1*

We find evidence for high allelic diversity among *P. vivax* parasites in Sri Lanka, following sequence analysis of the entire coding region of *PvAMA-1*. Fifteen haplotypes were identified among the 23 Sri Lankan isolates analyzed. There are a total of 34 polymorphic sites and 35 nucleotide polymorphisms. Only one of these sites was trimorphic, whereas the rest were dimorphic. Twenty-seven of the 35 nucleotide polymorphisms uncovered here caused amino acid replacements (nonsynonymous changes). All but one (site 188) of the amino acid substitutions in domain

I were previously reported from worldwide isolates (Figtree et al. 2000; Chesne-Seck et al. 2005). Two thirds of the sites previously defined in domain I of 219 isolates (Figtree et al. 2000) were also found in the present study, despite the fact that sample size was 10-fold lower here, suggesting that single nucleotide polymorphism (SNP) hotspots are conserved across isolates from different geographic locations.

The recent modeling of the *PfAMA-1* sequence onto the crystal structure of *PvAMA1* (Chesne-Seck et al. 2005) made it possible for us to compare the distribution of SNPs between the 2 species (59% sequence identity). Nonsynonymous variations uncovered in domains II and III were specifically compared, because other studies had previously focused on domain I (Chesne-Seck et al. 2005). We found that 5 (residue 253, 277, 352, 380, and 384) of the 9 polymorphic sites in domain II and 1 (residue 438) of the 2 polymorphic sites in domain III were also polymorphic in *PfAMA-1*. The hitherto uncovered substitution at position 188 (Domain I) of *PvAMA-1* was also polymorphic at its comparable residue in *PfAMA-1*, position 243 (Bai et al. 2005; Chesne-Seck et al. 2005); interestingly residue 243 was 1 of 5 sites categorized as “highly polymorphic” in *PfAMA-1* (Bai et al. 2005).

Sequence Diversity among Ectodomains of *PvAMA-1*

When diversity across each of the 3 ectodomains of *PvAMA-1* was compared, we found SNPs to be most concentrated in Domains I and II of the antigen (table 2). It follows that pairwise diversity (π) was highest in these 2 regions (0.0092 and 0.0097, respectively). Previous estimates of diversity in Domain I of *PvAMA-1* (0.0174) were slightly larger and may reflect the larger number of isolates and hence SNPs sampled in Figtree et al. (2000). Diversity across the most polymorphic domain of the Duffy-binding protein in *P. vivax* (Cole-Tobian and King 2003) was much higher ($\pi = 0.0184$; $n = 9$), as were diversity estimates from a 249 bp region of *PvMSP1* (0.0451 $n = 175$) (Figtree et al. 2000), confirming earlier observations that *AMA-1* is relatively conserved in comparison to many blood stage antigens in *Plasmodium* (Cheng and Saul 1994; Escalante et al. 1998; Mahajan et al. 2005).

To determine whether natural selection contributes to the diversity in Domain I and II, the ratio of nonsynonymous to synonymous substitutions (dN/dS) was evaluated within species (table 2). The dN/dS ratio was significantly greater than 1 in Domain II of *PvAMA-1* (Z-test: $P < 0.03$), suggesting that this part of the protein is most subject to directional or diversifying selection. As found in a previous study (Figtree et al. 2000), the dN/dS ratio in Domain I was not significantly different from 1; this finding may reflect the fact that both purifying and positive selection are acting on this region.

McDonald–Kreitman Test

As a further test of selection, the McDonald–Kreitman test for neutrality was applied to the entire *PvAMA-1* sequence, as well as to each ectodomain (table 3). The sequence as a whole shows significant departure from neutrality on the basis of excess intraspecific nonsynonymous

polymorphisms relative to nonsynonymous divergence from *P. cynomolgi*. All 3 of the domains show the same trend, however, this excess is statistically significant in Domains I and II only. Within-population analyses of alleles from Thai (Polley et al. 2003) and Nigerian (Polley and Conway 2001) *P. falciparum* isolates yield significant evidence for balancing selection on polymorphisms within Domains I and III of *PfAMA1*. In contrast, our results suggest that polymorphisms found in Domain II (and perhaps to a lesser extent, those in Domain I) of *PvAMA-1* are maintained by positive selection, presumably due to host immune pressure.

