Polymorphism of the Pv200L Fragment of Merozoite Surface Protein-1 of *Plasmodium vivax* in Clinical Isolates from the Pacific Coast of Colombia

Augusto Valderrama-Aguirre, Evelin Zúñiga-Soto, Leonardo Mariño-Ramírez, Luz Ángela Moreno, Ananías A. Escalante, Myriam Arévalo-Herrera, and Sócrates Herrera*

Instituto de Inmunología del Valle, Universidad del Valle, Cali, Colombia; Malaria Vaccine and Drug Development Center, Cali, Colombia; National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland; School of Life Sciences, Arizona State University, Tempe, Arizona

Abstract. Merozoite surface protein 1 (MSP-1) is a polymorphic malaria protein with functional domains involved in parasite erythrocyte interaction. *Plasmodium vivax* MSP-1 has a fragment (Pv200L) that has been identified as a potential subunit vaccine because it is highly immunogenic and induces partial protection against infectious parasite challenge in vaccinated monkeys. To determine the extent of genetic polymorphism and its effect on the translated protein, we sequenced the Pv200L coding region from isolates of 26 *P. vivax*-infected patients in a malaria-endemic area of Colombia. The extent of nucleotide diversity (π) in these isolates (0.061 ± 0.004) was significantly lower ($P \leq 0.001$) than that observed in Thai and Brazilian isolates; 0.083 ± 0.006 and 0.090 ± 0.006 , respectively. We found two new alleles and several previously unidentified dimorphic substitutions and significant size polymorphism. The presence of highly conserved blocks in this fragment has important implications for the development of Pv200L as a subunit vaccine candidate.

INTRODUCTION

Plasmodium vivax is responsible for about 20% of the global malaria cases and more than half (56%) of the non-African malarial infections. Indeed, *P. vivax* has reemerged in many regions of the world where malaria was eliminated in the 1950–60s with 70–80 million cases per year and 2.6 billion people at risk of infection.^{1,2} Given its broad distribution, *P. vivax* co-exists with *Plasmodium falciparum* and in minor proportion with *P. malariae*.^{3,4} Regardless, the progress achieved in the development of a vaccine against *P. falciparum*, there is paucity of suitable vaccine candidates against *P. vivax*, with only two antigens currently under evaluation in human clinical trials and a few others in preclinical evaluation.^{2–6} The understandable bias in research effort toward *P. falciparum*, however, hampers our ability of developing a vaccine that can be effectively deployed against malaria outside Africa where these two parasites coexist.

We previously defined the Pv200L fragment of PvMSP-1, located toward the N-terminal end of the 83 kDa domain (Figure 1), as a *P. vivax* potential subunit vaccine based on several features of the protein. The Pv200L has significant homology to Pf190L, a well-defined *P. falciparum* vaccine candidate,^{7–9} and a Pv200L recombinant protein produced in *Escherichia coli*, displayed a high level of antigenicity in humans. Additionally, this protein fragment showed good immunogenicity in mice and primates, and the capacity to induce partial protection against a *P. vivax* blood-stage challenge in *Aotus* monkeys.⁷

The Pv200L fragment includes the entire blocks 2, 3, and 4, plus segments of blocks 1 and 5 of the PvMSP-1.¹⁰ Genetic polymorphism studies in *P. vivax* isolates have shown that blocks 1, 3, and 5 are conserved at the protein level and display a few dimorphic substitutions. Blocks 2 and 4 are the most variable, both in size and sequence, with basic and recombinant block types generated by intra- and inter-allelic recombination events.¹¹ Here, we describe the polymorphism of the Pv200L

gene fragment and its inferred amino acid sequence in 26 *P. vivax* clinical isolates from Buenaventura, a malaria-endemic area located on the southern Pacific coast of Colombia.⁴ We also examined 42 Pv200L fragments from PvMSP-1 sequences previously reported in GenBank and describe the phylogenetic analysis of all sequences.

MATERIALS AND METHODS

Study population and *P. vivax* isolates. The Colombian Pacific region is composed of four states—Chocó, Valle, Cauca, and Nariño—and is considered the second most malaria-endemic region of Colombia, as it accounts for about 30% of the country's disease burden.⁴ In Colombia *P. vivax* is the predominant malaria species and is responsible for more than 60% of the clinical cases reported every year; however, because of the high prevalence of Duffy-negative in the Afro-colombian habitants of this region, the predominant species in the Pacific region is *P. falciparum*.⁴ Blood samples were collected by convenience during 2004 and 2005 from symptomatic patients diagnosed by thick smear at the outpatient clinical facilities of the Malaria Vaccine and Drug Development Center (MVDC) in Buenaventura, the main seaport on the Colombian Pacific coast. Buenaventura, located in the Valle state, has ~400,000 inhabitants and conditions for low and unstable malaria transmission.¹² A total of 26 samples exclusively infected by *P. vivax*, as confirmed by polymerase chain reaction (PCR),¹³ were collected from patients from rural and urban communities. All adult participants and the parents, or legal guardians of minor patients, were asked to provide a signed written informed consent previously approved by the Institutional Review Board of Universidad del Valle.

