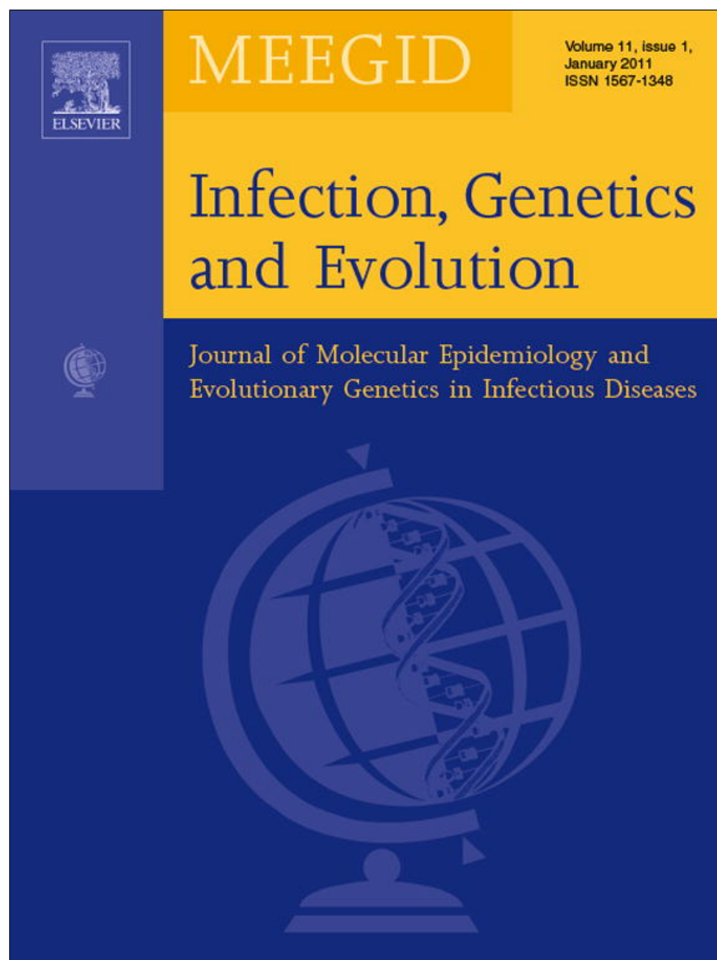
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Infection, Genetics and Evolution 11 (2011) 145–156

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Genetic diversity and recombination at the C-terminal fragment of the merozoite surface protein-1 of *Plasmodium vivax* (PvMSP-1) in Sri Lanka

Sajani Dias [a], Shirley Longacre [b], Ananias A. Escalante [c], Preethi V. Udagama-Randeniya [a,*]

[a] *Department of Zoology, Faculty of Science, University of Colombo, No 94, Cumaratunga Munidasa Mawatha, Colombo 03, Sri Lanka*
[b] *Laboratoire de Vaccinologie, Department de Parasitologie, Institut Pasteur, Paris Cedex 15, France*
[c] *School of Life Sciences, Arizona State University, Tempe, AZ, USA*

ARTICLE INFO

Article history:
Received 15 June 2010
Received in revised form 16 September 2010
Accepted 19 September 2010
Available online 8 October 2010

Keywords:
Sri Lanka
Malaria
Plasmodium vivax
Merozoite surface protein-1_{42}
Genetic diversity

ABSTRACT

Extensive polymorphism in the genes encoding for surface antigens of *Plasmodium falciparum* and *Plasmodium vivax* has been a serious impediment for malaria vaccine development. One such antigen is the merozoite surface protein-1 (MSP-1). The MSP-1 precursor after proteolytic cleavage generates a C-terminal fragment of 42 kDa (MSP-1_{42}), which subsequently produces 33 kDa (MSP-1_{33}) and 19 kDa (MSP-1_{19}) fragments. Since MSP-1_{42} is currently being considered as a candidate for vaccine development against blood stage malaria it is important to catalogue the existing diversity in this antigen in natural *P. vivax* infections. Here we investigated the level of genetic diversity in the PvMSP-1_{42} gene fragment in 95 single clone *P. vivax* infections in Sri Lanka. We observed that the PvMSP-1_{19} fragment was highly conserved among these samples, whereas the PvMSP-1_{33} fragment exhibited extensive diversity with 39 polymorphic amino acid positions (corresponding to 27 haplotypes, 19 of which were unique to Sri Lanka). Of these 27 PvMSP-1_{42} haplotypes, 24 belonged to hypervariable region (HVR) T1-T7 types, while 3 haplotypes were generated by interallelic recombination between T1/T3 (HVRT8-T9) and T2/T3 (HVRT10). In addition, we analysed 107 PvMSP-1_{42} sequences (corresponding to 62 haplotypes, H28 to H89) deposited in the NCBI GenBank database from other regions of the world. Seventy-four of these correspond to 9 of the 10 HVR types (HVR-T7 was unique to Sri Lanka). Two novel HVR types, T11 and T12, with a double recombination between HVR-T1/T3 and HVRT6/T2, were derived from South America and Thailand, respectively. T cell epitope polymorphism arising due to non-synonymous substitutions in PvMSP-1_{33} may result in differential binding of the polymorphic peptides to class II MHC alleles, inducing different host immune responses. In conclusion, under low transmission and unstable malaria conditions prevalent in Sri Lanka, extensive allelic polymorphism was evident at PvMSP-1_{33} due to recombination, mutation, and balancing selection. In contrast, PvMSP-1_{19} is highly conserved, greatly enhancing its suitability as a malaria vaccine candidate.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Malaria is endemic in 109 countries, with an estimated 247 million cases among 3.3 billion people at risk in 2006 (WMR, 2008). Although most of the malaria related deaths are due to *Plasmodium falciparum*, there have been several reports in the last few years which have documented severe complications and deaths due to *Plasmodium vivax* infections (Price et al., 2007). *P. vivax* is responsible for up to 400 million infections each year, representing the most widespread *Plasmodium* species. Although both species are endemic in Sri Lanka, the majority of reported malaria cases (65–80%) are due to *P. vivax* (Konradsen et al., 2000). Sri Lanka has two seasonal peaks, one at the beginning of the year and a larger one around June (Briët et al., 2003). The incidence of malaria in Sri Lanka has decreased significantly over the past few years, with only 196 cases reported for both species in 2007 (Annual Report of the Anti-malaria Campaign, 2007).

The malaria control program has traditionally relied on vector control (e.g. insecticides and bed nets) and case management through chemotherapy. Unfortunately, because of increasing resistance to both insecticides and anti-malarials, these two measures alone may not be sufficient to durably reduce the global burden of malaria. Thus effective, long-lasting malaria control may depend on developing cheap, broadly protective vaccines to both species (Gardiner et al., 2005). A blood stage vaccine generally aims to prevent or significantly reduce blood stage parasitemia either by reducing merozoite invasion of red cells or by targeted destruction of parasitized red cells. While progress in the development of such

* Corresponding author. Tel.: +94 11 2503399; fax: +94 11 2503148.
E-mail address: dappvr@yahoo.com (P.V. Udagama-Randeniya).

1567-1348/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2010.09.007

Author's personal copy

vaccines has been hampered by a number of factors, antigenic diversity has played a major role (Ferreira et al., 2004). This diversity is generated and maintained by several factors, including genetic recombination during the sexual phase of the parasite reproduction in the mosquito and positive natural selection by the host immune system (Chen et al., 2000; Escalante et al., 2004). This variation may partly explain why the acquisition of natural immunity to malaria is slow since the immune system is exposed to a constantly changing parasite population.

The merozoite is the parasite stage that invades circulating erythrocytes and reticulocytes during the parasite life cycle (Cowman and Crabb, 2006). Vaccine development efforts have focused on merozoite surface proteins (MSPs) because they are accessible to antibodies and complement, and they play critical roles in erythrocyte invasion (Holder, 2009). *P. vivax* merozoite surface protein-1 (PvMSP-1), like *P. falciparum* MSP-1 (PfMSP-1), after proteolytic processing generates a C-terminal fragment of 42 kDa (MSP-1_{42}), which subsequently produces 33 kDa (MSP-1_{33}) and 19 kDa (MSP-1_{19}) fragments. The MSP-1_{19} remains on the merozoite surface during and just after erythrocyte invasion (Holder, 2009). Both MSP-1_{42} and MSP-1_{19} fragments are under consideration for vaccine development (Galinski and Barnwell, 2008; Holder, 2009). In an immuno-epidemiological study carried out in Sri Lanka, individuals responded more to PvMSP-1_{42} than to PvMSP-1_{19} (Wickramarachchi et al., 2007). The baculovirus produced PcMSP-1_{19} antigen, closely related to PvMSP-1_{19}, was highly protective in vaccination trials carried out in the natural simian host-parasite system involving *Plasmodium cynomolgi* and the toque monkey *Macaca sinica* (Perera et al., 1998; Amaratunga, 2004).