Alternatively, the observed excess of nonsynonymous polymorphisms in Domains I and II may indicate a relaxation of purifying selection in *P. vivax*, rather than the action of diversifying selection. Indeed, Leclerc et al. (2004) report evidence for a recent population bottleneck in *P. vivax*, a demographic event that could result in relaxed purifying selection. However, this study, which is based on the paucity of microsatellite polymorphisms (dinucleotide repeats) in the *vivax* genome, has since been challenged by a more recent paper (Imwong et al. 2006) that reports a very high level of microsatellite diversity when a longer set array length is chosen. The lack of robust evidence for a recent bottleneck in *P. vivax*, in addition to observations of balancing selection at the *AMA-1* locus in *P. falciparum* (Polley and Conway 2001; Polley et al. 2003), suggests that the significant McDonald–Kreitman result reported in the current study may be most parsimoniously explained by balancing selection, rather than relaxed purifying selection.

Recombination and Linkage Disequilibrium

We find evidence for a high rate of recombination at the *PvAMA-1* locus in our Sri Lankan study population. The 4-gamete test (Hudson and Kaplan 1985) suggests that at least 9 recombination events have occurred among the 23 sequences. When meiotic recombination drives allelic diversity, LD between polymorphic sites (measured by both D' and R^2) is expected to decline with increasing distance along the chromosome. Both R^2 and D' were calculated for all polymorphic sites across the gene and statistical significance values were calculated for all pairwise combinations. For both R^2 and D' , a weak negative correlation was observed with nucleotide distance between pairs of sites (see supplementary fig. II, Supplementary Material online). To confirm this relationship, we compared the number of significant comparisons among sites <300 bp apart and >300 bp apart. Both the average value of R^2 and $|D'|$ were significantly larger among sites <300 bp apart (see table 4) and the percentage of significant pairwise comparison was also significantly higher ($P < 0.0001$) in this group. Together, these data suggest that recombination plays a role in generating haplotype diversity across *PvAMA-1*. Similar results were observed in *PfAMA-1* from Nigerian and Thai isolates (Polley et al. 2003; Polley and Conway 2001).

Estimating Diversifying Selection in the Presence of Recombination

The abundant evidence of recombination within *PvAMA-1* suggests that different regions of the locus

Table 2
Nucleotide Diversity at Different Regions of the PvAMA-1 Gene

Region	Residues	π (S.D) ^a	S^b	H^c	dN (s.e) ^d	dS (s.e) ^e	dN/dS
Entire gene ^f	1–501	0.0067 (0.00065)	34	15	0.008 (0.002)	0.006 (0.002)	1.33
Domain I	43–248	0.0092 (0.00138)	20	11	0.008 (0.003)	0.012 (0.006)	0.67
Domain II	249–385	0.0097 (0.00075)	12	13	0.012 (0.004)	0.002 (0.002)	6*
Domain III	386–487	0.0016 (0.00045)	2	3	0.0019 (0.0013)	0	—

NOTE.—“**” indicates that dN and dS were significantly different from one another, based on the Z-test of selection; $P < 0.05$.

^a Nucleotide diversity (all sites—synonymous and replacement mutations—are considered in the calculation) and its standard deviation in parentheses.

^b Number of polymorphic sites.

^c Number of haplotypes.

^d Number of nonsynonymous mutations per nonsynonymous site using the Nei–Gojobori method with the Jukes–Cantor correction and standard error in parentheses.

^e Number of synonymous mutations per synonymous site using the Nei–Gojobori method with the Jukes–Cantor correction and standard error in parentheses.