PCR, cloning, and sequencing. The primers 200L-1 (5'-GCC AAG CTG GAC AAG TTA GA-3') and 200L-2 (5'-AAG GTT GGA ACT GTC TTT CC-3') were designed to amplify by PCR the Pv200L coding region (bp 143–1,333) from the PvMSP-1 of Salvador 1 strain (GenBank accession no. AF435593). The primers were confirmed to align with all available complete sequences of PvMSP-1 in GenBank.¹⁴ The PCR reactions were performed with iProof high-fidelity DNA polymerase (Bio-Rad, Hercules, CA) in a PTC-100 thermal cycler (MJ Research, Watertown, MA) as follows: 35 cycles of

*Address correspondence to Sócrates Herrera, Malaria Vaccine and Drug Development Center, Carrera 37 - 2Bis No. 5E - 08, Cali, Colombia. E-mail: sherrera@inmuno.org

FIGURE 1. Nucleotide diversity (π) scores along the *Pv*200L gene fragment. The scores are shown for three groups of isolates: 26 from Colombia, 20 from Thailand, and 8 from Brazil. The sequences from Thai and Brazil isolates were extracted from complete *Pv*MSP-1 cds sequences available in GenBank. Relative position of blocks is indicated in the bar. Variable blocks are indicated in black.

30 sec at 94°C, 60 sec at 55°C, and 60 sec at 72°C. The PCR products (1–1.2 Kb) were ligated to pCR4-TOPO (Invitrogen, Carlsbad, CA) according to manufacturer instructions, and then used to transform chemically competent *E. coli* One-Shot (Invitrogen). Kanamycin-resistant clones were confirmed by restriction enzyme analysis with *Eco*RI (Fermentas, Hanover, MD). We sequenced only one clone per isolate. Sequencing was performed with BigDye Terminator version 3.1 kit and the M14 forward and reverse primers in an ABI-PRISM AVANT 3100 sequencer (Applied Biosystems, Foster City, CA). Every run was performed a minimum of two times and repeated as much as needed to obtain quality values > 20 (< 4 Ns/20 bases, < 10% Ns, and maximum percent of mixed bases = 20.0%) as assessed with SeqScape Software (Applied Biosystems).

Sequences, sequence alignment, and statistical analysis. We analyzed the 26 *Pv*200L gene sequences from the Colombian Pacific coast, and the 42 *Pv*200L fragments from *Pv*MSP-1 sequences available in GenBank, accession numbers: AF435593–AF435599, AF435601–AF435620, AF435622–AF435625, AF435627, AF435629–AF435632, and AF435634–AF435639 (October 2006).¹⁰ These GenBank sequences were collectively named as non-Colombian. Sequences were aligned using CLUSTAL X2,¹⁵ and the alignment was used to calculate nucleotide diversity ($\pi \pm SE$) with MEGA 4.0 using the Jukes-Cantor model and retaining all gaps.¹⁶ The π value obtained for *Pv*200L and each block was submitted to one-way analysis of variance statistical analysis. DNA polymorphism was analyzed with DnaSP 4.10 with a sliding window of 100 bases and a step size of 20 bases for a haploid genome.¹⁷ To determine the influence of geographical origin of the isolates on nucleotide diversity and DNA polymorphism, we performed the same analysis for groups of GenBank sequences from Thailand (20 isolates) and Brazil (8 isolates). The output data were exported to STATA 8.0 (Stata Corp., College Station, TX) to plot overlapping π curves with standardized scales for the

different groups. Boundaries of conserved and variable blocks were determined from nucleotide sequence homology, as previously reported.¹⁰ Subsequently, π and its standard error were computed for each block. The number of synonymous (d_s) and non-synonymous (d_n) substitutions was estimated for each block, to avoid bias caused by size polymorphisms, by Nei and Gojobori's method with the Jukes-Cantor correction in MEGA 4.0.^{18–20} We estimated the difference between D_s and D_n and its standard deviation was calculated using bootstrap with 500 pseudo-replications for D_s and D_n , and two-tail Z-test on the difference between D_s and D_n .²¹ The null hypothesis is that $D_s = D_n$; thus, we assumed as null hypothesis that the observed polymorphism was neutral. *In silico* translated DNA sequences were used to identify the types of variable blocks (2 and 4), as previously defined,¹⁰ and point mutations in conserved blocks. Alleles were identified by the specific combination of variable block types as proposed by Putaporntip and others.¹⁰