Although sequence variation in *PvMSP-1* genes has been studied extensively, few studies have focused on the diversity in the regions coding for PvMSP-1_{42} (PvMSP-1_{33} and PvMSP-1_{19}) (Tanabe et al., 1987; Putaporntip et al., 2002, 2006; Escalante et al., 1998; Pacheco et al., 2007; Thakur et al., 2008; Sawai et al., 2010). This lack of interest in vaccine candidate diversity is surprising, given that host immune responses have specifically maintained it as an effective parasite survival strategy (Tanabe et al., 2007), and that subunit vaccines based on polymorphic polypeptides are destined to have short-lived efficacy at best. Like *P. falciparum* MSP-1_{42}, *P. vivax* MSP-1_{42}, exhibits extensive genetic polymorphism in natural infections (Escalante et al., 1998; Conway et al., 2000; Putaporntip et al., 2002; Pacheco et al., 2007; Thakur et al., 2008). In previous studies it has been shown that in *P. falciparum*, MSP-1_{19} fragment was under positive selection and the MSP-1_{33} was neutral or under purifying selection, while the opposite pattern was observed in *P. vivax* (Pacheco et al., 2007; Thakur et al., 2008).

Protection from infectious disease by the host immune response requires specific molecular recognition of unique epitopes of a given pathogen. The complex interplay of B and T cell epitopes of a parasite antigen, with relevant host MHC molecules are central to the specific stimulation of humoral and cell mediated host immune response(s). Therefore, polymorphism in predicted B and T cell epitopes of a parasite antigen in different parasite strains will enable parasites to escape host immune responses (Tanabe et al., 2007).

In this study we aimed to evaluate how the genetic diversity in the PvMSP-1_{42} locus is generated and maintained in natural *P. vivax* infections in Sri Lanka, where low transmission and unstable malaria prevails. Tests of diversity and of neutrality, statistical analysis of recombination and linkage disequilibrium, phylogenetic analysis, fixation index values, and polymorphism of predicted B and T cell epitopes were examined. Furthermore, we compared our data with worldwide PvMSP-1_{42} sequences obtained from the NCBI GenBank database to understand the global picture of PvMSP-1_{42} diversity.

2. Materials and methods

2.1. *P. vivax* isolates

This study was approved by the ethics review committee of the University of Colombo, Sri Lanka (EC/04/103). Following informed voluntary consent from patients tested positive for *P. vivax* infection via Giemsa stained thick and thin blood smears, 5 ml of venous blood was collected from each patient (age >15) prior to anti-malarial therapy. The samples were collected from December 1998 to March 2000 from three different regions; (i) General Hospital, Anuradhapura (8°22′N, 80°20′E; $N = 42$); (ii) Malaria Research station, Kataragama (6°25′N, 81°20′E; $N = 73$) and (iii) National Hospital of Sri Lanka, Colombo (7°55′N, 79°50′E; $N = 52$) (Wickramarachchi et al., 2007). Anuradhapura and Kataragama are regions endemic to vivax malaria, situated 250 km apart from each other. As samples collected from Anuradhapura were from individuals residing within that district, and those from Kataragama were from residents from a cluster of 7 contiguous villages comprising an area 10 km^2 (Mendis et al., 1990), these two sets of isolates constitute discrete study populations. Conversely, individuals living in non-endemic Colombo had contracted the infections during visits to areas of the island with vivax transmission (Fonseka and Mendis, 1987). The origins of these infections ($N = 52$) were traced to eleven administrative districts of the island including the two endemic areas ($N = 22$) of this study. During 1995–2000 the annual parasite incidence (API) per 1000 due to *P. vivax* was 0–1.25, 20–40 and 80–160 in Colombo, Anuradhapura and Kataragama, respectively (Briët et al., 2003). Parasitemias ranged from 0.0001 to 0.025%.

2.2. PvMSP-1_{42} fragment amplification and sequencing

Parasite DNA was extracted from 5 ml of venous blood as previously described (Gunasekera et al., 2007). Since patients could be infected with more than one *P. vivax* strain, samples were genotyped at the polymorphic PvMSP-3α locus using nested PCR and restriction fragment length polymorphism as described by Bruce et al. (1999) to select single clone infections. Only single clone infections were considered for PvMSP-1_{42} amplification and sequencing. We used two rounds of PCR to amplify 1162 bp PvMSP-1_{42} fragment (Sal-I nucleotide positions 4051–5213). First round of amplification was done using 5′GATGACGAC GGGGA GGAAGACC3′ (P1) and 5′AAGCTCCATGCACAGGAGG3′ (P2) primers whereas second round of amplification was done using 5′GACCAA GTAACAACGGGAG3′ (N1) and 5′GGACAAGCTTAGGAA GCTGG3′ (N2) primers (Manamperi, 2002). Cycling parameters for both primer pairs were: 92 °C-2 min (1 cycle), 91 °C-10 s, 59 °C-25 s, and 72 °C-25 s (30 cycles), with a final extension at 72 °C-4 min (1 cycle). Primary reactions contained 1 μl of DNA, 0.5 units of *Taq* polymerase (Promega, USA), 0.2 μM of each primer, 0.1 mM dNTPs in a 25 μl of reaction mix with 1.5 mM $MgCl_2$. Nested reactions contained 1 μl of the primary PCR product as template in a 50 μl reaction mix as above. Amplified products were resolved on a 1% agarose gel containing 0.25 μg/ml of ethidium bromide, visualized under UV and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, USA). Sequencing of the PCR products were done on both strands using N1 and N2 primers at Macrogen, Korea. Some isolates underwent four-fold sequence coverage to confirm rare polymorphisms.

2.3. Data analysis

Sequences were aligned and verified using Seqman II (DNAStar, Madison, WI). DnaSP version 4.50 (Rozas et al., 2003) and MEGA

Author's personal copy

4.1 (Tamura et al., 2007) were used for various population genetic analyses. The two discrete study populations from Anuradhapura and Kataragama were compared with respect to the different genetic indices. However, the mixed population from Colombo that overlapped with isolates from the two endemic study areas was only included in the local test population to draw comparisons with other global isolates.

We estimated the numbers of synonymous nucleotide substitutions per synonymous site (Ds), the number of non-synonymous substitutions per non-synonymous site (Dn), and the difference between the non-synonymous and synonymous substitutions (Dn – Ds) by Nei and Gojobori's method to get evidence of natural selection at PvMSP-1_{42} in MEGA 4.1. The McDonald–Kreitman test was applied as a second test of natural selection, which compares the intra- and interspecific number of synonymous and non-synonymous sites. Ten *P. cynomolgi* sequences were used as the out group and Fisher's exact test was applied to the data to test for significant non-randomness ($P < 0.05$), and the skew from randomness was calculated as the Neutrality Index in DnaSP. Phylogenetic analysis was performed by neighbor-joining (NJ) method with Kimura 2-parameter distance matrix in MEGA 4.1 (Tamura et al., 2007). All sequences ($N = 95$) reported in this paper were submitted to the NCBI GenBank database under the accession numbers GU175174-GU175268. PvMSP-1_{42} sequences ($N = 107$) corresponding to Sal I amino acids 1384–1722 from diverse geographical regions were obtained from the NCBI GenBank database to compare polymorphism with the Sri Lankan isolates.

2.4. Prediction of B and T cell epitopes at PvMSP-1_{33} and PvMSP-1_{19}

Prediction of linear B cell epitopes was carried out on all 27 amino acid haplotypes of the PvMSP-1_{33} and the single amino acid haplotype of the PvMSP-1_{19} using the BcPred software (EL-Manzalawy et al., 2008). The 12 amino acid length epitopes with a specificity of 75% were generated (EL-Manzalawy et al., 2008). Subsequently, differential binding of T cell epitopes spanning the 33 kDa and 19 kDa fragments of PvMSP-1_{42} were predicted for four MHC class II HLA alleles, HLA-DRB1*0101, DRB1*0401, DRB1*0701 and DRB1*1101 by the ProPred MHC class II binding peptide prediction server (Singh and Raghava, 2001). These four HLA alleles were selected because these were among the 10 most prevalent in Afro-American and Asian populations (Southwood et al., 1998).

3. Results

Of 167 *P. vivax* infected blood samples initially tested, only 95 (Colombo, $N = 37$; Anuradhapura, $N = 22$ and Kataragama, $N = 36$) were successfully amplified for the 1016 bp PvMSP-1_{42} fragment. We performed all genetic analyses for the PvMSP-1_{42}, PvMSP-1_{33} and PvMSP-1_{19} fragments separately in the two endemic study populations, and in the entire local population and compared the latter with previously published global isolates.

3.1. PvMSP-1_{19} polymorphism

The PvMSP-1_{19} fragment was found to be highly conserved in Sri Lankan isolates, except for two synonymous mutations at positions 4941 and 4959 (Sal I numbering). These mutations occurred together or individually in 6 isolates (Kataragama, $N = 5$ and Anuradhapura, $N = 1$). The nucleotide diversity per site (π) values for PvMSP-1_{19} reported in previous studies (Manamperi, 2002; Pacheco et al., 2007; Thakur et al., 2008) were considerably lower than that observed here due to the above two synonymous substitutions (Table 1). However, Tajima's D for this fragment was negative indicative of an excess of low frequency variants in the sample (Table 1). The neutrality indices, obtained using the McDonald–Kreitman test with *P. cynomolgi* as the out group, showed that the PvMSP-1_{19} fragment was under purifying selection, such that polymorphism was actively suppressed by natural selection (Table 2).