^f From a total of 562 residues 501 were analyzed above.

may have experienced different evolutionary histories, and therefore the inference of selection at this locus using maximum likelihood phylogenetic approaches (e.g., Nielsen and Yang 1998) may be inappropriate. Analysis of *PvAMA-1* using omegaMap (Wilson and McVean 2006) revealed patterns of selection and recombination within the locus congruent with estimates from more traditional approaches. Figure 2 shows the distribution of ω and ρ under the subjective prior set. Posterior distributions of ω and ρ across the protein were qualitatively similar regardless of the set of priors used (supplementary fig. III, Supplementary Material online), suggesting that these results are not significantly biased by the priors. There are 3 distinct peaks in ω , 2 within domain I (at approximately residue positions 110–150 and 190–240) and 1 in domain II (at approximately residue positions 350–390). Thus, the omegaMap analysis reveals a heterogeneous mixture of sites subject to positive selection or purifying selection along the length of the gene and may explain why the regional estimate of dN/dS in domain I was not significantly greater than 1 (table 2). Variation in ρ (although not as pronounced as that of ω) tends to track variation in ω , suggesting that recombination may facilitate diversification at this locus by shuffling SNPs both within and among hotspots of selection.

Three-Dimensional Distribution of Polymorphic Sites

The 3-dimensional distribution of SNPs as well as selection intensity across the *PvAMA-1* locus was examined in detail. The crystal structure of *PvAMA-1* (Pizarro; PDB file: 1W8K) was color coded according to the posterior probability of positive selection at each codon calculated using omegaMap (fig. 3), with dark blue coloring indicating strong functional constraint and lighter coloring indicating

increasing likelihood of diversifying selection. As expected, SNPs (colored dark brown) are distributed in areas where positive selection is inferred to be relatively intense. Similar to observations in *PfAMA-1* (Bai et al. 2005), the location of polymorphic sites is highly biased to one side of the molecule, suggesting that this surface may be more exposed to the exterior environment (i.e., immune pressure) and less susceptible to functional constraints.

Discussion

We report high allelic diversity within a Sri Lankan population of parasites and find evidence that this diversity is maintained by balancing selection at the *PvAMA-1* locus, presumably to evade immune recognition by the host. We tested for selection in *PvAMA-1* by traditional analyses as well as a novel approach for estimating population genetic parameters that is resilient to intralocus recombination (Wilson and McVean 2006). All methods find evidence for strong diversifying selection in Domain II of *PvAMA-1*. Diversity studies in the *PfAMA-1* locus also showed significant departure from neutrality for the molecule as a whole (Verra and Hughes 1999, 2000; Escalante et al. 2001; Polley and Conway 2001; Polley et al. 2003); however, in this case, positive selection was strongest in domain I and III of the protein (Polley and Conway 2001; Polley et al. 2003). Thus, different regions in *PvAMA-1* versus *PfAMA-1* could be under different pressures in terms of immune selection, and structural and/or functional constraints, as has been proposed for PvTRAP and PfTRAP (Putaporntip et al. 2001).

Despite differences in relative selection intensities across the 3 domains of *PvAMA-1* and *PfAMA-1*, the distribution of polymorphic sites in the 3-dimensional

Table 3
Nucleotide Diversity and Selection Pressure on Different Domains of PvAMA-1

Region	Fixed Differences between Species ^a		Polymorphic Changes within Sri Lankan Isolates		Neutrality Index	Fisher's Exact test
	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous	M–K test	(2-tailed)
Entire coding sequence	111	61	5	27	9.826	<0.000001
Domain I	45	23	4	13	6.359	0.0022
Domain II	31	13	1	12	28.615	0.00007
Domain III	18	20	0	2	—	Not significant

^a The AMA-1 sequence from *Plasmodium cynomolgi* (Dutta et al. 1995) was used as the outgroup species in the McDonald–Kreitman test.

