Plasmodium vivax Duffy binding protein: a modular evolutionary proposal

P. MARTINEZ[1,2]†, C. F. SUAREZ[1]†, P. P. CARDENAS[1,2] *and* M. A. PATARROYO[1,2]*

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

(*Received 20 August 2003; revised 21 October 2003; accepted 23 October 2003*)

SUMMARY

The population of malaria-causing parasites is characterized by great genetic diversity. Knowledge of the polymorphism generation mechanism is a central issue for developing effective vaccines against malaria and understanding the parasite population structure. *Plasmodium vivax* genetic diversity has been explained in terms of two major factors: natural selection and intragenic recombination. A modular organization was found within *P. vivax* Duffy binding protein in the present work. Four Colombian isolates have identical sequences to Salvador-1 strain amongst *dpb* regions III–VI analysed, suggesting a high identity between Central and South American isolates. Geographically clustered sectors, corresponding to cysteine-rich regions (II and VI), show a high sequence diversity that could reflect a possible immune response evasion mechanism; both positive and negative selection were detected in these regions. In contrast, other *dbp* gene regions display a non-geographical clustering pattern, lower sequence diversity and predominant negative selective pressure. Recombination was homogeneously detected all along the molecule. These findings suggest that diversification vs. homogenizing forces, drive *dbp* gene evolution and determine its mosaic region organization.

Key words: *Plasmodium vivax*, genetic diversity, Duffy binding protein, geographical variation, cysteine-rich motifs, Colombian isolates.

INTRODUCTION

Malaria, caused by protozoan parasites belonging to the *Plasmodium* genus, is characterized by repeated cycles of parasite growth and subsequent host erythrocyte destruction. Each blood-stage generation produces motile merozoites that must recognize and enter new erythrocytes to guarantee parasite survival (Chitnis, 2001).

P. vivax triggers erythrocyte invasion through expressing several surface and apical proteins on the merozoite which bind to erythrocyte surface proteins (Adams *et al*. 1992; Barnwell, Nichols & Rubinstein, 1989). One well-characterized ligand-receptor interaction involves the Duffy binding protein (DBP) expressed on the *P. vivax* merozoite and its corresponding erythrocyte receptor, the Duffy Antigen Receptor for Chemokines (DARC) (Adams *et al*. 1992; Wertheimer & Barnwell, 1989). African populations lacking DARC erythrocyte expression are resistant to *P. vivax* merozoite infection (Miller *et al*. 1976). This suggests that the DBP-DARC interaction is critical for *P. vivax* erythrocyte invasion. No alternative erythrocyte invasion pathways have been identified to date (Barnwell & Galinski, 1995). The above-mentioned characteristics make *P. vivax* DBP one of the most promising vaccine candidates against vivax malaria.

P. vivax DBP is a 140 kDa microneme-secreted protein belonging to the large erythrocyte binding-protein (EBP) family. This family includes the *P. vivax*, *P. knowlesi* α, *P. knowlesi* β and *P. knowlesi* γ Duffy-binding proteins and *P. falciparum* sialic acid-binding protein, also known as EBA-175 (Adams *et al*. 1992). This gene family's common features include: similar multi-exon structures, conserved exon/intron boundaries, single copy genes and amino and carboxyl cysteine-rich domains (Adams *et al*. 2001; Adams *et al*. 1992). The *Plasmodium* Erythrocyte Binding Protein family structure is characterized by sequence similarity in one or several regions amongst paralogous and homologue species. The extra-cellular portion of this protein family has been divided into six regions according to their sequence homology. Members of the EBP family share two cysteine-rich domains, located in regions II and VI (Adams *et al*. 2001; Adams *et al*. 1992; Fang *et al*. 1991), containing conserved cysteines and hydrophobic amino acid residues.

An additional characteristic is the presence of a functional binding domain in the conserved N-terminal cysteine-rich region (region II) (Chitnis

* Corresponding author: Molecular Biology Department, 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
† Both of the above-mentioned made equal contributions as first author.

Parasitology (2004), **128**, 353–366.
DOI: 10.1017/S0031182003004773 Printed in the United Kingdom

& Miller, 1994). These functional domains are referred to as Duffy-binding-like (DBL) domains and are also present in *P. falciparum* (Adams *et al*. 2001). DBL domains are found in parasite ligands mediating two crucial processes in malaria pathogenesis: erythrocyte invasion and cytoadherence (Ranjan & Chitnis, 1999).