Phylogenetics and epitope conservation analysis. The distances of the *P. vivax* isolates were inferred from phylogenetic analysis of the translated *Pv*200L sequences using MEGA 4.0. Because of size polymorphism in variable blocks, distances were calculated only with sequence alignments of conserved edited-joined blocks 1, 3, and 5. Trees were constructed using the neighbor-joining (NJ) method, excluding gaps by pairwise deletion, with the Kimura p-distance model.²² The reliability of the trees was assessed by the bootstrap method with 1,000 pseudo-replications. Phylogenetic analysis was performed with the 26 Colombian sequences and the 42 sequences available in GenBank. Complementary analyses to determine genetic divergences was performed using the Fstat program (Fst). Finally, we determined the conservation pattern of previously defined promiscuous T-helper epitopes contained in the *Pv*200L fragment.²³ To this purpose we constructed multiple sequence alignments with the 68 *Pv*200L available sequences and then realigned the T-helper epitope sequences in CLUSTAL X2.

RESULTS

Geographical origin of Colombian isolates. The *Pv*200L gene fragment was sequenced and analyzed in a total of 26 isolates from the Colombian Pacific coast obtained from *P. vivax*-infected patients. Eighteen (69.2%) of the isolates were collected from patients who preferred to live in rural areas of Buenaventura (Table 1). The majority of rural isolates were from La Delfina and San Cipriano villages. The urban isolates were predominantly from Commune 12, of Buenaventura

TABLE 1
Origin of the Colombian *Plasmodium vivax* isolates

Place	Locality	n	Isolates
Comuna 12	U*	6	CU45, CU57, CU65, CU66, CU81, CU83
La Delfina	R†	6	D20, D48, D61, D85, D102, D103
San Cipriano	R	4	SC84, SC92, SC93, SC1
Buenaventura	U	2	B74, B30
La Laguna	R	1	L8
La Gloria	R	1	G91
Cordoba	R	1	Co68
Zacarias	R	1	Z80
San Francisco	R	1	SF56
Triana	R	1	T28
Citronela	R	1	Ci22
Potodo	R	1	P54

* Urban.

† Rural.

TABLE 2
Nucleotide diversity ($\pi \pm \text{SE}$)* per blocks of the *Pv200L* gene fragment

Groups	n	Overall	Block 1	Block 2	Block 3	Block 4	Block 5
All	68	0.088 $\pm$ 0.006	0.020 $\pm$ 0.004	0.233 $\pm$ 0.020	0.049 $\pm$ 0.017	0.160 $\pm$ 0.021	0.049 $\pm$ 0.010
Non-Colombian†	42	0.086 $\pm$ 0.005	0.022 $\pm$ 0.004	0.212 $\pm$ 0.018	0.052 $\pm$ 0.017	0.162 $\pm$ 0.024	0.051 $\pm$ 0.011
Colombian	26	0.061 $\pm$ 0.004‡	0.011 $\pm$ 0.002‡	0.155 $\pm$ 0.014‡	0.040 $\pm$ 0.014	0.134 $\pm$ 0.022	0.039 $\pm$ 0.010
Thailand	20	0.083 $\pm$ 0.006	0.022 $\pm$ 0.004	0.228 $\pm$ 0.020	0.042 $\pm$ 0.015	0.126 $\pm$ 0.019	0.038 $\pm$ 0.008
Brazil	8	0.090 $\pm$ 0.006	0.021 $\pm$ 0.004	0.203 $\pm$ 0.019	0.056 $\pm$ 0.020	0.166 $\pm$ 0.026	0.035 $\pm$ 0.015

*As calculated with MEGA 3.0 using the Jukes-Cantor model.

† *Pv200L* gene fragment trimmed out from the 42 *Pv* MSP-1 cds available in GenBank by June 2009.

‡ $P \leq 0.001$ as calculated by one-way analysis of variance (ANOVA) Colombia vs. Thailand and Colombia vs. Brazil.

town. This commune is the closest community to the humid rain forests in the Pacific coast.

Nucleotide diversity. We found a nucleotide diversity value of 0.088 ± 0.006 for the *Pv200L* gene fragment for all the isolates, including Colombian and non-Colombian (Table 2). Such π value was significantly lower in the Colombian group than in either Thai or Brazilian subgroups of sequences. The highest contribution for nucleotide diversity was observed in the Brazilian group followed by the Thai group. As expected, the variable blocks 2 and 4 had the highest π values across all of the groups. Block 2 from Colombian isolates had a significantly lower π value than Thai and Brazilian isolates (Figure 1 and Table 2); whereas Thai sequences displayed the lowest π value for block 4. Regarding conserved blocks 1, 3, and 5, all of them displayed the expected conservation. In the case of block 1, the nucleotide diversity was significantly lower in the Colombian isolates than in the Thai and Brazilian ones. Although block 3 had also the lowest π value in the Colombian parasites, the difference was not statistically significant. Conserved block 5 displayed π values below 0.040 in all isolates, with the lowest diversity in the Brazilian isolates (Table 2).