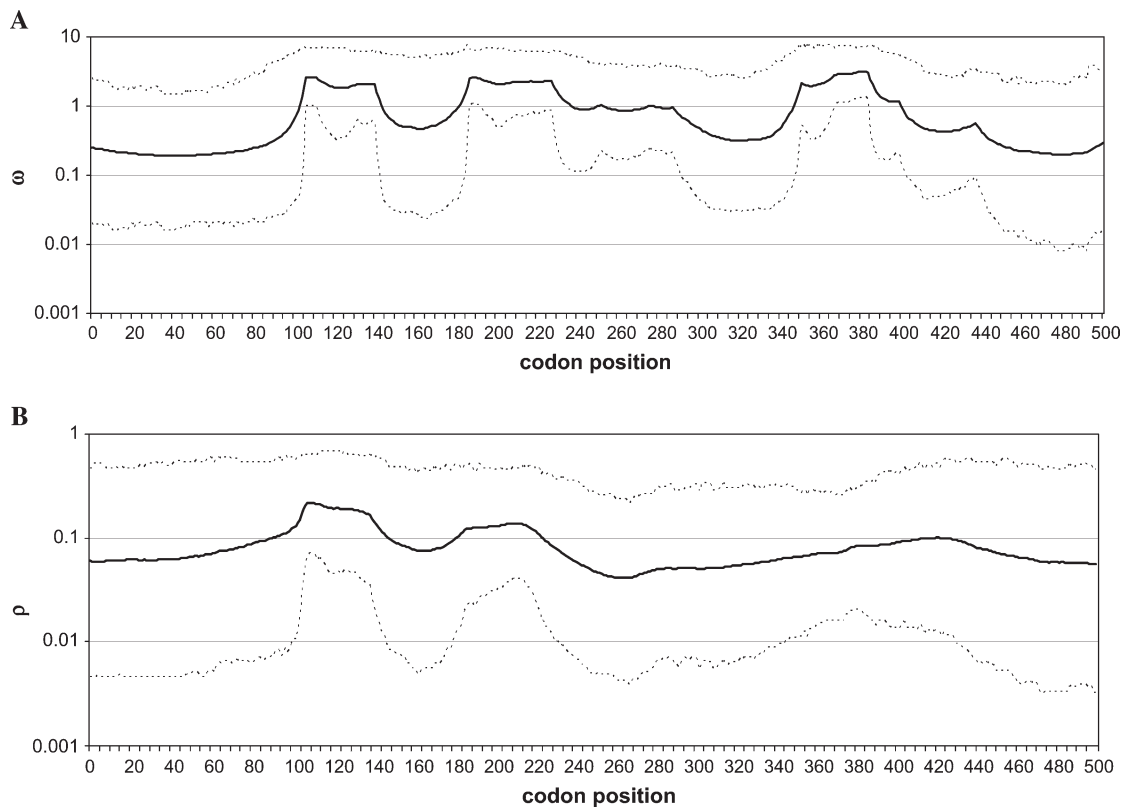


FIG. 2.—Spatial variation in selection, ω (A) and recombination ρ (B) across the *AMA-1* locus in *P. vivax*. Parameter estimates were carried out using the omegaMap software package using the “Subjective” set of prior distributions (Wilson and McVean 2006). The sitewise mean (solid line) and 95% HPD intervals (dotted lines) are shown in each case.

low transmission rates (Anderson et al. 2000). Our results may indicate that rare recombinant haplotypes generated within the Sri Lankan population are highly selectively advantageous or that new haplotypes generated by recombination in areas of higher endemicity sweep into the Sri Lankan population. If *AMA-1* is indeed subject to balancing selection, recombinant *AMA-1* haplotypes would be expected to accumulate in the worldwide population over long periods of time, and their diffusion into hypoendemic regions would bring the benefits of outcrossing to largely clonal populations. Alternatively, these observations may be related to the unique biological features of *P. vivax* infection, such as earlier gametogenesis and relapse, which increase the chance of mixed strain infections and sexual outcrossing, even in areas of low endemicity. Indeed, approximately 13% of infections in the Sri Lanka population were multiclonal in nature (data not shown).