P. vivax DBP cysteine-rich region II (330 amino acids) has been implicated in the erythrocyte binding process (Chitnis & Miller, 1994; Wertheimer & Barnwell, 1989); this region contains critical binding motifs mapped between amino acids 291 and 460 (Ranjan & Chitnis, 1999). Previous studies have shown that although the positions of cysteine residues are conserved, other amino acids are highly polymorphic (Ampudia *et al*. 1996; Tsuboi *et al*. 1994), being subjected to positive selection pressure (Xainli, Adams & King, 2000). Different region II analysed haplotypes have similar adhesion to DARC-positive erythrocytes (Xainli, Adams & King, 2000), suggesting that such polymorphism does not significantly alter host–parasite binding. Most of the antibodies found in infected people recognize this region, more than other DBP regions (Michon, Fraser & Adams, 2000; Tsuboi *et al*. 1994).

Region VI has approximately 100 amino acids and contains 8 cysteine residues. These Cys residues may imply a particular structural arrangement associated with function (similar to what occurs in region II). The amino acid conservation amongst all region VI paralogues in other species also suggests an important structural-functional role (Adams *et al*. 1992).

Little is known about *P. vivax* DBP regions III–VI function and genetic variability. Cole-Tobian & King (2003) have calculated nucleotide diversity and selective pressure on *dbp* gene regions II–VI in a study including 12 Papua New Guinean (PNG) and 1 Salvadorian sequences.

The present study describes the genetic diversity, selective pressure, recombination sites and cluster analysis for the *P. vivax dbp* encoding gene in 23 Colombian isolates and their comparison with previously reported Korean, Salvadorian and PNG sequences.

MATERIALS AND METHODS

P. vivax *samples*

Blood samples from patients diagnosed as having *P. vivax* malaria by microscope analysis were collected for this study. Samples from 23 patients were extracted, anti-coagulated with EDTA, processed and stored for a 1-year period (2001–2002). These samples were obtained from areas to the east of the Andes mountain range in Colombia: Llanos Orientales (Guaviare, Villavicencio and nearby locations) and Norte de Santander (Cúcuta and neighbouring areas) (Fig. 1A).

DNA preparation

Plasma and leukocytes were removed after spinning. DNA was obtained by phenol-chloroform extraction, precipitated in isopropanol and hydrated in Tris–EDTA buffer for use as template in PCR.

PCR and fragment isolation

The Duffy binding protein gene was amplified with specific primers for regions III–VI (forward: 5′-GAT AAG TCA GCC TGT AGA TAG 3′ and reverse: 5′ AGC ACT TGT AGG ATG GAT C 3′, respectively) (Fig. 1B). The reaction mixture contained 10 mM Tris, 50 mM KCl, 3 mM $MgCl_2$, 200 nM deoxynucleotide triphosphates, 2·5 U *Taq* polymerase and 10–40 ng DNA template at a final volume of 25 μl. The thermal profile used was: one 5-min cycle at 95 °C and 35 cycles of 1 min at 95 °C, 1 min at 50 °C and 1 min at 72 °C.

PCR products were purified with Wizard PCR preps (Promega Corp, Madison, WI) and cloned into pGEMT-Easy Vector System (Promega Corp, Madison, WI). DNA sequencing was performed by the dideoxynucleotide method in an ABI Prism 310 Automated Sequencer (Applied Biosystems, Foster City, CA) in both directions. Three clones were sequenced per isolate, each one came from an independent PCR amplification. For those sequences belonging to the same isolate but displaying differences among them, two extra clones were sequenced to discard errors. Clinical isolate nucleotide sequences were compared to *P. vivax* Sal-1 (GenBank Accession number M61095).

Sequence analysis

Colombian *P. vivax* Duffy binding protein region III–VI sequences were analysed, as well as previously reported region I–VI sequences from Korea (GenBank Accession numbers AF215737-38 (Kho *et al*. 2001) and AF220657-68 (Suh *et al*. 2001)), Papua New Guinea (AF469515-602) (Cole-Tobian *et al*. 2002), L23069-75 and U10103-07 (Tsuboi *et al*. 1994) (AF289480-83, AF289635-53 and AF291096) (Xainli, Adams & King, 2000) and Colombian regions I–II sequences, previously identified by our group, plus the Belem strain (Brazil) sequence for this region (U50575-91) (Ampudia *et al*. 1996).

