High polymorphism in *Plasmodium vivax* merozoite surface protein-5 (MSP5)

A. GOMEZ[1,3], C. F. SUAREZ[2], P. MARTINEZ[1], C. SARAVIA[1] *and* M. A. PATARROYO[1,3]*

[1] *Molecular Biology Department, Fundacion Instituto de Inmunologia de Colombia, Carrera 50#26-00, Bogota, Colombia*
[2] *Biomathematics Department, Fundacion Instituto de Inmunologia de Colombia, Carrera 50#26-00, Bogota, Colombia*
[3] *Universidad Nacional de Colombia, Carrera 30, Calle 45, Bogota, Colombia*

(*Received 29 March 2006; revised 16 June 2006; accepted 21 June 2006; first published online 18 October 2006*)

SUMMARY

A key issue relating to developing multi-component anti-malarial vaccines, lies in studying *Plasmodium vivax* surface proteins' genetic variation. The present work was aimed at amplifying, cloning and sequencing the gene encoding *P. vivax* merozoite surface protein 5 (*Pv*MSP5) in samples obtained from infected patients from Colombian areas having varying malaria transmission rates.† Our results have revealed that *Pv*MSP5 is one of the *P. vivax* surface proteins having greater polymorphism, this being restricted to specific protein regions. The intron and exon II (which includes the GPI anchor and EGF-like domain) were both highly conserved when compared to exon I; exon I displayed the greatest variation and most of the recombination events occurred within it. No geographical grouping was observed. The Nei-Gojobori test revealed significant positive selection in the samples analysed here, whereas Tajima and Fu and Li tests presented a neutral selection pattern. The results reflected a localized variation pattern, recombination between *Pv*MSP5 alleles and also functional and immune pressures, where stronger selective forces might be acting on exon I than on exon II, suggesting that the latter could be an important region to be included in an anti-malarial vaccine.

Key words: antigenic diversity, merozoite surface protein 5 (MSP5), *Plasmodium vivax*, polymorphism, selection.

INTRODUCTION

According to the World Health Organization (WHO), 1·1 to 2·7 million deaths and more than 500 million clinical cases are caused by malaria per year around the world (WHO, 2002; Snow *et al*. 2005). Although most of the deaths are caused by *P. falciparum*, one of the most prevalent and geographically widespread parasite species producing this disease is *Plasmodium vivax*, undoubtedly representing a major public health problem in many countries in South America, Central America, Asia and the south-western Pacific (Polley *et al*. 2004). The development of an effective vaccine against malaria has been mainly focused on *Plasmodium falciparum*; nevertheless, a vaccine is also required for controlling *P. vivax* and has been greatly hampered by many factors such as drug resistance, the lack of continuous culture of this parasite, its complex life-cycle and the antigenic variation which it displays (Good, 2001; Carvalho *et al*. 2002).

Merozoites represent the invasive form of the malarial blood-stage life-cycle, exposing the merozoite surface protein (MSP) complex on their surface that is involved in initial attachment to the erythrocyte surface by means of specific molecular interactions. The merozoite stage is one of the few stages where the parasite lies outside the host cell, making it a visible target for antibodies; a lot of work in malaria vaccine development has thus been directed towards producing antibodies able to inhibit merozoite attachment to, entrance or development within the erythrocyte (Carvalho *et al*. 2002; Gaur *et al*. 2004).

Several studies have shown that MSPs are targets for recognition by a host's immune system and they accumulate polymorphism as a means of evading a host's defences. These proteins are highly polymorphic compared to those expressed during the sexual stage or inside the parasite (Escalante *et al*. 1998). Elucidating the polymorphism-generating mechanism is a central issue for developing effective anti-malarial vaccines and understanding the parasite's population structure (Cui *et al*. 2003*a*).

The *Plasmodium vivax* merozoite surface protein 5 (*Pv*MSP5) is an asexual-stage vaccine candidate; the encoding gene is located downstream of the adenylosuccinate lyase (ASL) gene near MSP4, having 862 bp and 299 bp exons, separated by a 401 bp intron containing 5 tandem repeats of the TGTGATGAC(T/G)GCTCCAC(A/G)TGGCC-GCGAAA(A/C)(A/G) sequence. *Pv*MSP5 consists of 387 amino acids having a 41 kDa calculated

* Corresponding author: Fundacion Instituto de Inmunologia de Colombia, Carrera 50#26-00, Bogota, Colombia. Tel: +57(1)3244672 Ext 141. Fax: +57(1) 4815269. E-mail: mapatarr@fidic.org.co
† Nucleotide sequence data reported in this paper are available in the GenBank, EMBL and DDBJ databases under Accessions numbers DQ341586 to DQ341601.

 © 2006 Cambridge University Press
doi:10.1017/S0031182006001168 Printed in the United Kingdom

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

molecular mass; it is anchored to the membrane through a GPI (glycosyl-phosphatidylinositol) motif located close to the C-terminus. It also contains an EGF-like (epidermal growth factor-like) domain just upstream of the putative GPI attachment site. Different to other malarial surface antigens, *Pv*MSP5 does not have repeat amino acid motifs and, rather than being localized all along the merozoite surface, this protein has been found to be associated with the apical organelles or located over the apical polar prominence (Black *et al*. 2002).

Recent studies have shown that *Pv*MSP5 homologue proteins in *P. yoelii* (*Py*MSP4/5) (Kedzierski *et al*. 2000) or *P. chabaudi* (*Pch*MSP4/5) (Rainczuk *et al*. 2004) are able to protect vaccinated mice against experimental challenge with the parasite species from which the proteins were derived. In the former study, protective efficacy was enhanced when *Py*MSP4/5 was delivered in conjunction with $PyMSP1_{19}$ (Kedzierski *et al*. 2002; Wang *et al*. 2004). The present paper is the first to describe analysis of polymorphism in *Plasmodium vivax* MSP5 from Colombian isolates as a contribution towards the development of a multicomponent vaccine against *P. vivax* malaria.

MATERIALS AND METHODS

P. vivax *samples*

The blood samples from which parasite DNA was obtained were collected from regions in eastern and western Colombia: Norte de Santander (Cúcuta; $n=10$), Nariño (Tumaco; $n=4$), Llanos Orientales (Villavicencio; $n=8$). Twenty-two samples were extracted from patients diagnosed as having *P. vivax* malaria by microscopic analysis (with patients' informed consent), anti-coagulated with EDTA, processed and stored from 2001 until 2004. DNA was obtained with the Wizard Genomic DNA Purification kit (Promega, Madison WI) following manufacturer's instructions for use as a template in PCR.

PCR and fragment isolation

The *Pv*MSP5F (5′-CGCGTCGTGTTAGCTATC-3′) and *Pv*MSP5R (5′-CAACACAGCCAGAGCAAAG-3′) primers were used for amplifying the *Pv*MSP5 gene from genomic DNA, based on the GenBank Accession AF403476 sequence. PCR was performed with 10 mM Tris, 50 mM KCl, 3 mM $MgCl_2$, 20 mM of each dNTP, 1 pmol of each primer, 0·25 U of *Taq* polymerase and 10–40 ng of DNA template in a final volume of 25 μl. The thermal profile was as follows: 1 cycle of 5 min at 95 °C and 30 cycles of 1 min at 95 °C, 30 sec at 52·5 °C and 1 min at 72 °C and a final 5 min extension cycle at 72 °C.