PvAMA-1 polymorphic sites in the Sri Lankan population, as well as corresponding substitutions, were well conserved among *P. vivax* isolates worldwide. A gene genealogy using the *PvAMA-1* sequences of Sri Lanka isolates in the current study and 245 previously deposited sequences in GenBank (www.ncbi.nlm.nih.gov), revealed no significant geographic clustering of the samples from Sri Lanka, despite the barrier to gene flow imposed on this population by its island status (data not shown). Our data together with previous analyses at the *PvAMA-1* locus (Figtree et al. 2000) predict that a multiallelic vaccine based on the rational identification of *AMA-1* sequences will be

equally effective in protecting against *vivax* malaria in different endemic settings around the world.

It is predicted that those regions of *PvAMA-1* which are functionally restricted yet remain sufficiently immunogenic will provide effective targets for vaccine development. Half of the polymorphic residues uncovered in the present investigation mapped to secondary structure components of the PAN folding motif (a motif associated with adhesion functions) in Domain I and II. However, consistent with results in *P. falciparum* (Chesne-Seck et al. 2005), we failed to detect polymorphisms in a segment of Domain II (the domain II loop) that was shown to include the epitope

Table 4
Summary of LD Analysis for *PvAMA-1*

	<i>N</i> (pairs)	<i>ID'</i> ¹	<i>R</i> ²	% of Significant Pairwise Tests
<300 bp ^a	243	0.849	0.171	29
>300 bp ^b	285	0.767	0.068	11
<i>P</i> value for indices of LD ^c	—	0.0038	<0.0001	—

NOTE.—*N* (pairs), Number of pairwise comparisons between sites; *ID'*¹, Average of the absolute values for *D'*; *R*², Average of *R*²; % of Significant Pairwise Tests, The percent of significant pairwise tests by χ^2 analysis in each distance group. The probability that the portion of significant pairwise test in each distance group (<300 bp and >300 bp) was significantly different was $P < 0.0001$ (χ^2 statistic: 24.89).

^a Pairwise comparisons between sites less than 300 bp apart.

^b Pairwise comparisons between sites greater than 300 bp apart.

^c The probability that the mean values for indices of LD (*D'* and *R*²) are significantly different between distance groups (<300 bp and >300 bp) as evaluated by independent sample *t*-tests.

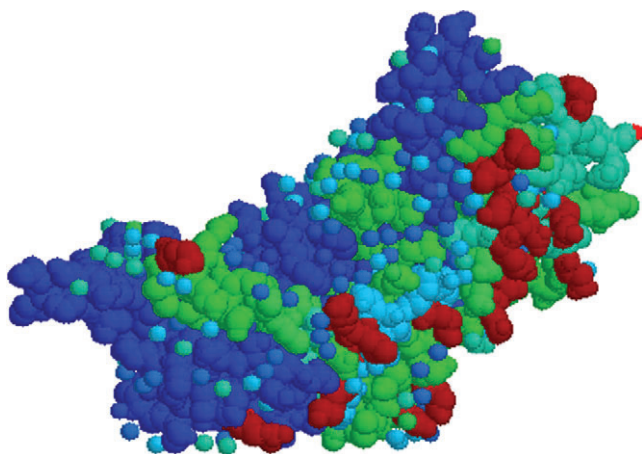


FIG. 3.—Three-dimensional structure of *PvAMA-1* color coded according to selection intensity. The molecular structure of *PvAMA-1* was color coded according to the posterior probability of positive selection at each codon, calculated using omegaMap. Dark blue coloring indicates strong functional constraint whereas lighter colors indicate increasing likelihood of diversifying selection. Positions of amino acid replacements are indicated in dark brown. This image was produced using Protein Explorer (Martz 2002; <http://proteinexplorer.org>) and is represented as a space-filling model.

of an anti-*PfAMA-1* invasion-inhibitory antibody (Pizarro et al. 2005). Functional constraints may limit antigenic variation in the loop, which is proposed to form a putative ligand-binding site within the PAN domain. Together, these data suggest that the constrained areas defined by the domain II loop may serve as a potential vaccine component against *P. vivax*. Our data underscore the need for immunologic evaluation of these and other potential target sequences of acquired immunity within the *PvAMA-1* locus.