Preliminary sequence alignment and sequence comparison were performed using ClustalX software (Thompson *et al*. 1997). Later, the sequences that possessed an intraspecific identity value higher than 99%, detected by GeneDoc software (Nicholas, Nicholas & Deerfield, 1997), were eliminated.

MEGA 2.1 software (Kumar *et al*. 2001) was used for parsimony (Fitch, 1971) and distance methods for sequence tree construction. The distance methods included Minimum-Evolution and Neighbour-Joining (Rzhetsky & Nei, 1993). Genetic distances

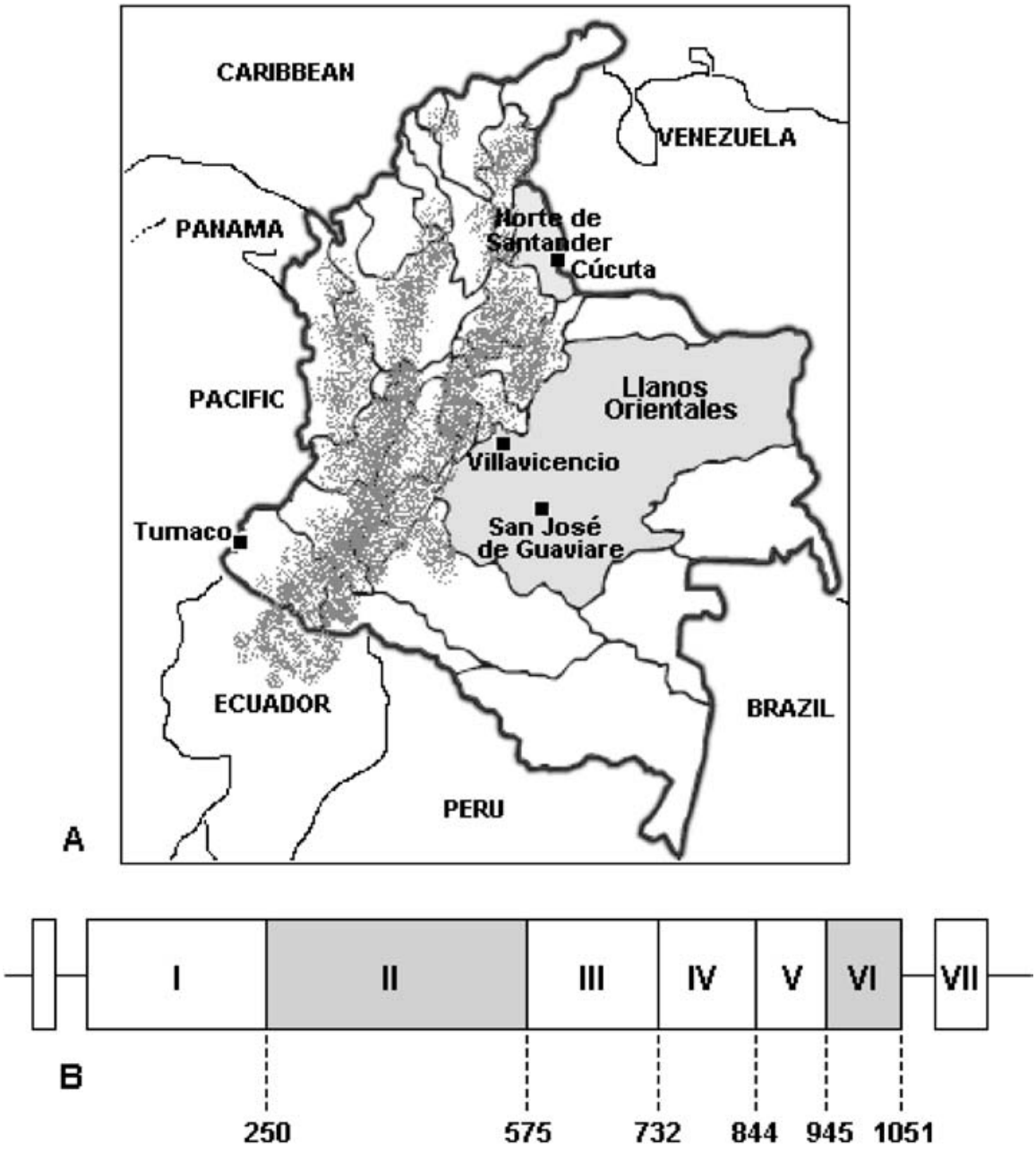

Fig. 1. (A) Map of Colombia for studied locations. Shaded parts of the map represent the areas from which patients came who donated the infected blood samples; the black squares represent sample reception centres; stippled areas represent mountain ranges. (B) *Plasmodium vivax dbp* region structure. Grey represents cysteine-rich regions; numbers represent amino acid location, according to Fang's subdivision of regions (Fang *et al*. 1991).