TABLE 3

Synonymous (d_s) and non-synonymous (d_n) nucleotide substitutions in *Pv200L* fragment

Groups	Block	$d_s^* \pm \text{SE}$	$d_n^* \pm \text{SE}$	Z
All ($N = 68$)	1	0.053 $\pm$ 0.016	0.012 $\pm$ 0.003	-2.5944†
	2	0.168 $\pm$ 0.036	0.259 $\pm$ 0.032	-2.5930†
	3	0.165 $\pm$ 0.084	0.020 $\pm$ 0.011	-1.6597
	4	0.110 $\pm$ 0.044	0.174 $\pm$ 0.026	1.1718
	5	0.141 $\pm$ 0.048	0.023 $\pm$ 0.008	-2.5111†
Non-Colombian ($N = 42$)	1†	0.057 $\pm$ 0.017	0.012 $\pm$ 0.004	-2.6581†
	2	0.132 $\pm$ 0.028	0.244 $\pm$ 0.028	-2.67242†
	3	0.182 $\pm$ 0.088	0.019 $\pm$ 0.011	-1.7733
	4	0.088 $\pm$ 0.038	0.184 $\pm$ 0.030	1.891
	5	0.142 $\pm$ 0.046	0.026 $\pm$ 0.010	-2.4249
Colombian ($N = 26$)	1	0.030 $\pm$ 0.009	0.006 $\pm$ 0.002	-2.6278†
	2	0.142 $\pm$ 0.029	0.161 $\pm$ 0.020	-2.13556†
	3	0.116 $\pm$ 0.066	0.020 $\pm$ 0.010	-1.2895
	4	0.116 $\pm$ 0.052	0.139 $\pm$ 0.023	0.3967
	5	0.118 $\pm$ 0.043	0.017 $\pm$ 0.006	-2.1372†
Thailand ($N = 20$)	1	0.053 $\pm$ 0.016	0.013 $\pm$ 0.004	-2.49773†
	2	0.152 $\pm$ 0.034	0.259 $\pm$ 0.030	-2.42017†
	3	0.141 $\pm$ 0.078	0.016 $\pm$ 0.009	-1.6835
	4	0.079 $\pm$ 0.034	0.139 $\pm$ 0.022	1.4794
	5	0.105 $\pm$ 0.034	0.019 $\pm$ 0.008	-2.3153†
Brazil ($N = 8$)	1	0.060 $\pm$ 0.018	0.010 $\pm$ 0.003	-2.7435†
	2	0.140 $\pm$ 0.030	0.229 $\pm$ 0.028	-2.5384†
	3	0.184 $\pm$ 0.090	0.024 $\pm$ 0.013	-1.5954
	4	0.108 $\pm$ 0.043	0.183 $\pm$ 0.033	1.3086
	5	0.190 $\pm$ 0.064	0.033 $\pm$ 0.013	-2.4627†

* d_s and d_n are per every 100 sites.

† Indicates significant with $P < 0.05$.

Synonymous and non-synonymous substitutions. As expected, the d_n value was higher in blocks 2 and 4 (Table 3); however, the value was only significant in block 2. This was true for each country and the combined sample. The d_s value was significantly higher than d_n for blocks 1 and 5 using the Z-test, indicating that some blocks involved in this fragment could be under negative selection.

Non-synonymous substitutions in conserved blocks of Colombian isolates. We found a total of 26 dimorphic substitutions across conserved blocks 1, 3, and 5, and seven of them are newly identified (Table 4). Some substitutions were consistently linked within isolates, such as SC84, SC93, SC1, and Co68, most of them collected from inhabitants from San Cipriano village, which shared four dimorphic substitutions, and isolates D61 and B74 that shared five dimorphic substitutions plus two insertions (block 1: 61[-/A] and 62[-/S]).

TABLE 4

Dimorphic substitutions in conserved blocks

Block	Mutation*	Pos†	Proportion‡	Isolates§
1	N/S	53	24:02:00	D61, B74
	K/Q	56	24:02:00	D61, B74
	V/G¶	57	25:01:00	B74
	D/E	58	21:05	D61, B74, SC84, SC93, SC1
	A/T	59	24:02:00	D61, B74
	G/S	83	24:02:00	D61, B74
	S/F¶	85	25:01:00	SC1
	F/V	105	24:02:00	D61, B74
	N/H¶	108	25:01:00	SC1
	H/R¶	152	25:01:00	CU45
3	I/V¶	155	25:01:00	D48
	T/S	156	19:07	<u>Co68, SC84, SC93, SC1</u> , D61, B74, SC92
	E/D	157	22:04	Co68, SC84, SC93, SC1
	A/T	186	24:02:00	Co68, SC92
	E/A	315	17:09	L8, D20, CU45, D48, D61, CU66, B74, G91, D103
	D/G	329	22:04	<u>Co68, SC84, SC93, SC1</u>
5	V/A	332	21:05	<u>Co68, SC84, SC93, SC1</u> , SF56
	D/Y¶	384	25:01:00	SC92
	D/N	389	13:13	Ci22, T28, B30, P54, SF56, CU57, CU65, Z80, CU81, CU83, D85, SC92, D102
	I/V	409	25:01:00	SC92
	S/N¶	412	25:01:00	D103
	K/S	414	25:01:00	SC92
	S/A	415	25:01:00	SC92
	A/S	416	25:01:00	SC92
	G/S	417	25:01:00	SC92
	P/T	419	25:01:00	SC92

*The least frequent amino acid residue is presented as denominator.