PCR products were purified with Wizard PCR Preps Systems (Promega, Madison WI) and cloned into a pGEM-T easy vector system (Promega, Madison WI); positive clones were selected using an ampicillin positive selection system and α-complementation of *lacZ* gene. Plasmid DNA was obtained by Wizard Plus Minipreps DNA Purification System kit (Promega, Madison WI). DNA sequencing was performed in both directions with the BigDye Terminator v3.1 Cycle Sequencing kit in an ABI Prism 310 Automated Sequencer (Applied Biosystems) following manufacturer's instructions. T7, SP6 and internal *Pv*MSP5 IF (5′-GAC GGC AAT GAT GAG AAC-3′) and *Pv*MSP5 IR (5′-TGG AGC AGT CAT CAC AGT-3′) primers were used due to the length of the sequence. Two clones were sequenced per isolate; each one came from an independent PCR amplification. The polymorphism observed was shared by more than one isolate, this being strong evidence that they are valid polymorphisms and not the result of PCR error. Two extra clones were sequenced for those isolates which displayed different sequences for both clones previously sequenced.

Sequence analysis

Colombian *P. vivax* MSP5 sequences were compared and analysed with the previously described Thai-NYU (GenBank Accession number AF403476) (Black *et al*. 2002) and Sal-I (Tentative annotation number Q8T6V0) strain sequences.

Sequences were checked and assembled using Sequence Navigator v1.0.1 (Applied-Biosystems, 1993). Preliminary alignment and comparison of sequences was performed using Clustal X software (Thompson *et al*. 1997). GeneDoc software was used for editing sequences and establishing the identity amongst them (Nicholas *et al*. 1997).

MEGA v2.1 software was employed for examining natural selection. The Nei and Gojobori method with the Jukes and Cantor correction was used for comparing the difference between non-synonymous and synonymous substitutions (positive values for this difference indicate positive selection and negative values indicate negative selection) (Nei and Gojobori, 1986). Statistical significance of the observed differences was estimated using a *t*-test. Nucleotide diversity (π) as well as Tajima's test were calculated using the DnaSP v4 software (Rozas *et al*. 2003). The latter test is based on neutral model prediction, proposing the difference (statistic D) between the number of segregating sites (S), and the average number of nucleotide differences between pairs of random sequences (Π), considering that θ can be estimated from S and Π. Parameter θ is related to the expected amount of genetic variation per nucleotide. A positive D value indicates possible balancing selection (reflecting the presence of large numbers of medium frequency polymorphisms) or population subdivision. A negative value suggests

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

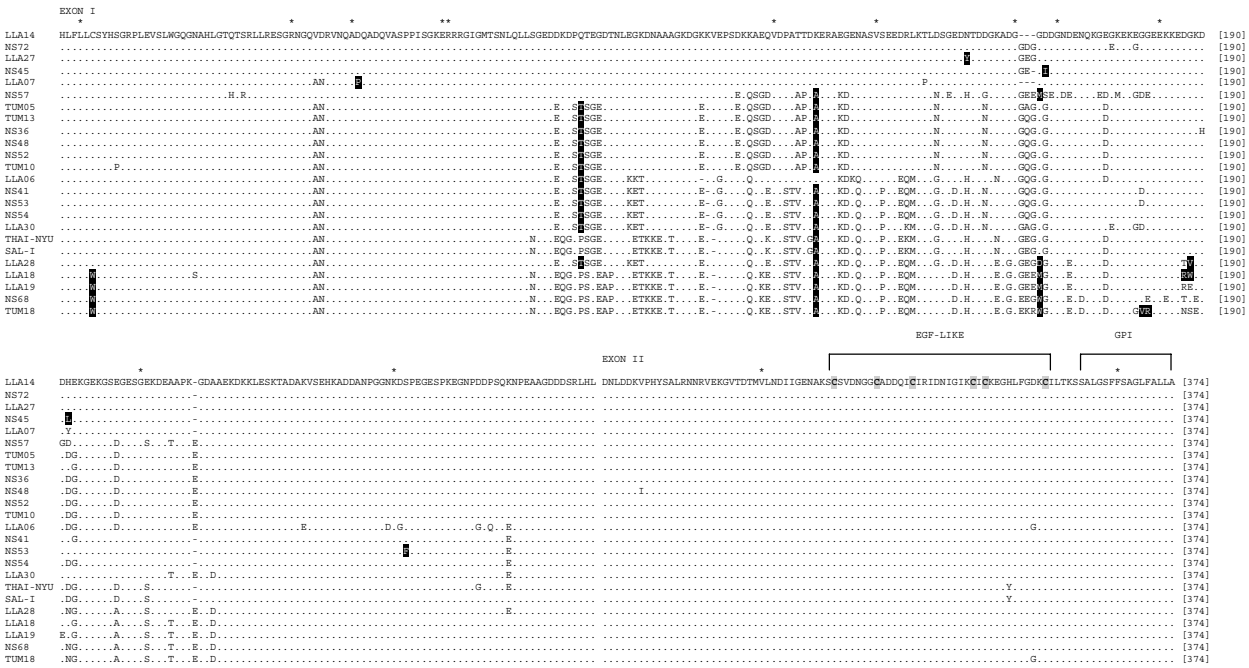


Fig. 1. Alignment of deduced amino acid sequences of *Pv*MSP5 comparing the Thai-NYU and Sal-I strains with Colombian isolates. An asterisk (*) above the alignment indicates those positions where synonymous mutations occurred. White letters on a black background represent non-conserved amino acid replacements; black letters on a grey background represent the six conserved cysteine residues. Dots indicate conserved residues and dashes represent gaps introduced for alignment. Lla = Llanos Orientales, NS = Norte de Santander and Tum = Tumaco.

a high number of low-frequency polymorphisms, indicating a growth in population size, a recent selective sweep, or purifying selection (Tajima, 1989). Fu and Li D* and F* tests were also performed; these reflect the same trends as Tajima and estimate θ from the number of mutations in external and internal phylogeny branches. The number of mutations in external phylogeny branches was estimated from singleton sites (S) and the total number of mutations giving the D* index, or between S and Π giving the F* index (Fu and Li, 1993). An excess of intermediate frequency polymorphism and a lack of rare substitutions (singletons) result in positive values for D* and F*. The interpretation of D* and F* values were the same as the Tajima test (described above). If non-significant values distinct from zero are found in both tests, the null hypothesis is accepted and neutral selection is thus occurring. A sliding window was used for Tajima's *D* and Fu and Li's values to examine whether selective forces might be acting in different directions in *Pv*MSP5.

Other parameters considered were the number of variant sites (VS) and, for these positions, the number of parsimony informative sites (PI) and singleton sites (S). These parameters were calculated with MEGA 2.1 software (Kumar *et al*. 2001). The number of variant positions with conservative (C) and non-conservative changes (NC) were also considered for amino acid sequences. The nature of amino acid replacement (conserved or non-conserved) was examined by using the PAM250 similarity matrix (Dayhoff, 1978). Zero or positive values indicated that variable amino acid residues could be functionally interchanged (conservative changes) whilst negative values indicated that it was not common for 2 amino acids to become replaced (non-conservative changes). The minimum number of recombination events test (Rm) (Hudson and Kaplan, 1985), included in DnaSP v4 software (Rozas *et al*. 2003), as well as the LDhat package (McVean *et al*. 2002) which implements a coalescence-based method for detecting recombination, were used when searching for evidence of recombination.

Phylogenetic trees were constructed using MEGA 2.1 software (Kumar *et al*. 2001) for parsimony (Fitch, 1971) and distance methods such as neighbour-joining and minimum-evolution (Rzhetsky and Nei, 1993). Genetic distances were estimated by using P and Tamura-3 parameters for nucleotides and P and Poisson-corrected distances for amino acid sequences. Bootstrap analysis (Hillis, 1993) and internal branch test (IB) (Sitnikova, 1996) (with 1000 replicates each) were used for assigning confidence levels to branch nodes.