Supplementary Material

Supplementary figures I–III are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>). All sequence data obtained in the current study are available in GenBank (accession number: EF218679–EF218701).

Acknowledgments

We are grateful to all those who donated blood for this study. The assistance rendered by the medical staff of the National Hospital Colombo, Sri Lanka, and by the staff of the Malaria Research Unit, Department of Parasitology, Faculty of Medicine, Colombo, Sri Lanka, is deeply appreciated. We also wish to thank Dr Johanna Daily, Dr Nadira Karunaweera, Dr Dan Milner, Dr Kevin Militello, Jennifer Sims, and Julia Fisher for their careful review of the manuscript. The work was funded by the 2004 Burroughs Wellcome Fund–Ellison Medical Foundation–American Society of Tropical Medicine and Hygiene Postdoctoral Fellowship awarded to A.M.G.; by an Ellison Senior Scholar Award to D.F.W.; and also by the Harvard Malaria Initiative, the Burroughs Wellcome Fund, and the Ellison Medical Foundation, Program in Career Development, Research and Training in Global Infectious Diseases. T.W. was supported by the National Science Foundation, Sri Lanka.

Literature Cited

- Anderson TJ, Haubold B, Williams JT, et al. (16 co-authors). 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol.* 17:1467–1482.
- Bai T, Becker M, Gupta A, Strike P, Murphy VJ, Anders RF, Batchelor AH. 2005. Structure of AMA1 from *Plasmodium falciparum* reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proc Natl Acad Sci USA.* 102:12736–12741.
- Bruce MC, Galinski MR, Barnwell JW, Snounou G, Day KP. 1999. Polymorphism at the merozoite surface protein-3 α locus of *Plasmodium vivax*: global and local diversity. *Am J Trop Med Hyg.* 61:518–525.
- Cheng Q, Saul A. 1994. Sequence analysis of the apical membrane antigen I (AMA-1) of *Plasmodium vivax*. *Mol Biochem Parasitol.* 65:183–187.
- Chesne-Seck ML, Pizarro JC, Vulliez-Le Normand B, Collins CR, Blackman MJ, Faber BW, Remarque EJ, Kocken CH, Thomas AW, Bentley GA. 2005. Structural comparison of apical membrane antigen 1 orthologues and paralogues in apicomplexan parasites. *Mol Biochem Parasitol.* 144:55–67.
- Cole-Tobian J, King CL. 2003. Diversity and natural selection in *Plasmodium vivax* Duffy binding protein gene. *Mol Biochem Parasitol.* 127:121–132.
- Conway DJ, Cavanagh DR, Tanabe K, et al. (12 co-authors). 2000. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat Med.* 6:689–692.
- Crewther PE, Matthew ML, Flegg RH, Anders RF. 1996. Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. *Infect Immun.* 64:3310–3317.
- Cui L, Escalante AA, Imwong M, Snounou G. 2003. The genetic diversity of *Plasmodium vivax* populations. *Trends Parasitol.* 19:220–226.
- Dutta S, Malhotra P, Chauhan VS. 1995. Sequence analysis of apical membrane antigen 1 (AMA-1) of *Plasmodium cynomolgi bastianelli*. *Mol Biochem Parasitol.* 73:267–270.
- Escalante AA, Cornejo OE, Rojas A, Udhayakumar V, Lal AA. 2004. Assessing the effect of natural selection in malaria parasites. *Trends Parasitol.* 20:388–395.
- Escalante AA, Grebert HM, Chaiyaroj SC, Magris M, Biswas S, Nahlen BL, Lal AA. 2001. Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. X. Asembo Bay Cohort Project. *Mol Biochem Parasitol.* 113:279–287.
- Escalante AA, Lal AA, Ayala FJ. 1998. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics.* 149:189–202.
- Figtree M, Pasay CJ, Slade R, Cheng Q, Cloonan N, Walker J, Saul A. 2000. *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. *Mol Biochem Parasitol.* 108:53–66.
- Fraser TS, Kappe SH, Narum DL, VanBuskirk KM, Adams JH. 2001. Erythrocyte-binding activity of *Plasmodium yoelii* apical membrane antigen-1 expressed on the surface of transfected COS-7 cells. *Mol Biochem Parasitol.* 117:49–59.
- Hodder AN, Crewther PE, Matthew ML, Reid GE, Moritz RL, Simpson RJ, Anders RF. 1996. The disulfide bond structure of *Plasmodium* apical membrane antigen-1. *J Biol Chem.* 271:29446–29452.
- 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.