† Position. Referred as to the sequence of *Pv* MSP-1 of Salvador I strain.

‡ Exact number of Colombian isolates that presented the most frequent over the least frequent amino acid residue.

§ Only those isolates that contain the least frequent substitution are listed. Isolates with linked substitutions are shown with special characters (bold, italic, and italic/bold/underlined).

¶ Newly identified substitutions.

TABLE 5
Classification of variable blocks

Type	# (%)	Isolates
Block 2a		
Basic 1	1 (3.8)	SC92
Basic 2	4 (15.4)	Co68, SC84, SC93, SC1
Basic 3	2 (7.7)	D61, B74
Basic 4	7 (26.9)	L8, D20, CU45, D48, CU66, G91, D103
Rec 6	10 (38.5)	Ci22, T28, B30, P54, SF56, CU57, Z80, CU81, CU83, D102
Rec 7	2 (7.7)	CU65, D85
Block 2b		
Type 20*	1 (3.8)	SC92
Type 14	7 (26.9)	L8, D20, CU45, D48, CU66, G91, D103
Type 15	12 (46.2)	Ci22, T28, B30, P54, SF56, CU57, CU65, Z80, CU81, CU83, D85, D102
Type 16	2 (7.7)	D61, B74
Type 19	4 (15.4)	Co68, SC84, SC93, SC1
Block 2c		
Basic 2	1 (3.8)	Co68
Basic 3	19 (73.1)	L8,† D20,‡ CU45,‡ D48,‡ CU66,‡ G91,‡ D103,‡ Ci22, T28, B30, P54, SF56, CU57, CU65, Z80, CU81, CU83, D85, D102
Rec 5	2 (7.7)	D61, B74
Rec 7	1 (3.8)	SC92
Rec 10	3 (11.6)	SC84, SC93, SC1
Block 4		
Basic 1	7 (26.9)	L8, D20, CU45, D48, CU66, G91, D103
Basic 2	4 (15.4)	Co68, SC84, SC93, SC1
Basic 3	1 (3.8)	SC92
Rec 5	14 (53.9)	Ci22, T28, B30, P54, SF56, CU57, CU65, Z80, CU81, CU83, D85, D102, D61,‡ B74‡

= Number of block types.

* Newly identified insertion of GSSNS after fifth amino acid.

† Seven isolates had the P/S mutation in the fourth amino acid.

‡ Two isolates had the four linked substitutions I/V, D/E, V/E, and D/A at positions 3, 33, 39, and 40.

Characteristically, only isolate SC92 had seven out of nine of the dimorphic substitutions identified for block 5 (Table 4).

Variable blocks classification in Colombian isolates. Results were consistent with previously reported data.¹⁰ We identified six block 2a types (four basic and two recombinants), four block 2b types (three previously reported and a new one), five block 2c types (two basic and three recombinants), and four block 4 types (three basic and one recombinant) (Table 5). The new block 2b type named by us as type 20, had an extra insertion of the tandem repeat GSSNS in the SC92 isolate. Additionally, seven isolates shared the P/S substitution in block 2c type 3. These last two findings helped to identify the two new alleles reported here (see below). Several new substitutions were identified in each block. Interestingly, the recombinant types

identified contained fragments from basic types previously reported but not detected in the 26 Colombian isolates (see supplemental data).

Allelic distribution. The variable blocks included in the *Pv*200L fragment allowed us to classify the isolates in 12 potential alleles pooled in seven groups. Alleles are presented in groups because the definite allele needs of other variable blocks are not included in the *Pv*200L fragment (Table 6). We found two non-previously described alleles, which we have designated as 32 and 33, to continue with the nomenclature proposed previously.¹⁰ Allele 32 was designed as a new allele because the block 2b type 20 and only isolate SC92 had it. The reason why allele 33 was considered as new is because its mosaic organization of variable blocks 2 (a, b, and c) and 4 have not been previously described. A total of seven of the isolates (26.9%) were classified as allele 32 and all shared the dimorphic substitution P/S in the block 2c (Table 6). The majority of isolates (38.5%) were grouped as allele 6 or 20, meanwhile the isolate SC1 that is being used to develop a vaccine candidate was the third most frequent (11%).