3.2. PvMSP-1_{42}/PvMSP-1_{33} polymorphism

In the 33 kDa fragment, 56 polymorphic nucleotide positions and 61 distinct point mutations were evident from each of the two endemic sites. Differences between the nucleotide diversity of non-synonymous mutations per non-synonymous site (Dn) and that of synonymous mutations per synonymous site (Ds) were indicative of a balancing selection acting on the 33 kDa fragment in Anuradhapura (Dn – Ds = 0.0152, $P < 0.05$). Values of P less than 0.05 were considered significant at the 5% level where the probability of the null hypothesis of strict-neutrality (Dn = Ds) was rejected in favour of the alternative hypothesis (Dn > Ds). In contrast, a significant departure from neutrality for Tajima's D test in the 33 kDa fragment was observed in Kataragama (Table 1).

All 58 polymorphic MSP-1_{42} positions in the entire local population were parsimony informative sites (having a minimum

Table 1
DNA sequence polymorphisms in *P. vivax* MSP-1 among Sri Lankan field isolates.

	π (S.D.)	Ds	Dn	Dn – Ds (S.D.)	*Z*-test	Tajima's *D* test
42 kDa						
Entire population ($N = 95$)	0.0231 (0.0005)	0.0170	0.0253	0.0083 (0.0045)	$P > 0.05$	2.86687 ($P < 0.05$)
Anuradhapura ($N = 22$)	0.0231 (0.0015)	0.0150	0.0258	0.01076 (0.0056)	$P < 0.05$	1.51312 ($P > 0.10$)
Kataragama ($N = 36$)	0.0238 (0.0009)	0.0185	0.0258	0.0073 (0.0064)	$P > 0.05$	2.16328 ($P < 0.05$)
33 kDa						
Entire population ($N = 95$)	0.0310 (0.0007)	0.0213	0.0347	0.0134 (0.0078)	$P < 0.05$	2.97050 ($P < 0.01$)
Anuradhapura ($N = 22$)	0.0313 (0.0012)	0.0201	0.0353	0.0152 (0.0075)	$P < 0.05$	1.58063 ($P > 0.10$)
Kataragama ($N = 36$)	0.0319 (0.0011)	0.0231	0.0354	0.0122 (0.0085)	$P > 0.05$	2.26303 ($P < 0.05$)
19 kDa						
Entire population ($N = 95$)	0.0012 (0.0003)	0.0059	0.00	−0.0058 (0.0040)	$P > 0.05$	−0.24254 ($P > 0.10$)
Anuradhapura ($N = 22$)	0.0003 (0.0003)	0.0016	0.00	−0.0016 (0.0016)	$P > 0.05$	−1.16240 ($P > 0.10$)
Kataragama ($N = 36$)	0.0013 (0.0005)	0.0063	0.00	−0.0063 (0.0046)	$P > 0.05$	−0.51702 ($P > 0.10$)

π - nucleotide diversity with standard deviation (S.D.).
N - number of sequences.
Ds - nucleotide diversity of synonymous mutation per synonymous site, Dn - nucleotide diversity of non-synonymous mutation per non-synonymous site, using the Nei and Gojobori method
Dn – Ds - the difference of Dn and Ds with their standard deviation estimated by bootstrap with 1000 pseudo replicates. *Z*-test was applied to test significance
Tajima's *D* - to detect departure from the neutral mode.

Author's personal copy

Table 2
The McDonald and Kreitman test for different regions of MSP-1.

Region		Interspecific fixed differences	Intraspecific difference		Neutrality index[a]	Fisher's exact test[b]
			P. vivax (*N* = 95)	*P. cynamolgi* (*N* = 10)	M K test	(2-tailed)
Entire population						
42 kDa	Synonymous	49	8	42	2.042	0.010277*
	Non-synonymous	46	43	57		
33 kDa	Synonymous	38	6	35	2.219	0.010240*
	Non-synonymous	36	43	47		
19 kDa	Synonymous	11	2	7	1.222	1.000000
	Non-synoymous	10	–	10		
Anuradhapura						
42 kDa	Synonymous	49	7	42	2.062	0.009958**
	Non-synonymous	46	42	57		
33 kDa	Synonymous	38	6	35	2.192	0.0105*
	Non-synonymous	36	42	47		
19 kDa	Synonymous	11	1	7	1.375	0.7512
	Non-synoymous	10	–	10		
Kataragama						
42 kDa	Synonymous	49	7	42	2.019	0.0105*
	Non-synonymous	46	43	57		
33 kDa	Synonymous	38	5	35	2.192	0.0105*
	Non-synonymous	36	43	47		
19 kDa	Synonymous	11	2	7	1.222	1.0
	Non-synoymous	10	–	10		

[a] Neutrality index indicates the extent to which the levels of amino acid polymorphism depart from the expected in the neutral model.
[b] Fisher's exact test. *$0.01 < P < 0.05$; **$0.001 < P < 0.01$.

of two nucleotides each present at least twice), with 53 having two variants, four having three, and one having four (nucleotide position 4550). In particular, a polymorphic region of 113 bp at positions 4468–4581 having a clear excess of non-synonymous substitutions, confirmed previous reports as were overall π values for the PvMSP-1_{33} in Table 1 (Manamperi, 2002; Pacheco et al., 2007; Thakur et al., 2008). The haplotype diversity and nucleotide diversity for PvMSP-1_{42} in the entire local population were 0.962 ± 0.007 S.D. and 0.0231 ± 0.0005 S.D., respectively, and were also similar to those reported from India (Thakur et al., 2008). Comparatively elevated π values for Kataragama compared to Anuradhapura for both the 42 and 33 kDa fragments may point at a higher transmission level in Kataragama.

Taking *P. cynomolgi* as the out group, the McDonald–Kreitman test for neutrality was applied separately to the 42 and 33 kDa fragments (Table 2). In both endemic study areas, a significant departure from neutrality on the basis of excess inter-specific non-synonymous 42 kDa polymorphisms relative to non-synonymous divergence from *P. vivax* was evident, where Anuradhapura showed a particularly significant departure from neutrality ($NI = 2.062$, $P < 0.001$) compared to Kataragama ($NI = 2.019$). Surprisingly PvMSP-1_{33} had significantly less inter-specific non-synonymous polymorphism at both the endemic sites ($P < 0.01$), providing strong evidence for purifying selection.

Of the 17 minimum recombination events (Rm) observed in PvMSP-1_{33}, 11 were common to both the endemic study sites. The linkage disequilibrium index (R^2) plotted against the nucleotide distances showed decline across the entire 1016 bp region indicating that intragenic recombination may contribute to PvMSP-1_{42} diversity observed locally.

The degree of genetic differentiation between populations was estimated by the F_{ST} value. This value ranges from 0 to 1 and measures the amount of allelic diversity within and between geographically distinct populations. A value of 0 indicates all alleles are evenly distributed among populations, and 1 indicates an allele is found in only one population (Cole-Tobian and King, 2003). Comparison of the genetic differentiation within Sri Lankan PvMSP-1_{42} diversity showed low F_{ST} values of -0.0205 between Anuradhapura and Kataragama; negative F_{ST} values can be interpreted as "0" and therefore no genetic differentiation was observed. However, the F_{ST} value for the entire Sri Lankan population with other countries showed the highest F_{ST} values with Bangladesh ($F_{ST} = 0.1894$) followed by Thailand ($F_{ST} = 0.0751$), Brazil ($F_{ST} = 0.0305$) and India ($F_{ST} = 0.0039$).

3.3. Classification and identification of haplotypes

Ninety five PvMSP-1_{42} sequences in this study defined 27 amino acid haplotypes, identical neither to the Salvador I nor the Belem strains, across the entire 33 kDa fragment (Table 3) of which 14 were common to all three study sites (Table 4). A single 19 kDa amino acid haplotype generally confirmed previous results (Manamperi, 2002; Putaporntip et al., 2000, 2002; Pacheco et al., 2007).

Longacre (1995), divided the 33 kDa fragment into three segments based on its diversity: (i) N-terminal region I (RI) upstream of the HVR block with 7 variant positions, (ii) a central 38 amino acid hypervariable region (HVR or RII) including positions 1490-1527 with 29 variants, and (iii) region III (RIII) downstream of the HVR block and upstream of the 19 kDa fragment with 3 variant positions defined in this study (Table 3). All variant positions in RI and RIII and over half in the HVR (19/29) were dimorphic, where only two alternating amino acid residues were observed in local and global isolates (Table 3).

Five more variant positions at RI compared to Manamperi's (2002) Sri Lankan cohort were observed in the present study, but none of these RI haplotypes were identical to either the Sal-I or the Belem strains. Interestingly, 9 isolates defined two variant positions (1443-K>E and 1458-K>E) unique to this study (Tables 3 and 4), whereas four haplotypes from other studies (H70, 76, 77, 75) defined an eighth variant position (1468-L>V; Table 3).

In RIII, 15 of the 27 haplotypes from the present study corresponded to the single Belem-Sal-I prototype. Of the 3 variant positions, 1566-N>S was frequently observed in both Sri Lankan and other global isolates, while 1586-N>T was present in H25-H27 from Sri Lanka and in H66 from India, and 1595-Y>S was restricted to 6 Sri Lankan isolates (Tables 3 and 4).