RESULTS

Sixteen new Pv*MSP5 sequences were obtained*

Twenty-two Colombian isolates were analysed; 16 different sequences were obtained. Just one type of sequence was found for each isolate. Figure 1

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

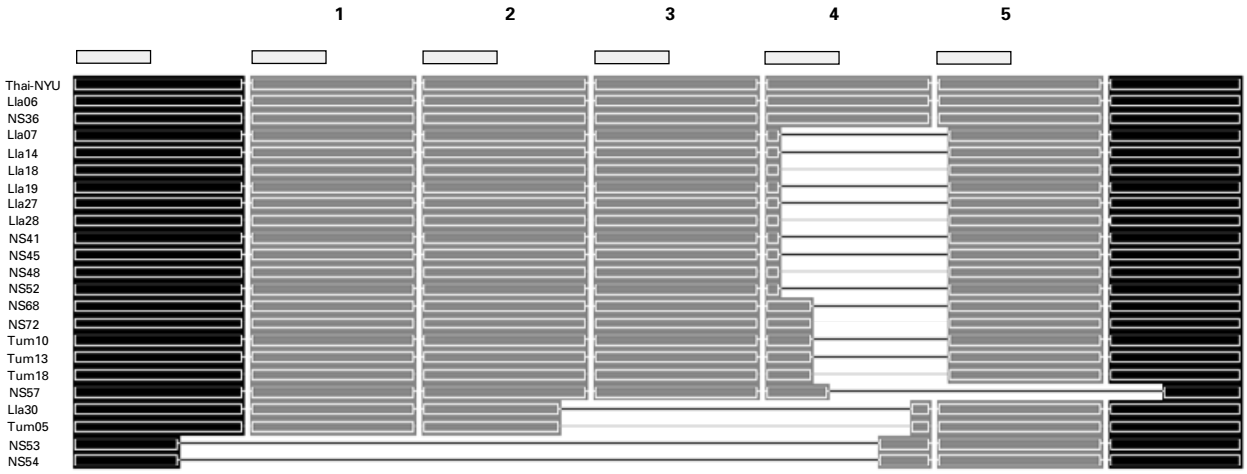


Fig. 2. *Pv*MSP5 intron repeat sequences. The 5 repeats (shown in grey) were virtually identical and are numbered at the top of the figure. The deletions are indicated by lines in the white spaces. The boxes at the top denote a shorter repeat motif, shared by 6 intron segments.

shows the amino acid alignment for *Pv*MSP5 from the Colombian sequences obtained. There were no great differences in composition between exons and introns (48% AT *vs*. 47% AT) or a marked bias towards AT content in *Pv*MSP5. Therefore, it was not necessary to make AT bias corrections, as opposed to what occurs with other *Plasmodium* proteins and species (Escalante *et al*. 1998).

In the Colombian sequences analysed here, 89–100% identity was obtained for exon I, 99–100% for exon II and 57–100% for the intron. In the case of the intron, its identity did not reflect the repeat motif's high level of conservation, meaning that the polymorphism observed corresponded to deletions of this motif or fragments of it (Fig. 2). The region having the greatest variation was localized between nucleotides 236 and 674 (exon I) where the greatest number of non-synonymous substitutions were observed. These changes were evenly distributed between the codon's first and second positions. Most of the changes observed were conservative, suggesting that these changes tended to maintain the protein's structural or functional characteristics. No preferential variant positions were observed according to geographical site.

The GPI as well as the EGF-like motifs within exon II were much more conserved compared to the rest of the protein (Fig. 1). In the case of the EGF-like motif it was observed that residues probably related to structural functions (such as the cysteines) were totally conserved in all sequences (Fig. 1).

Pv*MSP5 cluster analysis*

The topology obtained was similar for amino acids and nucleotides, independent of the method employed (Parsimony, NJ, ME), as well as when the intron was included for analysis. The trees constructed for exon II did not present statistically significant values due to the lower levels of parsimony informative sites in this region (data not shown). The topology obtained for exon I was identical to overall topology. Figure 3 shows sequence topology (in nucleotides) obtained for exons I and II from *Pv*MSP5 by the minimum evolution method with P distance. The results obtained revealed no clear geographical grouping; on the contrary, they showed clusters where sequences from different geographical regions were mixed (this was supported by statistically significant values). This phenomenon has also been observed in Thailand and Brazil (Jongwutiwes *et al*. 2002).

The Colombian sequences obtained, as well as those from the Thai-NYU and Sal-I (Tentative Annotation Number: Q8T6V0) strains, were used for cluster analysis. Four groups of highly related sequences could thus be distinguished: G1 (87–99% identity), G2 (95–99% identity), G3 (91–99% identity) and G4 (97–99% identity). Amongst those sequences that could not be assigned to any of the above-mentioned groups, 5 of them (NS57, Lla06, Thai-NYU, Sal-I, Lla28 and Lla30) showed clear mixtures of sequence segments between the different groups (Fig. 3 and recombination analysis).

Intron repeat variation

The *Pv*MSP5 intron exhibited high conservation in the repeat motif, agreeing with previous reports regarding its homologue in *P. falciparum* (Wu *et al*. 1999; Jongwutiwes *et al*. 2002). This intron includes 5 repeats in *P. vivax* (401 bp maximum length) (shown in Fig. 2 in grey) and several partial repeats localized at the 5′ region (Fig. 2, shown in rectangles). The variation observed was more related to the total or partial loss of repeats (24 bp to 126 bp) rather than to sequence changes. There were losses