Phylogenetic analysis. The NJ tree, created with distances between amino acid sequences of conserved blocks, showed that the 26 Colombian isolates clustered in agreement with the allele distribution, which was determined with the specific combination of variable blocks (Figure 2A). The isolate SC1 clustered with the isolates SC84 and SC93, all of them collected from inhabitants of San Cipriano village. The NJ tree created with the 68 *Pv*200L sequences showed that there is a slight trend to cluster in agreement with the geographical origin (Figure 2B), being more evident for the isolates from Colombia (COL), South Korea (SK), and Bangladesh (BD). Brazil (B) and Thailand (T) isolates displayed a more promiscuous clustering. The isolate SC92, collected from San Cipriano village, was the most distant of Colombian isolates, and clustered fairly close to Asian sequences (Figure 2B). A complementary analyses with *Fst* among Colombian, Brazilian, and Thai isolates revealed that there is a strong geographic structure; however, the lowest divergence (*Fst* = 0.06453) was observed between Brazilian and Thailand isolates, whereas Colombian isolates were clearly divergent (*Fst* = 0.1676 with Thailand and *Fst* = 0.2437 with Brazil).

Epitope conservation analysis. At least five different T-helper epitopes have been previously defined toward the N-terminal portion of *Pv*MSP-1, and four of them are included within the *Pv*200L fragment.²³ All of them showed a high conservation pattern along the 68 *Pv*200L amino acid sequences (Table 7).

TABLE 6
Frequency of the *Pv*Msp-1 *Pv*200L alleles identified in the Colombian Pacific coast

Block 2a	Block 2b	Block 2c	Block 4	Allele*	Isolates	Frequency (%)
Rec 6	Type 15	Basic 3	Rec 5	6, 20	Ci22, T28, B30, P54, SF56, CU57, Z80, CU81, CU83, D102	38.5
Basic 4	Type 14	Basic 3†	Basic 1	33‡	L8, D20, CU45, D48, CU66, G91, D103	26.9
Basic 2	Type 19	Rec 10	Basic 2	7, 21	SC84, SC93, SC1	11.5
Basic 3	Type 16	Rec 5	Rec 5	5, 11, 12, 13	D61, B74	7.7
Rec 7	Type 15	Basic 3	Rec 5	25	CU65, D85	7.7
Basic 1	Type 20§	Rec 7	Basic 3	32	SC92	3.8
Basic 2	Type 19	Basic 2	Basic 2	31	Co68	3.8

* Allele is defined by the specific arrangement of variable blocks. Several alleles mean that any of them contain that specific combination of blocks 2a, 2b, 2c, and 4. To define the exact allele needs variable blocks that are not included in the *Pv*200L sequence.

† P/S mutation in the fourth amino acid across the seven isolates.

‡ New alleles.

§ Newly identified insertion of repeat GSSNS after fifth amino acid.

FIGURE 2. Neighbor-joining (NJ) phylogenetic trees of *Pv200L*. Trees were built with edited-joined sequences of conserved blocks 1, 3, and 5. (A) NJ tree of the 26 *Pv200L* sequences from Colombian *Plasmodium vivax* isolates. Isolates alleles are denoted as follows: ▲ (Alleles 6 or 20), △ (Allele 33), ◆ (Alleles 7 or 21), ◇ (Alleles 8, 11, 12 or 13), • (Allele 31), and ■ (Allele 32). (B) NJ trees of trees of 68 *Pv200L* sequences, including 42 obtained from *PvMSP-1* cds available in Genbank and the 26 Colombian isolates (▼COL).

DISCUSSION

As expected, the *Pv200L* fragment of *PvMSP-1* showed to be polymorphic in sequence and in size, which resembles very well the mosaic structure previously described by others, with a polymorphism greatly concentrated in fragments 2 and 4, surrounded by the well-conserved blocks 1, 3, and 5.¹⁰ However, regardless of this polymorphism, there is some evidence of purifying selection. Although this study focused only in the *Pv200L* fragment, our results may indicate a more complex dynamic of selection acting on the *PvMSP-1* polymorphism

than previously thought.²⁴ The overall nucleotide polymorphism and the genetic diversity found in the isolates from the Colombian Pacific coast was substantially lower than in those from Thailand and Brazil, probably due in part to the lower/seasonal transmission and geographic isolation. Most malaria cases studied here occur in close rural communities with high internal mobility but low foreign influence from potentially *P. vivax*-infected international travelers. This geographic isolation also could explain the pattern of *Fst*. However, presence of imported parasites cannot be completely excluded. SC92

TABLE 7

Conservation pattern of promiscuous T-helper epitopes in Pv200L

Id§	Sequence conservation
PvT4	NFVGKFLQLQIPGHTDLLHL *:*.******
PvT6	FNQLMHVINFHFDLLRAKLH *****
PvT8	LDMLKKVVLGLYRPLDNIKD *****
PvT19	LEYYLREKAKMAGTLIIPES *****.*:**

§As reported by Caro-Aguilar and others.²³ Asterisk stands for 100% conserved residue across 68 sequences analyzed.

was the most distant Colombian isolate; it clusters with Asian sequences, and is a non-previously described allele. These observations might indicate the introduction of PvMSP-1 alleles from Asian *P. vivax*. This possibility is not extraordinary given that Buenaventura is the main entrance for most of the Colombian market arriving through the Pacific Ocean. Inter- and intra-allelic recombination can be observed among the Colombian isolates, a factor that can be attributable to the active internal migration of *P. vivax*-infected patients and multiplicity of infection within the study area, which might be also the explanation why there is a low correlation between Colombian alleles and the geographical origin.