- Imwong M, Pukrittakayamee S, Looareesuwan S, Pasvol G, Poirreiz J, White NJ, Snounou G. 2001. Association of genetic mutations in *Plasmodium vivax* dhfr with resistance to sulfadoxine-pyrimethamine: geographical and clinical correlates. *Antimicrob Agents Chemother.* 45:3122–3127.
- Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NP, White NJ, Anderson TJ. 2006. Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Mol Biol Evol.* 23:1016–1018.
- Jukes TH, Cantor RC. 1969. Evolution of protein molecules. In: Munro HN, editor. *Mammalian protein metabolism*. New York: Academic Press. p. 21–132.
- Kato K, Mayer DC, Singh S, Reid M, Miller LH. 2005. Domain III of *Plasmodium falciparum* apical membrane antigen 1 binds to the erythrocyte membrane protein Kx. *Proc Natl Acad Sci USA.* 102:5552–5557.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5:150–163.
- Lal AA, Hughes MA, Oliveira DA, Nelson C, Bloland PB, Oloo AJ, Hawley WE, Hightower AW, Nahlen BL, Udhayakumar V. 1996. Identification of T-cell determinants in natural immune responses to the *Plasmodium falciparum* apical membrane antigen (AMA-1) in an adult population exposed to malaria. *Infect Immun.* 64:1054–1059.
- Leclerc MC, Durand P, Gauthier C, Patot S, Billotte N, Menegon M, Severini C, Ayala FJ, Renaud F. 2004. Meager genetic variability of the human malaria agent *Plasmodium vivax*. *Proc Natl Acad Sci USA.* 101:14455–14460.
- Mahajan RC, Farooq U, Dubey ML, Malla N. 2005. Genetic polymorphism in *Plasmodium falciparum* vaccine candidate antigens. *Indian J Pathol Microbiol.* 48:429–438.
- Malkin EM, Diemert DJ, McArthur JH, et al. (18 co-authors). 2005. Phase 1 clinical trial of apical membrane antigen 1: an asexual blood-stage vaccine for *Plasmodium falciparum* malaria. *Infect Immun.* 73:3677–3685.
- Martz E. 2002. Protein Explorer: easy yet powerful macromolecular visualization. *Trends Biochem Sci.* 27:107–109.
- Mendis C, Gamage-Mendis AC, De Zoysa AP, Abhayawardena TA, Carter R, Herath PR, Mendis KN. 1990. Characteristics of malaria transmission in Kataragama, Sri Lanka: a focus for immuno-epidemiological studies. *Am J Trop Med Hyg.* 42:298–308.
- Mendis K, Sina BJ, Marchesini P, Carter R. 2001. The neglected burden of *Plasmodium vivax* malaria. *Am J Trop Med Hyg.* 64:97–106.
- Nair M, Hinds MG, Coley AM, Hodder AN, Foley M, Anders RF, Norton RS. 2002. Structure of domain III of the blood-stage malaria vaccine candidate, *Plasmodium falciparum* apical membrane antigen 1 (AMA1). *J Mol Biol.* 322:741–753.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 3:418–426.
- Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* 148:929–936.
- Pizarro JC, Vulliez-Le Normand B, Chesne-Seck ML, et al. (12 co-authors). 2005. Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science.* 308:408–411.
- Polley SD, Chokeyindachai W, Conway DJ. 2003. Allele frequency-based analyses robustly map sequence sites under balancing selection in a malaria vaccine candidate antigen. *Genetics.* 165:555–561.
- Polley SD, Conway DJ. 2001. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics.* 158:1505–1512.