The HVR (RII) displayed the most complex polymorphism, and was subdivided into 7 basic types (HVR-T1 to T7) seen in 24 of the 27 local haplotypes (Tables 3 and 4). HVR-T1 was essentially the Belem type (H1-H4), while HVR-T2 (H19-H27) corresponded to the Sal-I strain (Tables 3 and 4). Amino acid haplotypes H7-H11, H13-

Author's personal copy

H14, and H17-H18 defined HVR types T3, T4 and T5, respectively (Table 3 and 4). Altogether 94% of the isolates fell into one of the five HVR types, with a surprising 75% found in T1-T3. HVR-T6 of H5 differed from HVR-T1 by a single variant position (1511-K>E) and HVR-T7 with a single H15 isolate differed from HVR-T4 by a single variation (1516-E>V) (Tables 3 and 4). H6, appeared to be a double recombinant between Belem (HVR-T1) and HVR-T3 (T1 > T3 > T1) with a single variation (1520-T>A) and was observed in 3 different isolates representing HVR-T8. H12 of HVR-T9 represented a double recombinant between HVR-T3 and HVR-T1 (T3 > T1 > T3) with no other variations. The single H16 isolate appeared to be a recombinant between HVR-T2 (Sal-I) and T3, defining HVR type T10.

Markedly, of the 10 HVR types represented in the local isolates, 9 were observed in 74 global isolates, with the exception of HVR-T7 unique to Sri Lanka. The HVR-T1 (Belem) was observed in haplotypes H28-H30 ($N = 9$), while HVR-T2 (Sal I, H62–H70) constituted the majority of global isolates ($N = 17$). Finally, two new HVR types T11 and T12 representing a double recombination between HVR-T1/T3 and HVR-T6/T2, respectively were observed in haplotypes H44-H46 and H34-H35. The remaining 19 haplotypes (H71-H89) defined by 21 isolates from outside Sri Lanka, were generated by intragenic recombination (Table 3).

Altogether, 202 sequences defined 89 PvMSP-1_{42} amino acid haplotypes (Table 3), of which eight common to both local and global isolates represented HVR types T1, T3-T6 and T8-T9, mostly observed in India. Interestingly, H28 and H63 from Sri Lanka (Manamperi, 2002) were identical to the Belem and Sal I strains, respectively. Strikingly, 19 of the haplotypes described in this study are unique to the island and are reported for the first time.

The 47 PvMSP-1_{42} polymorphic amino acid positions identified in the 202 local and global isolates included 28 dimorphic, 10 trimorphic (T/E/A-1494, K/A/T-1505, Q/E/H-1506, K/E/Q-1511, V/D/E-1516, S/T/N-1518, S/T/A-1521, N/T/A-1524, E/Q/K-1525, F/S/L-1526), and 3 tetramorphic (N/S/A/V-1517, D/T/A/N-1520 and I/T/M/N-1527) a.a. positions (Table 3). Importantly, of these 28 dimorphic sites, our study revealed two novel sites (K/E-1443 K/E-1458) (Table 3).

3.4. Phylogenetic tree of Pvmsp-1_{42}

To investigate PvMSP-1_{42} phylogenetic relationships, we constructed a Neighbor-Joining tree based on the 27 haplotypes described here (Fig. 1). Interestingly, the haplotypes were clustered into eight groups as shown by the 8 different HVR types (HVR-T1-T8) at PvMSP-1_{33}. Haplotype H1 through H4 and H19 through H27 were closely related to the Belem and Sal I strains, respectively, and represented HVR types T1 and T2. Haplotype H7 through H11, H13 through H14, H17 through H18 and H6, H12, H16 defined six separate groups that represented the observed HVR types T3-T5 and T8-10, respectively. Though, haplotypes H5 and H15 represented unique HVR types they were clustered with

Table 3

Amino acid polymorphism in the *P. vivax* MSP-1_{33} among geographically diverse malaria endemic areas.

		1432	1443	1454	1458	1466	1468	1476	1482	1490	1491	1493	1494	1495	1496	1501	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1524	1525	1526	1527	1566	1573	1574	1586	1595	1597	1601	1612	1709	
Belem*		P	K	L	K	K	L	T	G	T	A	N	E	E	V	A	K	A	E	D	D	K	I	K	K	I	G	S	D	S	T	K	T	T	E	T	Q	S	M	N	K	E	N	Y	K	K	K	K	HVR types
Sal I[a]		S	K	I	K	M	L	T	G	D	G	K	T	E	I	D	K	K	Q	D	E	E	L	K	K	L	G	N	V	N	S	Q	D	S	K	N	E	F	I	N	K	E	N	Y	K	K	K	K	
Amino acid haplotypes A	N	RI								RII																														RIII									
H1	6	.	.	L	.	K	.	.	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	HVR-T1
H2	2	.	.	L	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	
H3	1	.	.	L	.	K	.	.	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	S	.	.	.	.	
H4	4	.	.	.	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H5	3	.	.	.	.	K	.	.	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	E	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	HVR-T6
H6	3	.	.	.	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	HVR-T8
H7	11	.	.	L	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	HVR-T3
H8	1	.	E	L	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	
H9	8	.	.	.	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H10	2	.	E	.	E	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H11	1	.	.	.	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	
H12	2	.	.	.	.	.	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	D	S	T	K	T	T	E	A	K	.	T	.	.	.	.	.	.	.	.	.	HVR-T9
H13	9	.	.	.	.	.	.	A	E	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	S	.	.	.	.	.	.	.	.	HVR-T4
H14	2	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	
H15	1	.	E	L	E	K	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	.	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	HVR-T7
H16	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	HVR-T10

Table 3 (*Continued*)

H17	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	S	.	.	.	.	HVR-T5
H18	5	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	
H19	2	.	.	.	.	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	HVR-T2
H20	9	.	.	.	.	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	
H21	2	.	.	.	E	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
H22	9	.	.	.	.	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
H23	1	.	.	.	E	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	
H24	1	.	E	.	E	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
H25	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	S	.	.	.	.	
H26	1	.	.	L	E	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	S	.	.	.	.	
H27	6	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	
B																																																	
H28	2	P	.	L	.	K	.	.	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	HVR-T1
H29	2	.	.	.	.	.	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H30	1	.	.	L	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	
H31	1	.	.	L	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	E	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	HVR-T6
H32	1	.	.	L	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	E	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	N	T	.	.	
H33	3	.	.	.	.	K	.	A	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	E	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H34	7	.	.	.	.	K	.	.	.	.	.	.	.	.	.	A	.	A	E	.	D	K	I	E	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	HVR-T12
H35	1	.	.	.	.	.	.	A	E	.	.	.	.	.	.	A	.	A	E	.	D	K	I	E	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H36	3	.	.	L	.	K	.	A	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	HVR-T8
H37	2	.	.	.	.	K	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	
H38	1	.	.	.	.	.	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H39	1	.	.	.	.	.	.	A	.	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H40	1	.	.	.	.	.	.	A	E	T	A	N	E	.	V	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	S	.	.	.	.	.	.	.	E	
H41	1	P	.	L	.	K	.	.	.	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	HVR-T3
H42	1	.	.	.	.	K	.	A	.	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	
H43	1	.	.	.	.	.	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H44	1	.	.	L	.	K	.	A	E	T	A	N	A	Q	.	.	.	.	.	.	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	R	.	HVR-T11
H45	2	.	.	L	.	K	.	A	E	T	A	N	A	Q	.	.	.	.	.	.	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	R	.	
H46	1	.	.	L	.	K	.	A	E	T	A	N	A	Q	.	.	.	.	.	.	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	.	.	.	.	.	.	.	.	.	

Author's personal copy

Table 3 (*Continued*)