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

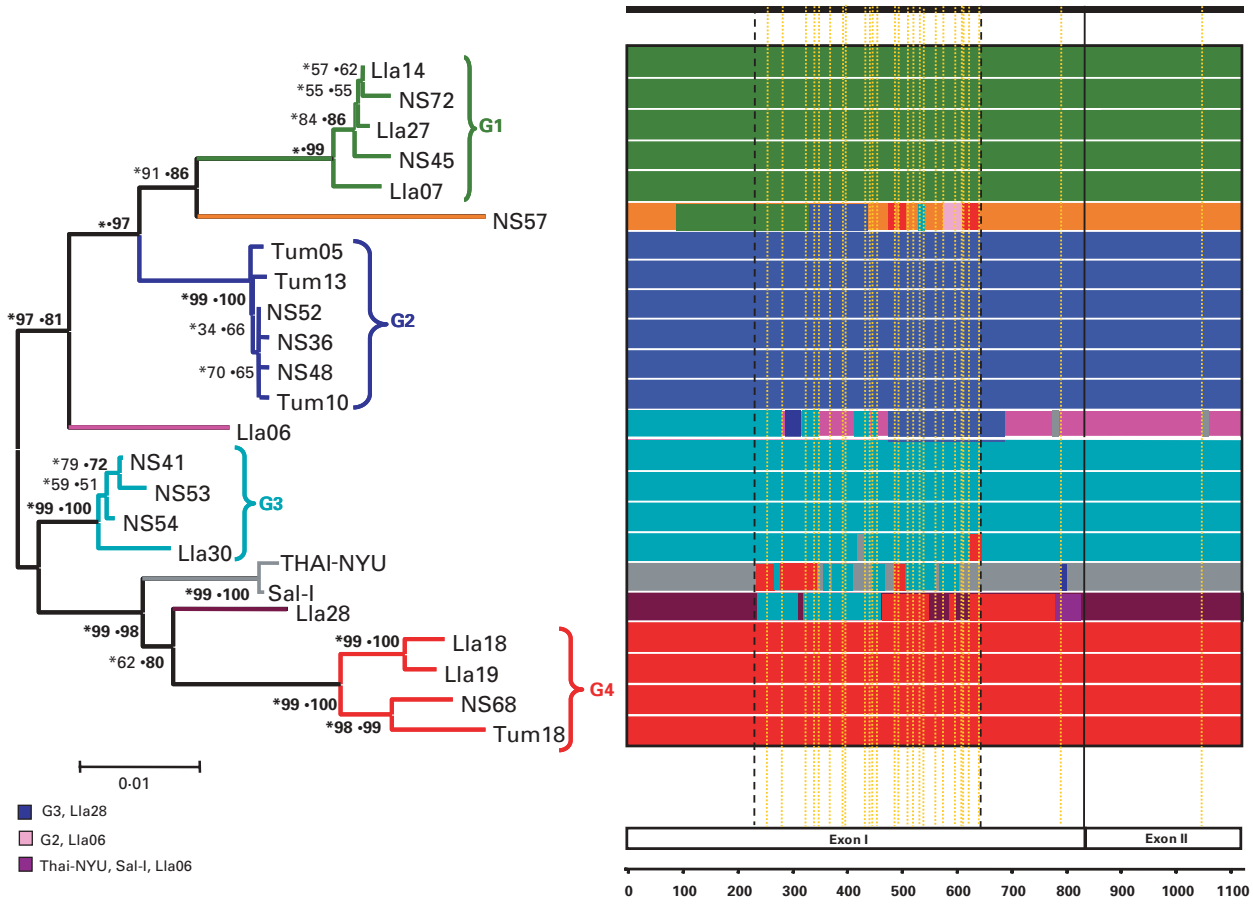


Fig. 3. Cluster analysis and schematic representation of the recombination segments in *Pv*MSP5. The left-hand side of the diagram shows a phylogenetic tree obtained by minimum evolution for *Pv*MSP5 exon nucleotide sequences, based on pair-wise differences using P distance. Numbers represent the branches' interior branch (*) and bootstrap (•) values with 1000 replicates each. The scale indicates 0·01 substitutions per site. Most of the nodes were supported by statistically significant IB (interior branch) (>95%) (Sitnikova, 1996) and bootstrap (>70%) values (Hillis, 1993). Significant values are shown in bold. The Thai-NYU and Sal-I sequences have been included in the analysis. NS = Norte de Santander, Lla = Llanos Orientales and Tum = Tumaco. The phylogenetic groups are represented by a characteristic branch colour for each of them. The *Pv*MSP5 exon division is shown at the bottom right-hand side of the diagram. A representation of conservative and variable regions is shown at the top right-hand side of the diagram by black and light grey bars (respectively); yellow lines indicate the recombination sites in DnaSP v4. The black dotted lines delimit the high diversity zone. Recombination fragments are represented within the schematic sequence by the branch colour assigned to the locus donating the tract.

towards the 3′ extreme which were not correlated to the isolates' geographical distribution. This variation was possibly due to the action of some DNA polymerase slippage mechanism, as occurs in P*f*MSP5 (Wu *et al*. 1999).

It is clear that the substitution pattern differs between exons and the intron, highlighting the possible role of selection in encoding and non-encoding sequences where little polymorphism has been observed in introns but extensive polymorphism in exons for *Plasmodium* (Hughes and Verra, 2001; Volkman *et al*. 2001).

Diversity analysis

Table 1 presents the calculations for the diversity parameters used. Exon I presented greater nucleotide diversity ($\pi = 0{\cdot}04983$) and a greater number of parsimonious sites compared to exon II which presented little variation ($\pi = 0{\cdot}00212$), preferentially exhibiting singleton sites. The great variation in exon I suggests that this sector of the gene has been under diversifying pressure compared to exon II which seemed to have highly restricted variation. The same trend was observed when geographical regions were analysed.

The diversity obtained for the groups was low (represented mostly by singleton changes) (Table 1), compared to that obtained by geographical area. Group 4 showed the greatest diversity ($\pi = 0{\cdot}01653$), displaying only parsimonious sites.

The nucleotide diversity obtained for *Pv*MSP5 was considerable. This has also been observed for *P. vivax* in low transmission rate countries with

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

Table 1. Diversity, selection and neutrality tests for *Pv*MSP5 coding regions

(% ident = % identity between sequences. Nei-Gojobori test: Sites = number of sites included, d*S* = synonymous substitution and d*N* = non-synonymous substitution. d*N*−d*S* (+) represents positive selection pressure and (−) represents negative selection pressure; significant values according to *t*-test are shown in bold. Diversity analysis: Sites = number of sites included, VS = number of variant positions, S = number of singleton sites, PI = number of parsimony informative sites and π = average number of nucleotide substitutions. Neutrality test for Tajima, Fu and Li: positive values correspond to positive selection pressure and negative values correspond to negative selection pressure. * Neutrality test for Tumaco has been omitted due to sample size.)

Origin	Region	% Iden	Sites	d*S*	d*N*	d*N*−d*S*	Sites	VS	S	P.I.	π	Tajima D	Fu and Li D*	Fu and Li F*
Llanos	Exon I	89–99%	279	0·0503	0·0635	+	838	108	18	90	0·05648	1·004597	0·38414	0·42576
	Exon II	99–100%	95	0·0000	0·0011	+	284	2	1	1	0·00088	−1·054819	−1·12639	−1·20353
	All	92–97%	374	0·0363	0·0472	+	1122	110	19	91	0·04214	0·97609	0·36204	0·40158
Santander	Exon I	89–100%	279	0·0311	0·0559	**+**	838	114	14	100	0·04829	0·12828	−0·19357	−0·17823
	Exon II	99–100%	95	0·0072	0·0009	−	284	3	2	1	0·00235	−0·18393	−0·28020	−0·28650
	All	83–99%	374	0·0245	0·0416	**+**	1122	118	14	101	0·03654	0·12097	−0·19809	−0·18320
Tumaco	Exon I	91–99%	279	0·0383	0·0488	+	838	79	7	72	0·04391	—*	—*	—*
	Exon II	99–100%	95	0·0077	0·0023	−	284	3	2	1	0·00352	—*	—*	—*
	All	91–99%	374	0·0297	0·0367	+	1122	82	9	73	0·03366	—*	—*	—*
Colombia	Exon I	89–99%	279	0·0385	0·0582	**+**	838	138	36	102	0·04983	0·64584	−0·35021	−0·25869
	Exon II	99–100%	95	0·0052	0·0012	−	284	5	3	2	0·00212	−1·26827	−0·81047	−1·08394
	All	87–99%	374	0·0292	0·0434	**+**	1122	144	40	104	0·03752	0·57112	−0·37717	−0·29889
Group 1	Exon I	97–99%	279	0·0023	0·0078	+	838	14	14	0	0·00630	−1·21039	−1·21039	−1·28541
	Exon II	99–100%	95	0·0062	0·0000	−	284	1	1	0	0·00141	−0·8165	−0·8165	−0·77152
	All	87–99%	374	0·0033	0·0058	+	1122	15	15	0	0·00505	−1·21472	−1·21472	−1·29226
Group 2	Exon I	99%	279	0·0000	0·0025	**+**	838	5	5	0	0·00199	−1·33698	−1·36825	−1·45012
	Exon II	99–100%	95	0·0510	0·0015	−	284	2	2	0	0·00235	−1·13197	−1·15529	−1·19511
	All	95–99%	374	0·0014	0·0023	+	1122	7	7	0	0·00208	−1·39031	−1·42394	−1·52011
Group 3	Exon I	98–99%	279	0·0028	0·0095	+	838	13	12	1	0·00801	−0·60528	−0·60528	−0·6148
	Exon II	100%	95	0·0000	0·0000		284	0	0	0	0·00000	0·00000	0·00000	0·00000
	All	91–99%	374	0·0021	0·0071	**+**	1122	13	12	1	0·00597	−0·60528	−0·60528	−0·6148
Group 4	Exon I	97–99%	279	0·0285	0·0194	−	838	10	0	10	0·02096	1·09849	0·09587	0·02752
	Exon II	99–100%	95	0·0077	0·0023	−	284	0	0	0	0·00352	−0·70990	−0·70990	−0·60427
	All	97–99%	374	0·0228	0·0151	−	1122	10	0	10	0·01653	0·97316	0·04112	−0·02601

Table 2. Nucleotide diversity for *Plasmodium vivax* and *P. falciparum* antigens

(*n* = number of isolates, Sites = number of sites included and π = nucleotide diversity.)