- Putaporntip C, Jongwutiwes S, Tia T, Ferreira MU, Kanbara H, Tanabe K. 2001. Diversity in the thrombospondin-related adhesive protein gene (TRAP) of *Plasmodium vivax*. *Gene.* 268:97–104.
- Rodrigues MH, Rodrigues KM, Oliveira TR, Comodo AN, Rodrigues MM, Kocken CH, Thomas AW, Soares IS. 2005. Antibody response of naturally infected individuals to recombinant *Plasmodium vivax* apical membrane antigen-1. *Int J Parasitol.* 35:185–192.
- Rozas J, Rozas R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics.* 15:174–175.
- Saul A, Lawrence G, Allworth A, et al. 2005. (15 co-authors). A human phase 1 vaccine clinical trial of the *Plasmodium falciparum* malaria vaccine candidate apical membrane antigen 1 in Montanide ISA720 adjuvant. *Vaccine.* 23:3076–3083.
- Sina B. 2002. Focus on *Plasmodium vivax*. *Trends Parasitol.* 18:287–289.
- Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A. 2006. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature.* 439:741–744.
- Thomas SM, Ndir O, Dieng T, Mboup S, Wypij D, Maguire JH, Wirth DF. 2002. In vitro chloroquine susceptibility and PCR analysis of pfcrt and pfmdr1 polymorphisms in *Plasmodium falciparum* isolates from Senegal. *Am J Trop Med Hyg.* 66:474–480.
- Thomas AW, Trape JF, Rogier C, Goncalves A, Rosario VE, Narum DL. 1994. High prevalence of natural antibodies against *Plasmodium falciparum* 83-kilodalton apical membrane antigen (PF83/AMA-1) as detected by capture-enzyme-linked immunosorbent assay using full-length baculovirus recombinant PF83/AMA-1. *Am J Trop Med Hyg.* 51:730–740.
- Triglia T, Healer J, Caruana SR, Hodder AN, Anders RF, Crabb BS, Cowman AF. 2000. Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. *Mol Microbiol.* 38:706–718.
- Verra F, Hughes AL. 1999. Natural selection on apical membrane antigen-1 of *Plasmodium falciparum*. *Parassitologia.* 41:93–95.
- Verra F, Hughes AL. 2000. Evidence for ancient balanced polymorphism at the Apical Membrane Antigen-1 (AMA-1) locus of *Plasmodium falciparum*. *Mol Biochem Parasitol.* 105:149–153.
- Waters AP, Thomas AW, Deans JA, Mitchell GH, Hudson DE, Miller LH, McCutchan TF, Cohen S. 1990. A merozoite receptor protein from *Plasmodium knowlesi* is highly conserved and distributed throughout *Plasmodium*. *J Biol Chem.* 265:17974–17979.
- Wickramarachchi T, Premaratne PH, Perera KL, Bandara S, Kocken CH, Thomas AW, Handunnetti SM, Udagama-Randeniya PV. 2006. Natural human antibody responses to *Plasmodium vivax* apical membrane antigen 1 under low transmission and unstable malaria conditions in Sri Lanka. *Infect Immun.* 74:798–801.
- Wilairatana P, Silachamroon U, Krudsood S, Singhasivanon P, Treeprasertsuk S, Bussaratid V, Phumratanapapin W, Srivilirit S, Looareesuwan S. 1999. Efficacy of primaquine regimens for primaquine-resistant *Plasmodium vivax* malaria in Thailand. *Am J Trop Med Hyg.* 61:973–977.
- Wilson DJ, McVean G. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics.* 172:1411–1425.

Sarah Tishkoff, Associate Editor

Accepted January 9, 2007