were estimated by using P and Tamura 3-parameters (used to correct the *Plasmodium* genome A+T composition bias) (Tamura, 1992), distances for nucleotide and P and Poisson-corrected distances for amino acid sequences. Bootstrap analyses (Hillis, 1993), with 1000 replicates each, were used to assign confidence levels to branch nodes.

Several diversity parameters were determined for regions I–VI from all samples (Colombia, Papua New Guinea and Korea). π (nucleotide diversity) was calculated with DnaSP 3.51 software (Rozas & Rozas, 1999), while amino acid sequence diversity was estimated with p-distance (with pairwise deletion). Other parameters considered were the proportion of variant positions (V.S.) and, for these positions, the proportion of singleton positions (S) and the proportion of Parsimony Informative sites (P.I.). These parameters were calculated with MEGA 2.1 software (Kumar *et al*. 2001). The proportion of variant positions with conservative (C) and non-conservative changes (NC) was also considered for amino acid sequences. The set of sequences for calculations varied, according to availability (not all the sequences spanned the same regions).

The nature of amino acid replacement (*conserved* or non-*conserved*) was examined using the PAM250 similarity matrix (Dayhoff, Schwartz & Orcutt, 1978). Zero or positive values indicate that variable

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

amino acid residues can be functionally interchanged, whilst negative values indicate that it is not common for two amino acids to become replaced in nature. GeneDoc software was used to determine basic physicochemical properties (such as polarity, hydrophobicity, volume and charge) for amino acids in region VI (Nicholas *et al*. 1997). In addition, a Kyte–Doolittle hydropathy plot for region VI sequences from Colombia and Papua New Guinea was carried out (Kyte & Doolittle, 1982). We used the minimum number of recombination events test (RM) when searching for evidence of recombination (Hudson & Kaplan, 1985), included in DnaSP 3.51 software (Rozas & Rozas, 1999).

Natural selection was calculated with MEGA v.2.1 software using the Nei and Gojobori method that compares the difference between non-synonymous and synonymous substitutions. Positive values for this difference indicate positive selection, whilst negative ones indicate negative selection (Nei & Gojobori, 1986). Tajima's test (statistic D), also included in DnaSP 3.51 software (Rozas & Rozas, 1999), was also used. This test is based on the neutral model prediction, estimating nucleotide diversity based on the relationship between the number of segregating sites and the average number of nucleotide differences from pairwise comparison. A positive D value indicates possible balancing selection or population subdivision. A negative value suggests recent directional selection, a population bottleneck, or purifying selection (Tajima, 1989).

Amino acid sequences from different malarial species were used for comparing EBP region VI: *P. cynomolgi* EBP (T18373), *P. knowlesi* EBP beta (P50493), *P. knowlesi* EBP gamma (P50494), *P. knowlesi* EBP alpha (P22545), *P. falciparum* MAEBL (NP_701342), *P. vivax* MAEBL (AAL10508), *P. yoelii* EBP (T09127), *P. falciparum* EBA175 (P19214), *P. falciparum* EBP (AAM28887), *P. reichenowi* EBP (CAB96159), *P. berghei* EBL (AAC47736), *P. knowlesi* MAEBL (AAL10507), *P. falciparum* EBL-1 like (NP_703228), *P. chabaudi* EBL (AAC47737), *P. vinckei vinckei* EBL (AAC47738), *P. falciparum* EBP2 (AAL75954) and *P. falciparum* BAEBL (AAK49521).

RESULTS AND DISCUSSION

Nineteen new sequences of dbp *regions III–VI were obtained*

A single 1388 bp band was amplified. Nineteen different sequences are being reported here amongst the 23 Colombian isolates analysed, corresponding to Sal-1 strain sequence amino acid positions 582–1043 (Regions III to VI). Only one type of sequence was found for each isolate. Four sequences were identical to Sal-1 *dpb*, while the remaining 19 were geographically grouped and consecutively numbered within each region as follows: COLns01-COLns08 (Norte de Santander) (GenBank accessions AY341900-07), COLlla01-COLlla11 (Villavicencio, Llanos Orientales) (GenBank accessions AY341889-99). Nucleotide (ALIGN_000583) and amino acid (ALIGN_000584) alignments were reported in the EMBLALIGN database (http://www.ebi.ac.uk) (Lombard *et al*. 2002).