	Antigen	Origin	n	Sites	π	Reference
P. falciparum	EBA-175	—	14	1842	0,00366	Escalante *et al*. 1998
	AMA1	—	9	1866	0,01635	Escalante *et al*. 1998
	MSP1	—	40	1056	0,08792	Escalante *et al*. 1998
	MSP2	—	30	417	0,04409	Escalante *et al*. 1998
	MSP3	Asia	19	575	0,09694	Escalante *et al*. 1998
	MSP4	PNG	42	965	0,0031	Benet *et al*. 2004
	MSP5	Senegal, Cambodia	71	955	0,00036	Polson *et al*. 2005
P. vivax	DBP	Colombia	23	1388	0,01224	Martinez *et al*. 2004
	AMA1	Africa, Asia	219	461	0,0174	Figtree *et al*. 2000
	TRAP	Thai, Brazil	39	1470	0,00594	Putaporntip *et al*. 2001
	MSP1	Africa,Asia	175	417	0,0451	Figtree *et al*. 2000
	MSP3a	Thai	35	2577	0,04962	Mascorro *et al*. 2005
	MSP4	Colombia	30	846	0,00112	Martinez *et al*. 2005
	MSP5	Colombia	22	1524	0,03752	This paper

proteins such as *Pv*CSP, *Pv*MSP3-α (Cui *et al*. 2003*b*) and *Pv*TRAP (Putaporntip *et al*., 2001). A low transmission rate in *P. vivax* is thus not directly related to low nucleotide diversity for a particular protein, as has been observed in *P. falciparum* (Felger *et al*. 1994; Babiker *et al*. 1997).

Table 2 presents a comparison of nucleotide diversity values obtained for different *P. falciparum* and *P. vivax* proteins. *Pv*MSP5 showed 10-fold greater nucleotide diversity ($\pi = 0{\cdot}03752$) compared to its homologue in *P. falciparum* ($\pi = 0{\cdot}00036$) (Polson *et al*. 2005). There was greater difference in exon I (71 times greater, $\pi = 0{\cdot}04983$ for *Pv*MSP5 vs. $\pi = 0{\cdot}0007$ for *Pf*MSP5) than in exon II (1·7 times greater for *Pv*MSP5 $\pi = 0{\cdot}00212$ vs. $\pi = 0{\cdot}0012$ for *Pf*MSP5) (Jongwutiwes *et al*. 2002; Polson *et al*. 2005). There was no variation for MSP5 at intron level in both species.

Comparing *Pv*MSP5 to other surface proteins revealed that this was the third most diverse in *P. vivax* following MSP3 and MSP1 (Table 2). When comparing it to *P. falciparum* proteins, a clear difference was seen in nucleotide diversity regarding other falciparum surface proteins where it was the least diverse, even though other members of *Pf*MSP were as diverse as those in *P. vivax* (Table 2). This suggests differences at the functional level revealing the degree of differential restriction on this protein in both species.

Recombination analysis

A minimum of 16 observable recombination (Rm) events were detected with DnaSP v4 software between nucleotides (266,287), (329,352), (377,397), (409,437), (437,444), (451,460), (490,504), (522,536) (536,542), (542,565), (578,600), (600,612), (615,625), (625,647), (647,793) in exon I, coinciding with this protein's region having the greatest variation (Figs 3 and 4). One observable recombination event was also found between nucleotides (793,1454), corresponding to the limits between exon I and exon II, respectively.

Less observable recombination events were obtained for *Pv*MSP5 compared to proteins such as *Pv*MSP1 (Rm = 37) (Putaporntip *et al*. 2002); *Pv*MSP3α (Rm = 19–20) (Mascorro *et al*. 2005; Rayner *et al*., 2002), similar to *Pv*Duffy (Rm = 16) (Martinez *et al*. 2004) and greater than in *Pv*TRAP (Rm = 8) (Putaporntip *et al*. 2001).

Figure 3 gives a schematic representation of the interchange of segments observed between sequences (as detected by alignment inspection) where recombination fragments were detected between NS57, Lla06, Thai-NYU, Sal-I, Lla28 and Lla30 sequences and clusters of sequences. Thus, recombination between *Pv*MSP5 alleles seems to be an important source of diversity.

In addition, we searched for recombination events by examining the relationship between nucleotide distance and linkage disequilibrium using |D′| and r^2 statistics. Recombination should appear as a significant negative correlation between distance and either statistic. The permutation test implemented in the PAIRWISE software from the LDhat package (McVean *et al*. 2002) was used to assess the significance between |D′| or r^2 and nucleotide distance. Results obtained showed a significant negative relationship (corr r^2, $d = -0{\cdot}22675$, $P_{r2} = 0{\cdot}0000$; corr |D′|, $d = -0{\cdot}11847$, $P_{|D'|} = 0{\cdot}0230$) suggesting recombination in *Pv*MSP5.

Another analysis to determine the relative probability that any single nucleotide could be changed due to mutation versus recombination in *Pv*MSP5 genetic variation was carried out by calculating the recombination rate ρ (4N*e*r) relative to the rate of neutral mutation θ (4N*e*μ) (Guttman and Dykhuizen, 1994). Ne is the effective population size which cancels out when dividing ρ/θ. $\rho = 0{\cdot}0196$ and $\theta = 0{\cdot}02$ values were obtained using LDhat