The distribution of Colombian isolates in the NJ tree resembled very well the allele groups identified. The former was built based exclusively in conserved blocks 1, 3, and 5; meanwhile, the allele is defined exclusively in the specific combination of variable blocks 2a, 2b, 2c, and 4. Such a specific combination of variable blocks (allele), in some cases, was associated with specific dimorphic substitution observed in conserved residues of either variable or conserved blocks. For example, 1) all isolates having the block 2c basic 3 with the substitution P/S in the fourth position were finally classified as being allele 33, meanwhile the remaining basic 3 (without the P/S) substitution were classified as alleles 6 or 20; and 2) a similar effect was observed with isolates D61 and B74, which always matched together for variable block 2a–c and shared nine exclusively linked dimorphic substitutions, five of them in conserved block 1 and four in variable block 4. Despite no clear geographical relationship, the similarity between them is so strong that it is possible to propose that the isolate from Buenaventura (B74), where malaria transmission does not occur, was acquired in La Delfina (D61), which is a near rural spot for local tourism with active malaria transmission.

Several non-previously described dimorphic substitutions and two new alleles were found suggesting that Pv200L from PvMSP-1 is under evolutionary forces non-equally distributed along the whole fragment. Such evolutionary forces, which might be related to the immune response, may have generated these new alleles by selecting new combinations generated by mitotic recombination in the asexual blood stages.

Besides the presence of highly conserved fragments in this protein, several other factors indicate that this protein could be a suitable target for a vaccine: First, this protein is highly antigenic as indicated by its recognition by the great majority of individuals from endemic communities in Brazil and Colombia.⁷ Second, it has been demonstrated that naturally acquired IgG antibodies directed to the N-terminal region of PvMSP-1 are associated with clinical protection to *P. vivax*-malaria.²⁵ Third, previously defined promiscuous T-helper

epitopes are located at highly conserved portions of Pv200L and displayed a high conservation pattern across the 68 sequences of Pv200L.

This is the first study of Pv200L polymorphism in Colombia. Although the number of isolates is limited, this is the PvMSP-1 polymorphism study with the highest number of *P. vivax* isolates from the same geographic origin. Such an approach has allowed us to confirm the mosaic structure of PvMSP-1 and its inter- and intra-allelic recombinant nature, to illustrate its specific association with dimorphic substitutions in conserved blocks, and describe the wide geographic distribution of highly conserved epitopes, despite the high level of polymorphism in this fragment. Further studies with a larger number of isolates from other endemic regions of the country and the planet would be required to assess genetic diversity with greater accuracy, linkage disequilibrium, and population structure. The presence of highly conserved blocks in this fragment of the *P. vivax* MSP-1 protein has important implications for the development of Pv200L as a subunit vaccine candidate.

Received September 2, 2009. Accepted for publication December 23, 2009.

Acknowledgments: We thank Ingrid Felger from the Swiss Tropical Institute (Basel, Switzerland) and Hernando del Portillo from CRESIB (Barcelona, Spain) for their advice and criticism, and Suzanne Fischer from Family Health Organization (USA) for her assistance in revising and editing this manuscript.

Financial support: This study was supported through a *P. vivax* vaccine research program granted by the National Institute of Allergy and Infectious Diseases (NIAID grant no. A1-49486/ TMRC), National Bureau of Sciences, University of Valle State (contract no. 245-2004), COLCIENCIAS (grant 1106-04-16489), and the Colombian Ministry of Social Protection (grant 2304-04-19524) (contract no.253-2005). Ananías A. Escalante is supported by the grant R01GM080586 from the National Institutes of Health, USA and through an International Center of Excellence for Malaria Research NIAID/ICEMR grant no U 19AI089702.

Authors' addresses: Augusto Valderrama-Aguirre, Evelin Zúñiga-Soto, Luz Ángela Moreno, Myriam Arévalo-Herrera, and Sócrates Herrera, Instituto de Immunología, Facultad de Salud, Universidad del Valle, Cali, Colombia and Malaria Vaccine and Drug Development Center, Cali, Colombia, E-mails: avalderrama@inmuno.org, ezuniga@inmuno.org, luzangelmo@hotmail.com, marevalo@inmuno.org, and sherrera@inmuno.org. Leonardo Mariño-Ramírez, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Computational Biology Branch, Building 38A, Bethesda, MD, E-mail: marino@ncbi.nlm.nih.gov. Ananías A. Escalante, School of Life Sciences, Arizona State University, Tempe, AZ, E-mail: Ananias.Escalante@asu.edu.