Region I comprised residues 1 to 104; region II, residues 105 to 429; region III, 430 to 595; region IV, 596 to 707; region V, 708 to 808; and region VI, residues 809 to 901.

Fig. 2 shows the amino acid alignment for *P. vivax* DBP regions III–VI including new Colombian isolates. One sequence (COLns08, from Norte de Santander, Colombia) had a 27 pb (9 amino acid – **FAESTKSAE**) insertion in the middle of the central hydrophilic region (at the beginning of region IV); this insertion has been previously reported in some Papua New Guinean (Tsuboi *et al*. 1994) and Korean (Suh *et al*. 2001) *dbp* sequences. The inserted sequence is a tandem repeat created by partial duplication of the sequence immediately following the insertion site and is not repeated in any other part of the gene (Tsuboi *et al*. 1994).

Cluster and recombination analysis of P. vivax *DBP regions*

Sequences from Papua New Guinea, Salvador, Korea and Colombia were employed for cluster analysis. These sequences were divided into regions I–VI, according to the region delimitation previously reported by Fang (Fang *et al*. 1991) (Fig. 1B).

Fig. 3 shows the amino acid cluster trees constructed for each *dbp* region by the Neighbour-Joining method with 1000 bootstrap replicates. Maximum Parsimony and Minimum-Evolution methods also resulted in similar tree topology (not shown); nucleotides and amino acids show equivalent results. Non-significant bootstrap values were observed in all trees. This is explained by the scarcity of variant sites and the scattered pattern of the variation shown among *dbp* sequences (the majority of variant sites were singletons, see Table 1). These low bootstrap values are common to other studies with *P. vivax* antigens: AMA and MSP1 genes (Figtree *et al*. 2000), TRAP (Putaporntip *et al*. 2001), MSP3α (Rayner *et al*. 2002) and MSP1 (Putaporntip *et al*. 2002).

No clear geographical clustering was found amongst the Colombian samples; this might indicate constant intermingling of *P. vivax* in eastern Colombia (Fig. 2). Colombian and Salvador sequences frequently clustered together; moreover, 4 Colombian sequences found were identical to Sal-1 sequence. These observations indicate a close relationship between Central American and Colombian strains. The sequence from Belem strain