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

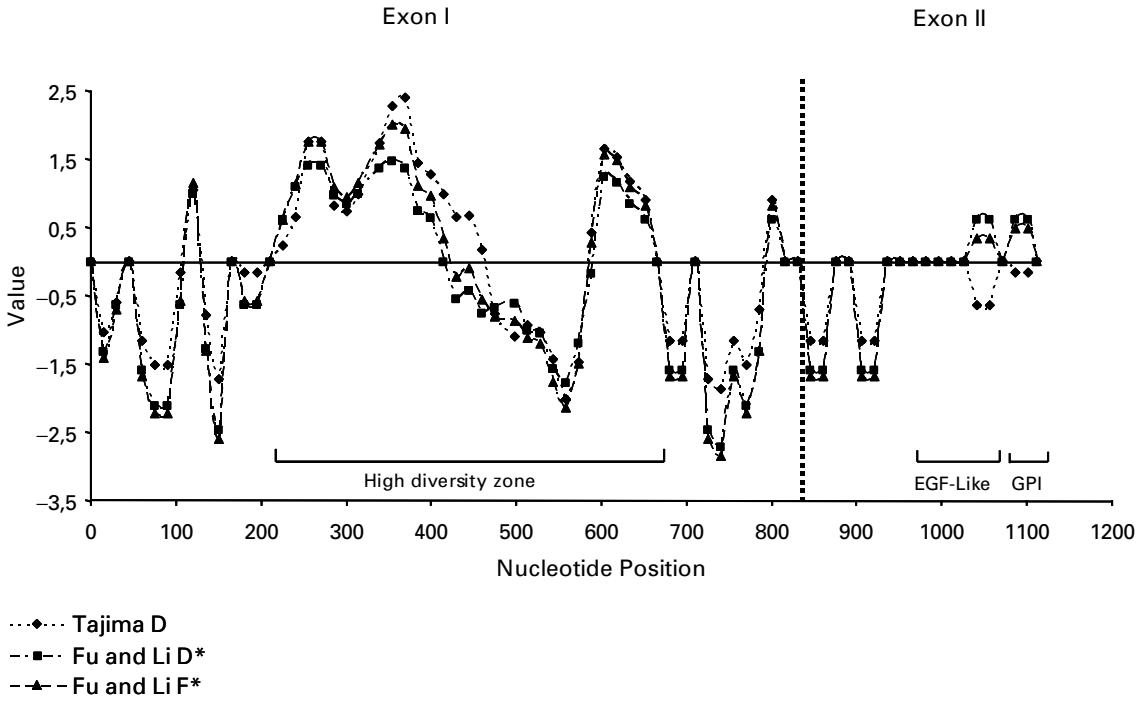


Fig. 4. Sliding-window analysis of Tajima's and Fu and Li's test along the *Pv*MSP5 sequence. The analysis was performed using a 30 nucleotide window size and 15 nucleotide step size.

package. Since these values are similar (and also in general agreement with those derived from DnaSP), this suggests that the variability observed in *Pv*MSP5 can be equally attributed to both recombination and mutation.

It has been suggested that *Pv*MSP4 and *Pv*MSP5 are the result of a duplication event given their high identity. As a consequence, it is important to establish whether these genes evolve independently or undergo any type of genetic exchange contributing to the allelic diversity. Cluster analysis using the above-mentioned genes and the *Pf*MSP4 and *Pf*MSP5, showed that orthologues group together revealing that genes in the same genome evolve independently, contrary to what would have been expected if a gene conversion event had occurred (Fitch *et al*. 1990; Aguileta *et al*. 2004;). In addition, a recombination analysis between *Pv*MSP4 and *Pv*MSP5 performed using the TOPALi software (Milne *et al*. 2004) did not detect any conversion events between the genes encoding these two proteins (data not shown).

Selection in Pv*MSP5*

Tajima (Tajima, 1989), Nei–Gojobori (Nei and Gojobori, 1986) and Fu and Li (Fu and Li, 1993) tests were used to test for departure from neutrality hypothesis (considering all exon I and II positions). Table 1 shows the results of this analysis for the obtained Colombian sequences.

The Nei-Gojobori test revealed significant positive selection for the set of Colombian sequences, when considering the whole transcript. When analysis was done by exons, significant positive selection was observed for exon I whilst exon II showed a higher number of d*S* over d*N*, this trend not being statistically significant. This test suggested that exon I has been subjected to some type of diversifying pressure and that this has been independent of the type of grouping of the sequences being considered (whether groups of geographical areas or in clusters of sequences).

The Tajima test does not present significant values different to zero for the combinations of sequences analysed, which might indicate a neutral selective pattern rather than a balanced or purifying selection. Similar results were obtained in the Fu and Li test (it is worth noting that these test results could be over-estimated due to the sample size).

As a whole, *Pv*MSP5 did not seem to be under any significant balanced or background selective pressure and only has a slight general tendency to negative values, as has been observed for *Pf*MSP5 (Polson *et al*. 2005). Positive selection was only observed when using the Nei-Gojobori test.

Fu and Li, and Tajima sliding window analysis for detecting selection, revealed different pressures acting in the area having greatest variability where positive selection was observed between nucleotides 255–414 and 603–651 and negative selection between nucleotides 417–573 (Fig. 4). All test values presented the same tendency. A pattern close to neutrality predominated in exon II or a tendency towards purifying selection was observed, whilst the whole of exon I has been under different

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

pressure, indicating possible areas where immune pressure had exerted different effects on *Pv*MSP5 evolution.

DISCUSSION

Studying the genetic variability of antigens relevant to the process of erythrocyte binding and invasion is a key aspect when developing anti-malarial vaccines. One of the proteins probably involved in such binding is *Pv*MSP5; its analysis has established that this is much more polymorphic when compared to other *P. vivax* proteins.

Cluster analysis has shown that there was no geographical grouping of the isolates being studied for *Pv*MSP5, a similar observation has been published for *Pv*DBP (Martinez *et al*. 2004) and *Pv*MSP3-α (Rayner *et al*. 2002). The grouping observed was related to the type of sequence variation, with clearly defined groups being shown. These presented less diversity compared to those obtained by geographical area. The lack of clustering amongst sequences from the same geographical region suggests either the presence of an ancient *P. vivax* population, the independent convergence of polymorphic variants in a recent population or the extensive exchange of parasite populations.

The sequences analysed came from samples from areas of Colombia having different transmission rates (Annual Parasite Index (API) 2003 = 1·32 for Norte de Santander, API 2003 = 18·10 Llanos Orientales and API 2003 = 16·74 Tumaco) (INS, 2003), even though the transmission rate in Colombia is considered to be low (Snewin *et al*. 1991) compared to areas of countries like India or Brazil (Martinez *et al*. 2005). Thus, the transmission rate of each geographical area was independent of the degree of diversity obtained.

Recombination in *Plasmodium* is another factor producing genetic variation (Conway *et al*. 1999). The results obtained with the coalescence method implemented in LDhat package (McVean *et al*. 2002) (according to the permutation test) support the notion that recombination is an important source of variation in *Pv*MSP5. In addition, the recombination to mutation ratio indicates that, in *Pv*MSP5, any particular nucleotide has an equal chance of being changed by mutation than by recombination. On the other hand, population interchange might be contributing towards creating the mosaic of groups (such as NS57, Lla06, Thai-NYU, Sal-I, Lla28 and Lla30) observed in the geographical areas sampled here.

The polymorphism observed in *Pv*MSP5 did not arise by chance throughout the whole gene but became localized in specific regions. Exon II has been shown to be more conserved in this work whilst the greater part of this protein's variation was located in exon I, this suggests that specific regions of this protein might be under selective pressure (as discussed below).

The conservation of the EGF-like motif in *Pv*MSP5 is in agreement with what has been observed in other surface proteins containing this motif, where some degree of divergence has been found in all regions, except in this one. It has also been shown that the EGF-like motif is an important conformational target for protective antibodies in naturally exposed individuals (Egan *et al*. 1999; O'Donnell *et al*. 2001) as well as in experimentally infected animals (Marshall *et al*. 1997; Wang *et al*. 1999; Stowers *et al*. 2001).

The differences exhibited between different types of proteins in both *P. falciparum* and *P. vivax* demonstrate the variety of such proteins at both functional and structural levels. Although proteins like EBA 175 in *P. falciparum* and DBP in *P. vivax* (both relevant in red blood cell invasion) (Barnwell *et al*. 1989; Sim, 1995) are under diversifying selection, they display differences regarding diversity due to *P. falciparum* having alternate invasion routes (Reed *et al*. 2000) and there is possibly less pressure, contrary to that observed in *P. vivax* where DBP is the only invasion route (Miller *et al*. 1976). Surface proteins present varying diversity levels but they are much higher compared to intracellular proteins such as AMA-1 in both *P. falciparum* and *P. vivax*. Most of them have repeats which are possibly used as 'smoke screens' for avoiding the host's immune system (Kemp *et al*. 1987). Positive selection is thought to be maintaining most of the observed polymorphism in malarial antigens, except for *Pf*MSP5 and *Pv*MSP4, suggesting that they might be exposed to different degrees of diversifying selection (such as immune system pressure).