H47	1	.	.	L	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	D	S	T	K	T	T	E	A	K	.	T	S	.	.	.	.	.	.	.	.	HVR-T9
H48	1	.	.	.	.	.	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	D	S	T	K	T	T	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H49	1	.	.	.	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	D	S	T	K	T	T	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H50	1	.	.	.	.	K	.	A	E	T	A	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	.	D	S	T	K	T	T	E	A	K	.	T	.	.	.	.	.	.	.	.	.	
H51	1	.	.	.	.	K	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	HVR-T4
H52	1	.	.	.	.	.	.	A	E	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	
H53	2	.	.	L	.	K	.	A	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	
H54	1	.	.	.	.	.	.	A	E	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	S	.	.	.	.	.	.	.	.	
H55	2	.	.	.	.	K	.	A	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	S	.	.	.	.	.	.	.	.	
H56	1	.	.	.	.	K	.	A	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	
H57	1	.	.	.	.	K	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	S	.	.	.	.	.	.	.	.	
H58	1	.	.	.	.	K	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	.	.	.	.	.	.	.	.	HVR-T5
H59	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	N	.	R	K	.	.	.	.	.	.	
H60	1	.	.	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	HVR-T10
H61	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	E	
H62	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HVR-T2
H63	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	
H64	4	.	.	.	.	.	.	A	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
H65	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	
H66	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	
H67	1	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	
H68	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	E	
H69	2	.	.	.	.	.	V	A	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	
H70	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	R	.	
C																																																	
H71	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	G	T	Q	L	N	.	.	.	.	.	.	.	.	.	
H72	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	
H73	1	.	.	.	.	.	.	A	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	
H74	1	P	.	L	.	K	.	.	.	T	A	N	A	.	V	A	N	.	.	.	D	K	I	.	.	I	.	S	D	A	T	K	T	T	G	T	Q	S	M	.	.	.	.	.	.	.	.	.	
H75	1	.	.	.	.	.	V	A	E	T	A	N	E	.	V	A	N	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H76	2	.	.	.	.	.	V	A	E	T	A	N	A	Q	.	A	.	A	E	.	D	K	I	.	T	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H77	2	.	.	.	.	.	V	A	E	T	A	N	A	Q	.	A	.	A	E	.	D	K	I	.	.	I	E	.	E	A	N	K	A	A	E	T	Q	S	M	S	.	.	.	.	.	.	.	.	
H78	1	.	.	.	.	.	.	.	.	.	.	N	A	Q	.	.	N	T	.	N	.	.	.	.	.	I	E	V	.	A	N	K	T	A	.	.	Q	.	.	.	.	.	.	.	.	.	.	.	
H79	1	.	.	.	.	.	.	A	E	T	A	N	A	Q	.	.	.	A	E	.	D	K	I	.	.	I	.	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	
H80	1	.	.	.	.	.	.	A	E	.	A	N	E	.	V	A	.	T	.	.	.	.	.	.	.	I	.	S	D	S	.	.	T	.	E	T	Q	L	.	S	.	.	.	.	.	.	.	.	
H81	1	.	.	L	.	K	.	.	.	T	A	N	E	.	V	A	.	T	.	.	.	.	.	.	.	I	.	S	D	S	.	.	T	.	E	T	Q	L	.	S	.	.	.	.	.	.	.	.	
H82	1	P	.	L	.	K	.	.	.	T	A	N	E	.	V	A	.	T	H	.	D	K	I	.	.	I	.	S	E	V	.	.	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	.	.	
H83	1	P	.	L	.	K	.	A	.	.	.	.	.	.	.	.	.	A	E	.	D	K	I	.	.	I	E	S	D	S	T	K	T	T	E	T	Q	S	M	.	.	.	.	.	.	.	R	.	
H84	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	K	.	Q	.	.	.	S	E	V	.	.	N	.	E	T	Q	L	.	.	.	.	.	.	.	.	.	.	
H85	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	E	.	D	S	T	K	T	T	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H86	1	.	.	.	.	K	.	A	E	.	.	.	.	.	.	.	.	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H87	1	.	.	.	.	K	.	A	E	.	.	.	.	.	.	.	.	T	.	N	.	.	.	.	.	I	E	.	E	A	N	K	T	A	E	A	K	.	T	S	.	.	.	.	.	.	.	.	
H88	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	H	.	.	.	.	.	.	.	E	S	.	.	.	.	.	.	G	T	Q	.	.	.	.	.	.	.	.	.	.	.	
H89	1	.	.	L	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	E	V	.	.	N	.	E	T	Q	S	N	.	.	.	.	.	.	.	.	.	

Note: Amino acid variant positions 1432–1709 corresponded to the two reference strains of PvMSP-1_{33} Belem* (Accession no. AF435594; Putaporntip et al., 2002) and Salvador I# (Accession no. M75674; Gibson et al., 1992). Positions of amino acid substitutions are specified vertically above with reference to the Salvador I strain. Pink and light blue correspond, respectively to Belem and Sal I specific residues; yellow residues are identical between Belem and Sal I; green residues differ from both Belem and Sal I. HVR-T1 toT12 represents the PvMSP-1_{33} hypervariable region sequence types of (A) Amino acid haplotypes from the current study, (B) Amino acid haplotypes of global isolates, and (C) Amino acid haplotypes manifested by intragenic recombination. "." represents those amino acids similar to the PvMSP-1_{33} Salvador I strain. (For interpretation of the references to color in this table legend, the reader is referred to the web version of the article.)

Author's personal copy

Table 4
The numbers of isolates present in each of the PvMSP-1_{42} amino acid haplotypes representing the three study areas of Sri Lanka.

Amino acid haplotypes	Study area N = number of isolates		
	Anuradhapura	Kataragama	Colombo
H1	√ (3)	√ (3)	
H2	√ (1)		√ (1)
H3		√ (1)	
H4		√ (1)	√ (3)
H5			√ (3)
H6	√ (1)	√ (1)	√ (1)
H7	√ (1)	√ (7)	√ (3)
H8		√ (1)	
H9	√ (3)	√ (2)	√ (3)
H10		√ (1)	√ (1)
H11			√ (1)
H12			√ (2)
H13	√ (2)	√ (5)	√ (2)
H14	√ (1)		√ (1)
H15	√ (1)		
H16			√ (1)
H17		√ (1)	
H18	√ (2)	√ (1)	√ (2)
H19	√ (1)		√ (1)
H20	√ (1)	√ (4)	√ (4)
H21		√ (1)	√ (1)
H22	√ (4)	√ (1)	√ (4)
H23			√ (1)
H24		√ (1)	
H25		√ (1)	
H26	√ (1)		
H27		√ (4)	√ (2)

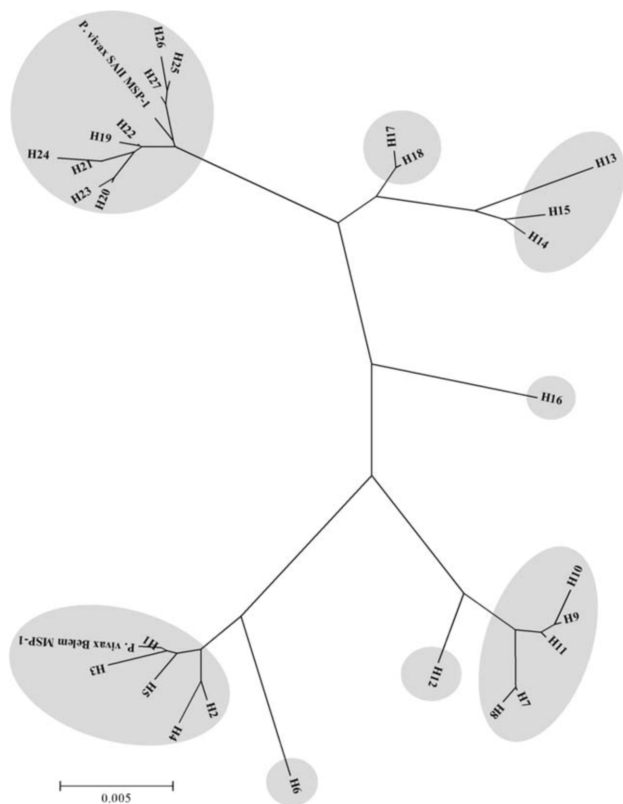


Fig. 1. Neighbor-joining tree of PvMSP-1_{42} amino acid sequence of 27 haplotypes observed among 95 Sri Lankan isolates.

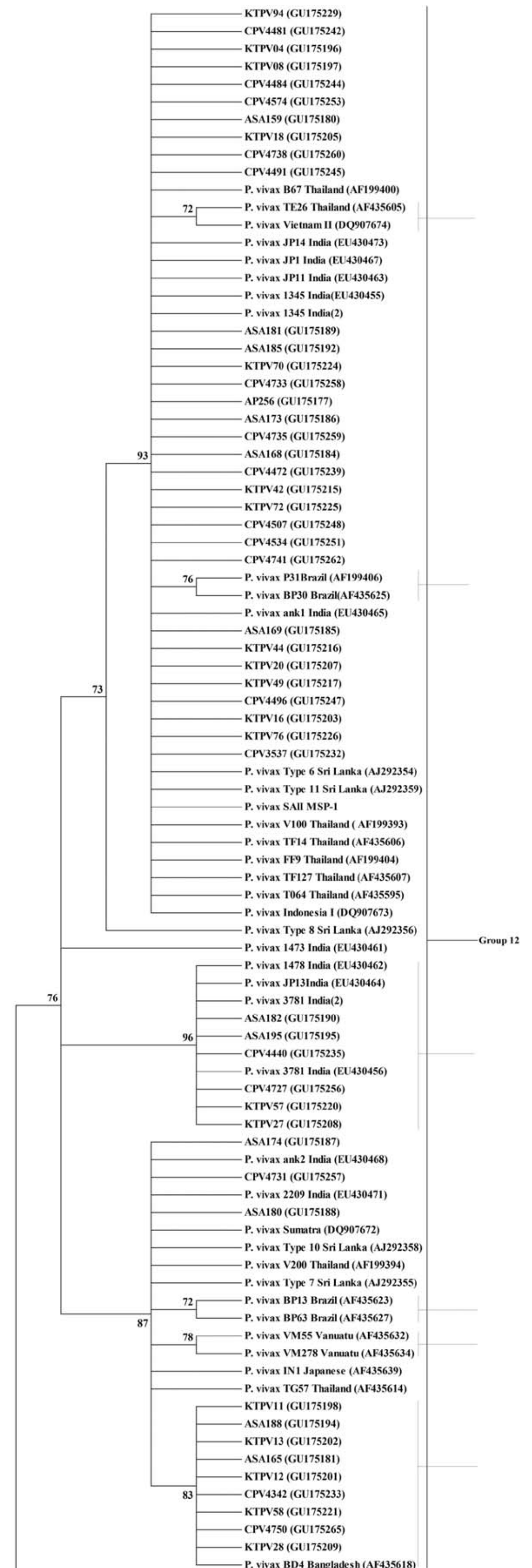


Fig. 2. Neighbor-joining tree of PvMSP-1_{42} nucleotide sequences using Kimura 2 parameter. Isolates of Sri Lanka and other geographically different countries are included. The numbers on the node of the tree are percent bootstrap values based on 1000 pseudo-replications, only boot strap values above 70% are displayed on the tree.