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

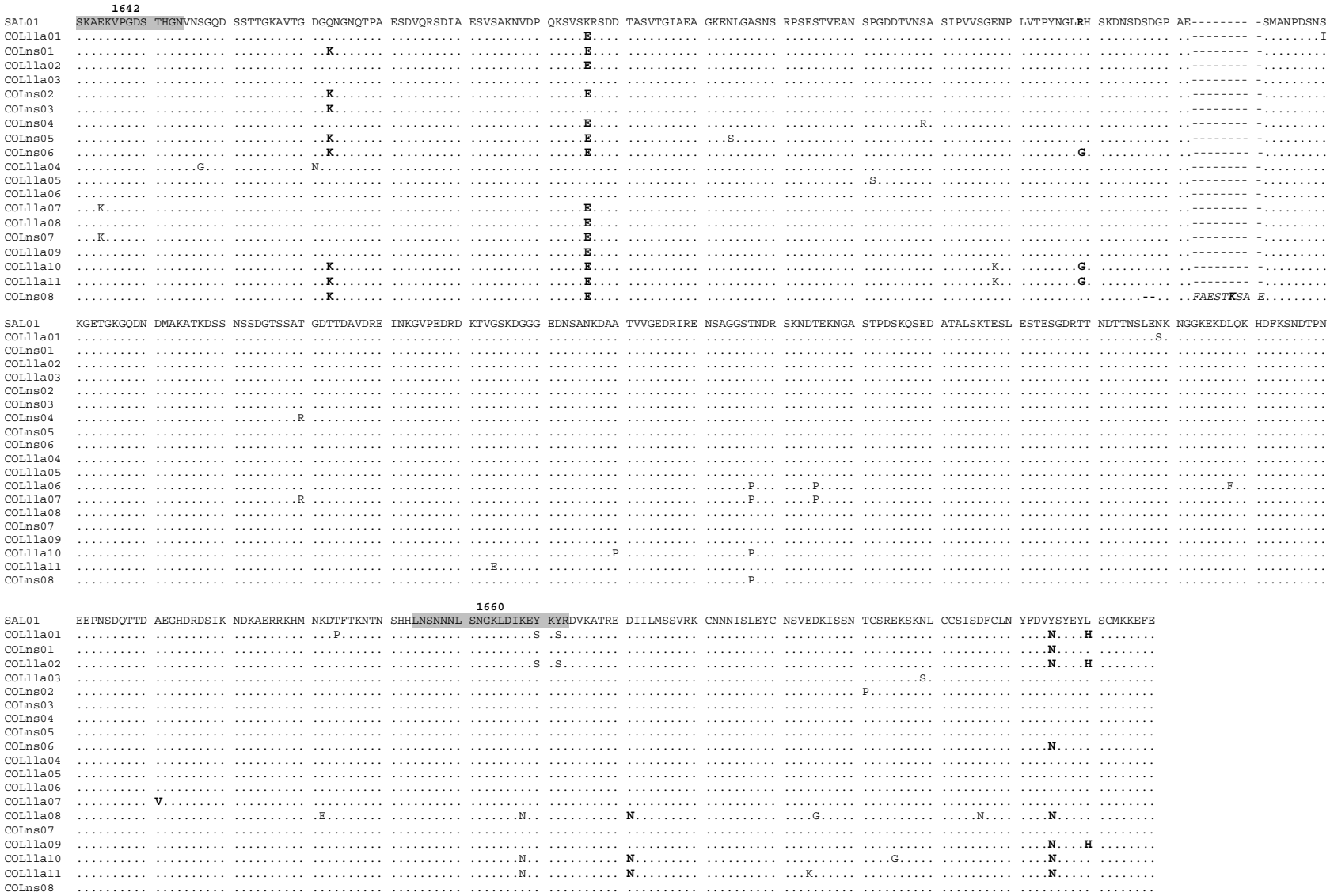

Fig. 2. Colombian isolate DBP region III–VI amino acid alignment. Grey shading indicates specific high binding affinity peptides (1640 and 1642); bold represents co-variant residues with Korean and Papua New Guinean sequences; italics show insertion in COLns08 sample. Dots indicate conserved residues and dashes represent gaps introduced for alignment.