Opposing selective forces (such as immune pressure versus functional constraint) acting on the same molecule have been described recently for DBP, where highly variable regions neighbouring the conserved and functionally important regions have been found experimentally (Singh *et al*. 2006).

The Nei-Gojobori test indicated positive selection action in *Pv*MSP5 exon I and a non-significant negative selection for exon II. Greater accumulation of dN than dS was also observed at exon I level for clusters of sequences. Tajima and Fu and Li tests did not present conclusive evidence that other forms of selection (such as balanced or purifying) were acting in an overall manner. Sliding window results for these tests showed that exon I was under a repertoire of more intense and diverse selective pressures than exon II. A partial explanation for the differences observed amongst the tests used, could be that Tajima's test is more sensitive to the demographic assumptions than the Nei-Gojobori test (Nielsen, 2005). Although we can not conclude with certainty the kind of selection acting on *Pv*MSP5 due to the contrasting results and the sample size,

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

according to the d*N* vs d*S* ratio it can be proposed that exon I is somewhat under immune system pressure (as opposed to exon II) and this variation might act as an evasion mechanism. This has also been observed in other sporozoite and merozoite surface proteins (Hughes and Hughes, 1995).

On the other hand, it is interesting that *Pv*MSP5 displays a greater variation than *Pf*MSP5. This seems odd, since *Pv*MSP5 is localized on the apical pole (Black *et al*. 2002) whilst *Pf*MSP5 is found on the merozoite surface (Marshall *et al*. 1998) and therefore, a greater immune system pressure should be expected to act on *Pf*MSP5. This could suggest that *Pv*MSP5 and *Pf*MSP5 are under different functional constraints.

A greater number of isolates from around the world and the simultaneous use of other proteins should be used to better understand the population structure of *P. vivax*. Functional studies should also be carried out for more precisely evaluating *Pv*MSP5 exon II's immunological role and its implications in developing a vaccine against *P. vivax* as this is the protein's most conserved region, making it a possible candidate in developing vaccines.

We would like to thank Stella Buitrago for supplying the *P. vivax*-infected blood samples, Jorge Egurrola and Ana Maria Rey for their technical support, and Jason Garry and Joanna Roocroft for reviewing the manuscript. We would also like to thank Professor Manuel Elkin Patarroyo and Professor Vladimir Corredor for their invaluable comments and suggestions. This work was supported by the President of Colombia's office and the Ministry of Social Protection.

REFERENCES

Aguileta, G., Bielawski, J. P. and Yang, Z. (2004). Gene conversion and functional divergence in the beta-globin gene family. *Journal of Molecular Evolution* **59**, 177–189.

Applied-Biosystems (1993). *Sequence Navigator v1.0.1*. Foster City CA (USA).

Babiker, H. A., Lines, J., Hill, W. G. and Walliker, D. (1997). Population structure of *Plasmodium falciparum* in villages with different malaria endemicity in east Africa. *American Journal of Tropical Medicine and Hygiene* **56**, 141–147.

Barnwell, J. W., Nichols, M. E. and Rubinstein, P. (1989). *In vitro* evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. *Journal of Experimental Medicine* **169**, 1795–1802.

Benet, A., Tavul, L., Reeder, J. C. and Cortes, A. (2004). Diversity of *Plasmodium falciparum* vaccine candidate merozoite surface protein 4 (MSP4) in a natural population. *Molecular and Biochemical Parasitology* **134**, 275–280.

Black, C. G., Barnwell, J. W., Huber, C. S., Galinski, M. R. and Coppel, R. L. (2002). The *Plasmodium vivax* homologues of merozoite surface proteins 4 and 5 from *Plasmodium falciparum* are expressed at different locations in the merozoite. *Molecular and Biochemical Parasitology* **120**, 215–224.

Carvalho, L. J., Daniel-Ribeiro, C. T. and Goto, H. (2002). Malaria vaccine: candidate antigens, mechanisms, constraints and prospects. *Scandinavian Journal of Immunology* **56**, 327–343.

Conway, D. J., Roper, C., Oduola, A. M., Arnot, D. E., Kremsner, P. G., Grobusch, M. P., Curtis, C. F. and Greenwood, B. M. (1999). High recombination rate in natural populations of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences, USA* **96**, 4506–4511.

Cui, L., Escalante, A. A., Imwong, M. and Snounou, G. (2003*a*). The genetic diversity of *Plasmodium vivax* populations. *Trends in Parasitology* **19**, 220–226.

Cui, L., Mascorro, C. N., Fan, Q., Rzomp, K. A., Khuntirat, B., Zhou, G., Chen, H., Yan, G. and Sattabongkot, J. (2003*b*). Genetic diversity and multiple infections of *Plasmodium vivax* malaria in Western Thailand. *American Journal of Tropical Medicine and Hygiene* **68**, 613–619.

Dayhoff, M. O. (1978). Protein segment dictionary 78. In *Atlas of Protein Sequence and Structure*, Vol. 5 pp. VII 470. National Biomedical Research Foundation, Silver Spring, Md, USA.

Egan, A. F., Burghaus, P., Druilhe, P., Holder, A. A. and Riley, E. M. (1999). Human antibodies to the 19 kDa C-terminal fragment of *Plasmodium falciparum* merozoite surface protein 1 inhibit parasite growth *in vitro*. *Parasite Immunology* **21**, 133–139.

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

Felger, I., Tavul, L., Kabintik, S., Marshall, V., Genton, B., Alpers, M. and Beck, H. P. (1994). *Plasmodium falciparum*: extensive polymorphism in merozoite surface antigen 2 alleles in an area with endemic malaria in Papua New Guinea. *Experimental Parasitology* **79**, 106–116.

Figtree, M., Pasay, C. J., Slade, R., Cheng, Q., Cloonan, N., Walker, J. and Saul, A. (2000). *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA1 and MSP1 genes. *Molecular and Biochemical Parasitology* **108**, 53–56.

Fitch, D. H., Mainone, C., Goodman, M. and Slightom, J. L. (1990). Molecular history of gene conversions in the primate fetal gamma-globin genes. Nucleotide sequences from the common gibbon, *Hylobates lar*. *Journal of Biological Chemistry* **265**, 781–793.

Fitch, W. (1971). Toward defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology* **20**, 406–416.

Fu, Y. X. and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Gaur, D., Mayer, D. C. and Miller, L. H. (2004). Parasite ligand-host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. *International Journal for Parasitology* **34**, 1413–1429.

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

Good, M. F. (2001). Towards a blood-stage vaccine for malaria: are we following all the leads? *Nature Reviews, Immunology* **1**, 117–125.

Guttman, D. S. and Dykhuizen, D. E. (1994). Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**, 1380–1383.

Hillis, D. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**, 182–192.

Hudson, R. R. and Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Hughes, A. L. and Verra, F. (2001). Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proceedings of the Royal Society, Biological Sciences* **268**, 1855–1860.

Hughes, M. K. and Hughes, A. L. (1995). Natural selection on *Plasmodium* surface proteins. *Molecular and Biochemical Parasitology* **71**, 99–113.

INS (2003). Datos de Vigilancia Epidemiológica: boletines de Enfermedades transmitidas por vectores. *Instituto Nacional de Salud Colombia*.