Author's personal copy

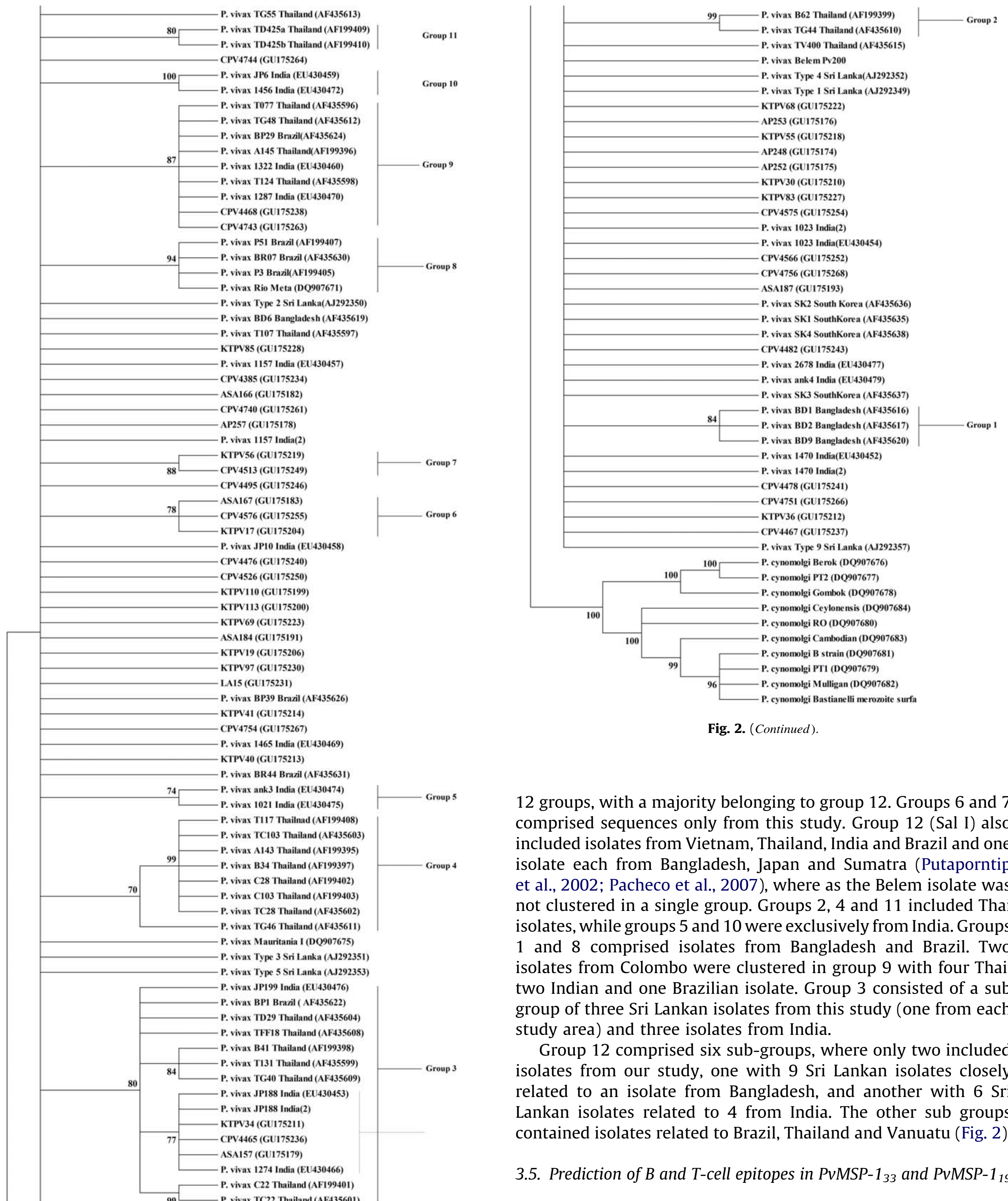


Fig. 2. (*Continued*).

Fig. 2. (*Continued*).

haplotypes of HVR-T1 and T4, respectively, due to the variation of a single amino acid each.

A phylogenetic tree constructed using all published 202 PvMSP-1_{42} nucleotide sequences defined 12 distinctive groups (Fig. 2). As indicated above, the Sri Lankan isolates were clustered in 5 of the 12 groups, with a majority belonging to group 12. Groups 6 and 7 comprised sequences only from this study. Group 12 (Sal I) also included isolates from Vietnam, Thailand, India and Brazil and one isolate each from Bangladesh, Japan and Sumatra (Putaporntip et al., 2002; Pacheco et al., 2007), where as the Belem isolate was not clustered in a single group. Groups 2, 4 and 11 included Thai isolates, while groups 5 and 10 were exclusively from India. Groups 1 and 8 comprised isolates from Bangladesh and Brazil. Two isolates from Colombo were clustered in group 9 with four Thai, two Indian and one Brazilian isolate. Group 3 consisted of a sub group of three Sri Lankan isolates from this study (one from each study area) and three isolates from India.

Group 12 comprised six sub-groups, where only two included isolates from our study, one with 9 Sri Lankan isolates closely related to an isolate from Bangladesh, and another with 6 Sri Lankan isolates related to 4 from India. The other sub groups contained isolates related to Brazil, Thailand and Vanuatu (Fig. 2).

3.5. Prediction of B and T-cell epitopes in PvMSP-1_{33} and PvMSP-1_{19}

Three linear B cell epitopes were predicted at amino acid positions 1393, 1542 and 1587 of the 33 kDa fragment common to all 27 local haplotypes. While H1 lacked a B cell epitope due to amino acid substitution at 1432, the other haplotypes included supplementary predicted B cell epitopes due to amino acid substitutions at several positions. The 19 kDa fragment comprised three predicted B cell epitopes at amino acid positions 1615, 1657 and 1678.

Author's personal copy

DRB1_0101:MYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKYSPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLATANDGVTYYNKM
GELYKTHLTAVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESLVSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNK
MDEKLEEYKKSEKKNEVKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNN
GGCAPEAECKMTDSNKIVCKCTKEGSEPL

DRB1_0401:MYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKYSPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLATANDGVTYYNKM
GELYKTHLTAVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESLVSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNK
MDEKLEEYKKSEKKNEVKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNN
GGCAPEAECKMTDSNKIVCKCTKEGSEPL

DRB1_0701:MYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKYSPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLATANDGVTYYNKM
GELYKTHLTAVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESLVSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNK
MDEKLEEYKKSEKKNEVKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNN
GGCAPEAECKMTDSNKIVCKCTKEGSEPL

DRB1_1101:MYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKYSPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLATANDGVTYYNKM
GELYKTHLTAVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESLVSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNK
MDEKLEEYKKSEKKNEVKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNN
GGCAPEAECKMTDSNKIVCKCTKEGSEPL

Fig. 3. Predicted T cell epitopes of the PvMSP-1_{42} Belem strain obtained for the HLA-DRB1*0101, HLA-DRB1*0401, HLA-DRB1*0701 and HLA-DRB1*1101 alleles. Predicted T cell epitopes to each MHC allele is highlighted in blue with the first position highlighted in red. The hypervariable region of PvMSP-1_{42} is underlined and the polymorphic epitopes are circled. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Predicted T cell epitopes on MSP-1_{33} shared by all 27 local haplotypes were clustered on either side of the hypervariable region and included 1 for HLA-DRB1*0101, 3 for HLA-DRB1*0401, 2 for HLA-DRB1*0701 and 4 for HLA-DRB1*1101. Additionally common T cell epitopes were predicted for H1-H6, as well as a novel H3 epitope. Polymorphism in the hypervariable regions of H7-H27 generated diverse predicted T cell epitopes for all of the HLA alleles, except HLA-DRB1*0701. While two MSP-1_{19} T cell epitopes were predicted for HLA-DRB1*0101 and one for both HLA-DRB1*0401 and 1101, the epitope LLNVQTQLL predicted for HLA-DRB*0701 was positioned at the end of MSP-1_{33} and beginning of MSP-1_{19} (Fig. 3).

All predicted B and T cell epitopes at PvMSP-1_{33} were polymorphic and had more non-synonymous than synonymous substitutions. The linear B cell epitope predicted on PvMSP-1_{19} (at position 1615) consisted more of synonymous mutations, while the other two B cell epitopes and all T cell epitopes at PvMSP-1_{19} were conserved even at the nucleotide level (data not shown).