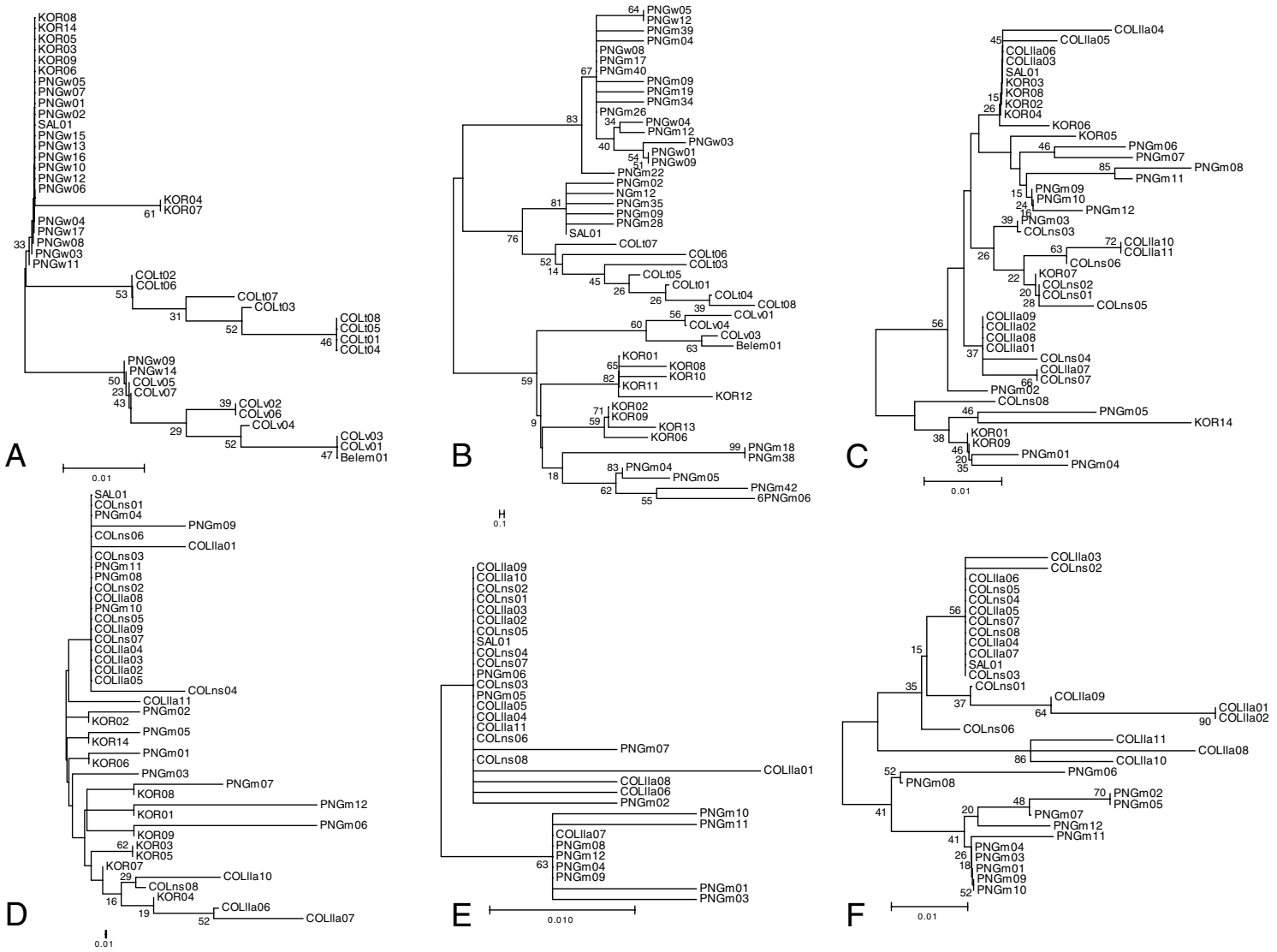

Fig. 3. Neighbour-Joining tree of amino acid sequences from *Plasmodium vivax* DBP regions (I–VI) from endemic populations based on pair-wise differences using P distance. The significance of the nodes was proved by Bootstrap calculated with 1000 replicates. The scale indicates 0·01 substitutions per site. (A) Tree for region I. (B) Region II. (C) Region III. (D) Region IV. (E) Region V. (F) Region VI. Colombia, COL; Norte de Santander, ns; Llanos Orientales, lla; Korea, KOR; Papua Nueva Guinea, PNG; Wosera, W; Madang, m.

Table 1. Diversity analysis of *dbp* gene

(n, number of sequences sampled; Sites, number of sites included; V.S., number of variant positions; S, number of singleton positions, P.I., number of parsimony informative sites; π, average number of nucleotide substitutions in DNAsp 3·51; C, conservative change in amino acid; NC, non-conservative change in amino acid; d, average number of sequence diversity for amino acid with p-distance.)

			Nucleotide					Amino Acid						
Origin	Region	n	Sites	V.S.	S	P.I.	π	Sites	V.S.	S	P.I.	C	NC	d
Colombia	RI	15	234	5	0	5	0·0102	78	5	0	5	4	1	0·0308
	RII	15	915	14	1	13	0·0062	305	14	1	13	12	2	0·0186
	RIII	19	456	18	13	5	0·0070	152	13	8	5	11	2	0·0171
	RIV	19	336	10	7	3	0·0044	112	6	3	3	4	2	0·0095
	RV	19	303	6	6	0	0·0020	101	5	5	0	5	0	0·0052
	RVI	19	279	14	8	6	0·0096	93	12	6	6	9	3	0·0267
PNG	RI	17	252	1	0	1	0·0009	84	1	0	1	1	0	0·0026
	RII	71	975	104	77	27	0·0130	325	65	43	22	48	17	0·0281
	RIII	12	498	26	15	11	0·0148	166	18	10	8	16	2	0·0310
	RIV	12	336	18	17	1	0·0093	112	13	13	0	11	2	0·0193
	RV	12	303	16	15	1	0·0100	101	7	6	1	4	3	0·0147
	RVI	12	279	13	10	3	0·0102	93	6	3	3	3	3	0·0178
Korea	RI	8	195	1	0	1	0·0021	65	1	0	1	1	0	0·0068
	RII	14	975	69	59	10	0·0145	325	48	9	39	33	15	0·0325
	RIII	10	498	20	12	8	0·0128	166	11	6	5	9	2	0·0212
	RIV	10	123	1	0	1	0·0025	41	1	0	1	1	0	0·0082
All	RI	42	252	7	0	7	0·0074	84	6	0	6	5	1	0·0210
	RII	102	975	165	124	41	0·0129	325	104	33	71	72	32	0·0314
	RIII	42	498	46	31	15	0·0116	166	30	18	12	27	3	0·0246
	RIV	42	336	29	24	5	0·0054	112	20	16	4	16	4	0·0116
	RV	32	303	21	20	1	0·0055	101	11	10	1	8	3	0·0103
	RVI	32	279	26	18	8	0·0136	93	17	9	8	12	5	0·0342