Reprint requests: Sócrates Herrera, Malaria Vaccine and Drug Development Center, Carrera 37 - 2Bis No. 5E - 08, Cali, Colombia, E-mail: sherrera@inmuno.org.

REFERENCES

- Guerra CA, Hay SI, Lucio-Parades LS, Gikandi PW, Tatem AJ, Noor AM, Snow RW, 2007. Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malar J* 6: 17.
- Mendis K, Sina BJ, Marchesini P, Carter R, 2001. The neglected burden of *Plasmodium vivax* malaria. *Am J Trop Med Hyg* 64: 97–106.
- Korenromp E, Miller J, Nahlen B, Wardlaw T, Young M, 2005. World malaria report. RBM/WHO/UNICEF, ed. Geneva, Switzerland: Roll Back Malaria, 1–214.
- Subdirección de Vigilancia y Control en Salud Pública INdS, 2006. Informe del VI período epidemiológico: malaria. SIVIGILA Semanas 1–24: 21–25.

5. Herrera S, Bonelo A, Perlaza BL, Fernandez OL, Victoria L, Lenis AM, Soto L, Hurtado H, Acuna LM, Velez JD, Palacios R, Chen-Mok M, Corradin G, Arévalo-Herrera M, 2005. Safety and elicitation of humoral and cellular responses in Colombian malaria-naïve volunteers by a *Plasmodium vivax* circumsporozoite protein-derived synthetic vaccine. *Am J Trop Med Hyg* 73: 3–9.
6. Polley SD, McRobert L, Sutherland CJ, 2004. Vaccination for vivax malaria: targeting the invaders. *Trends Parasitol* 20: 99–102.
7. Valderrama-Aguirre A, Quintero G, Gomez A, Castellanos A, Perez Y, Mendez F, Arévalo-Herrera M, Herrera S, 2005. Antigenicity, immunogenicity, and protective efficacy of *Plasmodium vivax* MSP1 Pv200L: a potential malaria vaccine subunit. *Am J Trop Med Hyg* 73: 16–24.
8. Genton B, Al-Yaman F, Betuela I, Anders RF, Saul A, Baea K, Mellombo M, Taraika J, Brown GV, Pye D, Irving DO, Felger I, Beck HP, Smith TA, Alpers MP, 2003. Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against *Plasmodium falciparum* in Papua New Guinean children. *Vaccine* 22: 30–41.
9. Guttinger M, Romagnoli P, Vandel L, Meloen R, Takacs B, Pink JR, Sinigaglia F, 1991. HLA polymorphism and T cell recognition of a conserved region of p190, a malaria vaccine candidate. *Int Immunol* 3: 899–906.
10. Putaporntip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho WG, 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 USA* 99: 16348–16353.
11. Putaporntip C, Jongwutiwes S, Tanabe K, Thaithong S, 1997. Interallelic recombination in the merozoite surface protein 1 (MSP-1) gene of *Plasmodium vivax* from Thai isolates. *Mol Biochem Parasitol* 84: 49–56.
12. Gonzalez JM, Olano V, Vergara J, Arévalo-Herrera M, Carrasquilla G, Herrera S, Lopez JA, 1997. Unstable, low-level transmission of malaria on the Colombian Pacific Coast. *Ann Trop Med Parasitol* 91: 349–358.
13. Snounou G, 1996. Detection and identification of the four malaria parasite species infecting humans by PCR amplification. *Methods Mol Biol* 50: 263–291.
14. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL, 2007. GenBank. *Nucleic Acids Res* 35: 21–25.
15. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
16. Kumar S, Tamura K, Nei M, 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5: 150–163.
17. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
18. Tajima F, 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
19. Fu YX, Li WH, 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
20. Nei M, Gojobori T, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.
21. Nei M, Kumar S, 2000. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.
22. Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
23. Caro-Aguilar I, Rodriguez A, Calvo-Calle JM, Guzman F, De la Vega P, Patarroyo ME, Galinski MR, Moreno A, 2002. *Plasmodium vivax* promiscuous T-helper epitopes defined and evaluated as linear peptide chimera immunogens. *Infect Immun* 70: 3479–3492.
24. 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 msp1 polymorphism in *Plasmodium vivax* and related simian malaria parasites. *Mol Biochem Parasitol* 156: 74–79.
25. Nogueira PA, Alves FP, Fernandez-Becerra C, Pein O, Santos NR, Pereira da Silva LH, Camargo EP, del Portillo HA, 2006. A reduced risk of infection with *Plasmodium vivax* and clinical protection against malaria are associated with antibodies against the N terminus but not the C terminus of merozoite surface protein 1. *Infect Immun* 74: 2726–2733.