3.6. Discussion

Many factors may contribute to genetic diversity in malaria parasites; mutation, recombination, gene flow between different populations and natural selection (Cole-Tobian and King, 2003). The overall aim of our study was to determine the factors that generated and maintained genetic diversity related to natural immunity in a key vaccine candidate antigen, PvMSP-1_{42}, in Sri Lanka and to compare these observations with the global scenario.

The level of polymorphism observed in PvMSP-1_{42} ($\pi = 0.0231$) in Sri Lanka was low compared to PvMSP-3α ($\pi = 0.059$) (Wickramarachchi et al., 2010), but higher than other merozoite proteins DBPII ($\pi = 0.0098$) and AMA-1 ($\pi = 0.0095$) (P.V. Udagama-Randeniya, unpublished data). In Sri Lanka, PvMSP-1_{33} exhibited extensive genetic diversity (with 39 polymorphic amino acid positions corresponding to 27 haplotypes) compared to reports from other regions of the world (Pacheco et al., 2007; Thakur et al., 2008). None of the Sri Lankan haplotypes were identical to either the Salvador I or the Belem reference strains. However, some of the sequences were closer to Sal I, especially in the hypervariable region (HVR) as reflected in the phylogenetic tree.

Four of the nine HVR types found in an earlier study in Sri Lanka (Manamperi, 2002) were also observed in this study. Nine of the 10 HVR types described here were also found in global isolates, and are reported for the first time in this study, where the HVR types T8-T10 resulted due to double interallelic recombination. Interestingly, the phylogenetic tree showed that isolates of HVR-T1 and T2 types are essentially of Asian origin, except for 2 Brazilian isolates of the T2 type, and also that the novel HVR types T11 and T12 were clustered into two clear groups.

Recombination sites were not uniformly distributed across PvMSP-1_{42}, with most recombination events (Rm) occurring at PvMSP-1_{33}, which corroborates previous findings (Putaporntip et al., 2002). These accounts for HVR types T3-T7 and the 19 haplotypes (H71-84) generated solely due to intragenic recombination and further confirm that the extensive polymorphism present at PvMSP-1_{33} is maintained by both interallelic and intragenic recombination. A recent study demonstrated that *P. falciparum* has a spectrum of population structures: strong linkage disequilibrium and low genetic diversity in areas with low levels of transmission; and linkage 'equilibrium' and high diversity in areas with high levels of transmission (Anderson et al., 2000). However, linkage disequilibrium of PvMSP-1_{42} in Sri Lankan isolates showed a significant decline with increasing nucleotide distance and a high level of haplotype diversity ($h = 0.962$). These observations indicate that new PvMSP-1_{42} alleles could be generated by recombination events even in areas with low transmission of malaria.

Haplotypes common to all three study areas may be due to Colombo patients contracting the disease during visits to endemic areas due to occupational commitments (Fonseka and Mendis, 1987; Briët et al., 2003), where the origin of these infections was traced to eleven administrative districts that included the two endemic areas of this study Though the two endemic populations living 250 km apart were at the onset defined as two discrete test populations, the negative F_{ST} values confirmed a gene flow of the parasites between the two populations and hence appear to be a single population.

The phylogenetic tree further confirmed the low genetic differentiation due to migration between Sri Lanka and neighboring India with a majority of common amino acid haplotypes, as well as the marked genetic differentiation between Sri Lanka and Bangladesh shown by the separate clustering of 3 of the 5 Bangladesh isolates. Overall, the phylogenetic tree showed no evidence of geographic clustering of alleles, but rather the interchange of *Plasmodium* strains due to migration of individuals and/or malaria transmitting vectors.

In Sri Lankan isolates, PvMSP-1_{33} was subjected to balancing selection, with frequent occurrences of non-synonymous substitutions relative to synonymous ones, presumably to avoid host

Author's personal copy

immune pressure. These results are consistent with previous studies that used a global sample of *P. vivax* isolates (Pacheco et al., 2007; Thakur et al., 2008). Consequently, these polymorphisms are actively maintained in *P. vivax* populations, and host immune responses likely play a role in maintaining the polymorphism of *P. vivax* MSP-1 alleles. Two possible mechanisms that enable PvMSP-1_{33} balancing selection in the local isolates are: (i) the presence of three of the five positive selection sites essential for immune evasion by the parasite, at amino acid positions 1490 (D/T), 1491 (G/A) and 1494 (T/A/E) in the 33 kDa fragment (Tanabe et al., 2007), and (ii) clusters of non-synonymous substitutions occurring in the predicted T cell epitopes in all 27 PvMSP-1_{33} amino acid haplotypes mediating differential binding of these polymorphic peptides to multiple class II MHC alleles.

Extensive genetic diversity at the PvMSP-1_{33} locus in Sri Lanka where low transmission and unstable malaria conditions persist highlights both interallelic and intragenic recombination as critical factors for the origin of new PvMSP-1_{42} amino acid haplotypes in local *P. vivax* populations. Moreover, balancing selection for polymorphism acting on PvMSP-1_{33} in response to host immune pressure creates new B cell epitopes requiring new antibody responses and provides for broad B and T cell recognition, which may contribute to the "immature" natural antibody response demonstrated against PvMSP-1_{42} in Sri Lanka (Wickramarachchi et al., 2007).

A single amino acid variant at position 1709 (K/E) in PvMSP-1_{19} seen in previous studies (Pasay et al., 1995; Soares et al., 1999; Putaporntip et al., 2000, 2002; Manamperi, 2002), was not observed in any of the 95 Sri Lankan isolates described here. Conversely, with the exception of a single recombination event, PvMSP-1_{19} at the amino acid level was highly conserved in all 95 Sri Lankan isolates, with confirmed lack of T cell epitope variations. Thus, PvMSP-1_{19} is clearly under purifying selection (Pacheco et al., 2007; Thakur et al., 2008), and the protective nature of the natural antibody response directed against the 19 kDa fragment previously characterized by the marked switch of anti-PvMSP-1_{19} IgM to functionally important cytophilic IgG1 and IgG3 antibodies with increasing exposure to malaria in the endemic areas of Sri Lanka (Wickramarachchi et al., 2007), emphasize the potential of an experimental vaccine based on PvMSP-1_{19} to be effective against *P. vivax* infections in Sri Lanka.

In addition, our results further confirmed previous observations that PvMSP-1_{33} is under balancing selection where as PvMSP-1_{19} is under purifying selection (Pacheco et al., 2007; Thakur et al., 2008). In contrast PfMSP-1_{19} and PfMSP-1_{33} are under balancing and purifying selection, respectively (Pacheco et al., 2007). The underlying different roles of this antigen in the host–parasite immune interaction in both *P. vivax* and *P. falciparum* therefore are highlighted (Pacheco et al., 2007; Thakur et al., 2008).

In conclusion, this study for the first time revealed in depth the high level of genetic diversity of PvMSP-1_{42} in Sri Lanka, where mutations, natural selection and recombination seem to fuel and sustain this phenomenon to evade the host immunity. Moreover, the greater diversity of PvMSP-1_{42} generated via natural selection appears to determine the diversity of predicted B and T cell epitopes for this antigen. Under local endemic settings, PvMSP-1_{33} under balancing selection, may generate an "immature" natural antibody response against PvMSP-1_{42} and also interfere with the development of a mature IgG dominated protective antibody response to PvMSP-1_{19} which is under purifying selection. Though the immunological interference between epitopes of PvMSP-1_{33} and PvMSP-1_{19} may complicate vaccine development based on PvMSP-1_{42}, the highly conserved PvMSP-1_{19} fragment alone (100% conserved in all Sri Lankan isolates) with the ensuing "protective" natural antibody response to it, could be considered important as a veritable vaccine candidate against *P. vivax* malaria in Sri Lanka.

Acknowledgements

Financial assistance by the National Science Foundation, Sri Lanka (grant no NSF/RG/2005/HS/06) and the National Research Council of Sri Lanka (grant no NRC-05-34) is acknowledged. The assistance rendered by Drs. Shiroma M. Handunnetti and Thilan Wickramarachchi, and Messers L. Perera and S. Bandara of the Malaria Research Unit, Department of Parasitology, Faculty of Medicine in sample collection (through PVUR's grant No. F/3008-1 from IFS, Sweden), and by Ms Anoma de Silva of the Department of Zoology, Faculty of Science, both of the University of Colombo, is deeply appreciated. A.A. Escalante is supported by the grant R01GM080586 from the US National Institutes of Health.

References

Amaratunga, C., 2004. The investigation in to merozoite surface protein-1 vaccination in a natural simian host-parasite system. Ph.D. dissertation. University of Colombo, Sri Lanka.

Anderson, T.J.C., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bocharie, M., Mokili, J., Mharakurwa, S., French, N., Whitworth, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K., 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol. 17, 1467–1482.

Annual Report of the Anti-malaria Campaign, 2007. Ministry of Healthcare and Nutrition, Sri Lanka.

Briët, O.J.T., Gunawardena, D.M., van der Hoek, W., Amerasinghe, F.P., 2003. Sri Lanka Malaria Maps. Malaria Journal. 2(22), www.malariajournal.com/content/2/1/22.

Bruce, M.C., Galinski, M.R., Barnwell, J.W., Snounou, G., Day, K.P., 1999. Polymorphism at the merozoite surface protein-3α locus of *Plasmodium vivax*: global and local diversity. Am. J. Trop. Med. Hyg. 61, 518–525.