Jongwutiwes, S., Putaporntip, C., Friedman, R. and Hughes, A. L. (2002). The extent of nucleotide polymorphism is highly variable across a 3-kb region on *Plasmodium falciparum* chromosome 2. *Molecular Biology and Evolution* **9**, 1585–1590.

Kedzierski, L., Black, C. G. and Coppel, R. L. (2000). Immunization with recombinant *Plasmodium yoelii* merozoite surface protein 4/5 protects mice against lethal challenge. *Infection and Immunity* **68**, 6034–6037.

Kedzierski, L., Black, C. G., Goschnick, M. W., Stowers, A. W. and Coppel, R. L. (2002). Immunization with a combination of merozoite surface proteins 4/5 and 1 enhances protection against lethal challenge with *Plasmodium yoelii*. *Infection and Immunity* **70**, 6606–6613.

Kemp, D. J., Coppel, R. L. and Anders, R. F. (1987). Repetitive proteins and genes of malaria. *Annual Review of Microbiology* **41**, 181–208.

Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.

Marshall, V. M., Silva, A., Foley, M., Cranmer, S., Wang, L., Mccoll, D. J., Kemp, D. J. and Coppel, R. L. (1997). A second merozoite surface protein (MSP-4) of *Plasmodium falciparum* that contains an epidermal growth factor-like domain. *Infection and Immunity* **65**, 4460–4467.

Marshall, V. M., Tieqiao, W. and Coppel, R. L. (1998). Close linkage of three merozoite surface protein genes on chromosome 2 of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **94**, 13–25.

Martinez, P., Suarez, C. F., Cardenas, P. P. and Patarroyo, M. A. (2004). *Plasmodium vivax* Duffy binding protein: a modular evolutionary proposal. *Parasitology* **128**, 353–366.

Martinez, P., Suarez, C. F., Gomez, A., Cardenas, P. P., Guerrero, J. E. and Patarroyo, M. A. (2005). High level of conservation in *Plasmodium vivax* merozoite surface protein 4 (PvMSP4). *Infection, Genetics and Evolution* **5**, 354–361.

Mascorro, C. N., Zhao, K., Khuntirat, B., Sattabongkot, J., Yan, G., Escalante, A. A. and Cui, L. (2005). Molecular evolution and intragenic recombination of the merozoite surface protein MSP-3alpha from the malaria parasite *Plasmodium vivax* in Thailand. *Parasitology* **131**, 25–35.

McVean, G., Awadalla, P. and Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241.

Milne, I., Wright, F., Rowe, G., Marshall, D. F., Husmeier, D. and McGuire, G. (2004). TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* **20**, 1806–1807.

Miller, L. H., Mason, S. J., Clyde, D. F. and McGinniss, M. H. (1976). The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *New England Journal of Medicine* **295**, 302–304.

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Nicholas, K., Hb, N. and Deerfield, D. (1997). GeneDoc: analysis and visualization of genetic variation. *EMBNEW.NEWS* **4**, 14.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197–218.

O'Donnell, R. A., De Koning-Ward, T. F., Burt, R. A., Bockarie, M., Reeder, J. C., Cowman, A. F. and Crabb, B. S. (2001). Antibodies against merozoite surface protein (MSP)-1(19) are a major component of the invasion-inhibitory response in individuals immune to malaria. *Journal of Experimental Medicine* **193**, 1403–1412.

Polson, H. E., Conway, D. J., Fandeur, T., Mercereau-Puijalon, O. and Longacre, S. (2005). Gene polymorphism of *Plasmodium falciparum* merozoite surface proteins 4 and 5. *Molecular and Biochemical Parasitology* **142**, 110–115.

Polley, S., MacRobert, L. and Sutherland, C. (2004). Vaccination for vivax malaria: targeting the invaders. *Trends in Parasitology* **20**, 99–102.

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

Putaporntip, C., Jongwutiwes, S., Tia, T., Ferreira, M. U., Kanbara, H. and Tanabe, K. (2001). Diversity in the thrombospondin-related adhesive protein gene (TRAP) of *Plasmodium vivax*. *Gene* **268**, 97–104.

Rainczuk, A., Scorza, T., Spithill, T. W. and Smooker, P. M. (2004). A bicistronic DNA vaccine containing apical membrane antigen 1 and merozoite surface protein 4/5 can prime humoral and cellular immune responses and partially protect mice against virulent *Plasmodium chabaudi adami* DS malaria. *Infection and Immunity* **72**, 5565–5573.

Rayner, J. C., Corredor, V., Feldman, D., Ingravallo, P., Iderabdullah, F., Galinski, M. R. and

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press

Barnwell, J. W. (2002). Extensive polymorphism in the *Plasmodium vivax* merozoite surface coat protein MSP-3alpha is limited to specific domains. *Parasitology* **125**, 393–405.

Reed, M. B., Caruana, S. R., Batchelor, A. H., Thompson, J. K., Crabb, B. S. and Cowman, A. F. (2000). Targeted disruption of an erythrocyte binding antigen in *Plasmodium falciparum* is associated with a switch toward a sialic acid-independent pathway of invasion. *Proceedings of the National Academy of Sciences, USA* **97**, 7509–7514.

Rozas, J., Sanchez-Delbarrio, J. C., Messeguer, X. and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Rzhetsky, A. and Nei, M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution* **10**, 1073–1095.

Sim, B. K. (1995). EBA-175: an erythrocyte-binding ligand of *Plasmodium falciparum*. *Parasitology Today* **11**, 213–217.

Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E. and Sharma, A. (2006). Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature, London* **439**, 741–744.

Sitnikova, T. (1996). Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol* **13**, 605–611.

Snewin, V. A., Herrera, M., Sanchez, G., Scherf, A., Langsley, G. and Herrera, S. (1991). Polymorphism of the alleles of the merozoite surface antigens MSA1 and MSA2 in *Plasmodium falciparum* wild isolates from Colombia. *Molecular and Biochemical Parasitology* **49**, 265–275.

Snow, R. W., Guerra, C. A., Noot, A. M., Myinme, H. Y. and Hay, S. I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature, London* **434**, 214–217.

Stowers, A. W., Cioce, V., Shimp, R. L., Lawson, M., Hui, G., Muratova, O., Kaslow, D. C., Robinson, R., Long, C. A. and Miller, L. H. (2001). Efficacy of two alternate vaccines based on *Plasmodium falciparum* merozoite surface protein 1 in an *Aotus* challenge trial. *Infection and Immunity* **69**, 1536–1546.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Volkman, S. K., Barry, A. E., Lyons, E. J., Nielsen, K. M., Thomas, S. M., Choi, M., Thakore, S. S., Day, K. P., Wirth, D. F. and Hartl, D. L. (2001). Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293**, 482–484.

Wang, L., Black, C. G., Marshall, V. M. and Coppel, R. L. (1999). Structural and antigenic properties of merozoite surface protein 4 of *Plasmodium falciparum*. *Infection and Immunity* **67**, 2193–2200.

Wang, L., Goschnick, M. W. and Coppel, R. L. (2004). Oral immunization with a combination of *Plasmodium yoelii* merozoite surface proteins 1 and 4/5 enhances protection against lethal malaria challenge. *Infection and Immunity* **72**, 6172–6175.

World Health Organization (2002). Epidemiological Bulletin. World Health Organization, Geneva.

Wu, T., Black, C. G., Wang, L., Hibbs, A. R. and Coppel, R. (1999). Lack of sequence diversity in the gene encoding merozoite surface protein 5 of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **103**, 243–250.

https://doi.org/10.1017/S0031182006001168 Published online by Cambridge University Press