Chen, Q., Schlichtherle, M., Wahlgren, M., 2000. Molecular aspects of severe malaria. Clin. Microbiol. Rev. 13 (3), 439–450.

Cole-Tobian, J., King, C.L., 2003. Diversity and natural selection in *Plasmodium vivax* Duffy binding protein gene. Mol. Biochem. Parasitol. 127, 121–132.

Conway, D.J., Cavanagh, D.R., Tanabe, K., Roper, C., Mikes, Z.S., Sakihama, N., Bojang, K.A., Oduola, A.M.J., Kremsner, P.G., Arnot, D.E., Greenwood, B.M., McBride, J.S., 2000. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nat. Med. 6, 689–692.

Cowman, A.F., Crabb, B.S., 2006. Invasion of red blood cells from malaria parasites. Cell 124, 755–766.

EL-Manzalawy, Y., Dobbs, D., Honavar, V., 2008. Predicting flexible length linear B-cell epitopes. In: 7th International Conference on Computational Systems Bioinformatics, Stanford, CA, USA, pp. 121–132.

Escalante, A., Lal, A.A., Ayala, F.J., 1998. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. Genetics 149, 189–202.

Escalante, A.A., Cornejo, O.E., Rojas, A., Udhayakumar, V., Lal, A.A., 2004. Assessing the effect of natural selection in malaria parasites. Trends Parasitol. 20, 388–395.

Ferreira, M.U., da Silva-Nunes, M., Wunderlich, G., 2004. Antigenic diversity and immune evasion by malaria parasites. Clin. Diagn. Lab. Immunol. 11, 987–995.

Fonseka, J., Mendis, K.N., 1987. A metropolitan hospital in a non-endemic area provides a sampling pool for epidemiological studies on vivax malaria in Sri Lanka. T. Roy. Soc. Trop. Med. H 81, 360–364.

Galinski, M.R., Barnwell, J.W., 2008. *Plasmodium vivax*: who cares? Malar J. 7 (1), doi:10.1186/1475-2875-7-S1-S9.

Gardiner, D.L., McCarthy, J.S., Trenholme, K.R., 2005. Malaria in the post-genomics era: light at the end of the tunnel or just another train? Postgrad. Med. J. 81, 505–509.

Gibson, H.L., Tucker, J.E., Kaslow, D.C., Krettli, A.U., Collins, W.E., Kiefer, M.C., Bathurst, I.C., Barr, P.J., 1992. Structure and expression of the gene for Pv200, a major blood stage surface antigen of *Plasmodium vivax*. Mol. Biochem. Parasitol. 50, 325–334.

Gunasekera, A.M., Wickramarachchi, T., Neafsy, D.E., Ganguli, I., Perera, L., Premaratne, P.H., Hartl, D., Handunnetti, S.M., Udagama-Randeniya, P.V., Wirth, D.F., 2007. Genetic diversity and selection at the *Plasmodium vivax* Apical Membrane Antigen-1 (*Pvama-1*) locus in a Sri Lankan population. Mol. Biol. Evol. 24 (4), 939–947.

Holder, A., 2009. Malaria vaccines: where next? Plos Pathogens 5 (10), e1000638, doi:10.1371/journal.ppat.1000638.

Konradsen, F., Amerasinghe, F.P., Van der Hoek, W., Amerasinghe, P.H., 2000. Malaria in Sri Lanka: Current Knowledge on Transmission and Control. IWMI, Colombo, Sri Lanka.

Longacre, S., 1995. The *Plasmodium cynomolgi* merozoite surface protein-1 C-terminal sequence and its homologies with other *Plasmodium* species. Mol. Biochem. Parasitol. 74 (1), 105–111.

Manamperi, A., 2002. Preliminary preparation towards malaria vaccine trials in Sri Lanka; studies on genetic polymorphism and immune responses against the C-terminal merozoite surface protein-1. Ph.D. Thesis, University of Colombo, Sri Lanka.

Author's personal copy

Mendis, C., Gamage-nendis, A.C., de Zoyza, A.P.K., Abhayawardena, T.A., Carter, R., Herath, P.R.J., Mendis, K.N., 1990. Characteristics of malaria transmission in Kataragama, Sri Lanka: a focus for immuno-epidemiological studies. Am. J. Trop. Med. Hyg. 42 (4), 298–308.

Pacheco, M.A., Poe, A.C., Collins, W.E., Lal, A.A., Tanabe, K., Kariuki, S.K., Udhayakumar, V., Escalante, A.A., 2007. A comparative study of the genetic diversity of the 42 kDa fragment of the merozoite surface protein-1 in *Plasmodium falciparum* and *P. vivax*. Infect. Genet. Evol. 7, 180–187.

Pasay, M.C., Cheng, Q., Rzepczyk, C., Saul, A., 1995. Dimorphism of the C-terminus of the *Plasmodium vivax* merozoite surface protein-1. Mol. Biochem. Parasitol. 70, 217–219.

Perera, K.L.R.L., Handunnetti, S.M., Holm, I., Longacre, S., Mendis, K., 1998. Baculovirus merozoite surface protein 1 C-terminal recombinant antigens are highly protective in a natural primate model for human *Plasmodium vivax* malaria. Infect. Immun. 66 (4), 1500–1506.

Price, R.N., Tjitra, E., Guerra, C.A., Yeung, S., White, N.J., Anstey, N.M., 2007. Vivax malaria: neglected and not benign. Am. J. Trop. Med. Hyg. 77, 79–88.

Putaporntip, C., Jongwutiwes, S., Seethamchai, S., Kanbara, H., Tanabe, K., 2000. Intragenic recombination in the 3′ portion of the merozoite surface protein-1 gene of *Plasmodium vivax*. Mol. Biochem. Parasitol. 109, 111–119.

Putaporntip, C., Jongwutiwes, S., Sakihama, N., Ferreira, M.U., Kho, W.G., Kaneko, A., Kanbara, H., Hattori, T., Tanabe, K., 2002. Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. Proc. Natl. Acad. Sci. U.S.A. 99, 16348–16353.

Putaporntip, C., Jongwutiwes, S., Iwasaki, T., Kanbara, H., Hughes, A.L., 2006. Ancient common ancestry of the merozoite surface protein-1 of *Plasmodium vivax* as inferred from its homologue in *Plasmodium knowlesi*. Mol. Biochem. Parasitol. 146, 105–108.

Rozas, J., Sánchez-delbarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP 4.50, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Sawai, H., Otani, H., Arisue, N., Palacpac, N., Martins, L., Pathirana, S., Handunnetti, S., Kawai, S., Kishino, H., Horii, T., Tanabe, K., 2010. Lineage-specific positive selection at the merozoite surface protein-1 (msp-1) locus of *Plasmodium vivax* and related simian malaria parasites. BMC Evol. Biol. 10(52), www.biomedcentral.com/1471-2148/10/52.

Singh, H., Raghava, G.P.S., 2001. ProPred: prediction of HLA-DR binding sites. Bioinformatics 17 (12), 1236–1237.

Soares, I.S., Barnwell, J.W., Ferreira, M.U., Gomes Da Cunha, M., Laurino, J.P., Laurino, J.P., Castilho, B.A., Rodrigues, M.M., 1999. A *Plasmodium vivax* vaccine candidate displays limited allele polymorphism, which does not restrict recognition by antibodies. Mol. Med. 5, 459–470.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., Sette, A., 1998. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 160, 3363–3373.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.1. Mol. Biol. Evol. 24, 1596–1599.

Tanabe, K., Mackay, M., Goman, M., Scaife, J.G., 1987. Allelic dimorphism in a surface gene of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 195, 273–287.

Tanabe, K., Escalante, A., Sakihama, N., Honda, M., Arisue, N., Horii, T., Culleton, R., Hayakawa, T., Hashimoto, T., Longacre, S., Pathirana, S., Handunnetti, S., Kishino, H., 2007. Recent independent evolution of MSP-1 polymorphism in *Plasmodium vivax* and related simian malaria parasites. Mol. Biochem. Parasitol. 156, 74–79.

Thakur, A., Alam, M.T., Sharma, Y.D., 2008. Genetic diversity in the C-terminal 42kDA region of merozoite surface protein-1 of *Plasmodium vivax* (*Pvmsp-1$_{42}$*) among Indian isolates. Acta Trop. 108, 58–63.

Wickramarachchi, T., Illeperuma, R.J., Perera, L., Bandara, S., Holm, I., Longacre, S., Handunnetti, S.M., Udagama-Randeniya, P.V., 2007. Comparison of naturally acquired antibody responses against the C-terminal processing products of *Plasmodium vivax* Merozoite Surface Protein-1 under low transmission and unstable malaria conditions in Sri Lanka. Int. J. Parasitol. 37, 199–208.

Wickramarachchi, T., Premaratne, P.H., Dias, S., Handunnetti, S.M., Udagama-Randeniya, P.V., 2010. Genetic complexity of *Plasmodium vivax* infections in Sri Lanka, as reflected at the merozoite-surface-protein-3α locus. Ann. Trop. Med. Parasitol. 104 (2), 95–108.

World Malaria Report, 2008. World Health Organization, Geneva, Switzerland.