(Brazil), only available for regions I and II, clustered with Villavicencio strains (Fig. 3A, B). These observations taken together suggest continuous dispersion of strains throughout the whole of the central and Southern American regions via Colombia, constituting a bridge between these two regions.

A geographical clustering pattern was only clearly observed in region VI (Fig. 3F). Region II topology (Fig. 3B) was similar to that reported by Cole-Tobian & King (2003), showing a pattern having a low degree of mix between sequences from Papuan and American strains. Regions I, III, IV and V show more mixing among strains from different geographical locations (Fig. 3A, C, D and E respectively).

This pattern (i.e. protein regions showing geographic clustering and others that do not) highlights the modular separation between protein regions, indicating the different extent of host specialization in these modules. This tendency was shared with other *P. vivax* antigens such as MSP-1 (Putaporntip *et al*. 2002) and MSP-3α (Rayner *et al*. 2002). The phenomena originating these modules could also be based on genetic drift, recombination processes and different selective immune pressures.

Sequences with a characteristic insertion at the end of region III do not show a unique cluster, indicating that there are characteristics in the rest of region III sequences driving the clustering pattern. This sector may be considered as evidence of a recombination process, or may represent a convergence event between Papuan and Colombian sequences. Evidence of recombination was found by using DnaSP 3.51 software (Rozas & Rozas, 1999). There were 9 recombination events (Rm) in region II between nucleotides (667, 676) (676, 707) (716, 718) (718, 722) (734, 737) (835, 847) (903, 904) (905, 1073) (1073, 1096) and 7 Rm in regions III–VI (1426, 1525) (1550, 1714) (1714, 1738) (1898, 2068) (2092, 2321) (2321, 2497) (2500, 2678). Only region I did not present recombination sites. Regions II and III displayed the highest recombination events in the *dbp* gene. This agrees with the recombination rate observed in *P. vivax* (Putaporntip *et al*. 2002; Rayner *et al*. 2002).

Geographical clustering has been reported in other *P. vivax* genes. For example, some evidence has been shown in the circumsporozoite protein (CSP) (Escalante *et al*. 2002; Qari *et al*. 1992) concerning the existence of strains having a specific geographical location. Polymorphic variants have been found in tropical regions of Brazil, Southeast Asia and Papua New Guinea (Escalante *et al*. 2002; Qari *et al*. 1992), but strikingly different strains have been described in China and Korea (Kim *et al*. 2002; Mann *et al*. 1994). This wide distribution of strains can be explained by the world's population mobility favouring the spread of different parasite strains.

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

Diversity analysis of P. vivax *DBP sequences*

Table 1 shows the calculations for the different diversity parameters used, sorted by region studied and geographical location. Different diversity patterns were observed between the geographical regions analysed.

Colombian sequences showed higher nucleotide diversity in region I than Papuan and Korean samples; on the other hand, Colombian isolates showed lower nucleotide diversity in all the other regions with respect to Papuan samples, displaying the lowest nucleotide diversity values for regions II and III with respect to Papuan and Korean samples. Korean samples showed higher nucleotide diversity in region II, but the lowest nucleotide diversity in region IV with respect to Papuan and Colombian samples. Altogether, Papuan samples were more diverse at the nucleotide level than Colombian and Korean samples. The diversity pattern was more similar between Papuan and Korean samples than Colombian samples. Sequence diversity at the protein level showed a similar pattern; the only difference was that Colombian samples showed higher sequence diversity in region VI. Calculations from all sets of sequences showed that *dbp* regions II and VI were the most variable; regions IV and V were the least variable. Region II was divided into portions described as being crucial for erythrocyte binding (II-1, II-2, II-3) (Ranjan & Chitnis, 1999; Singh *et al*. 2003). The highest diversity exhibited inside region II was observed within segment II-2 (comprising Cys5 to Cys8) (data not shown); this agrees with previous studies where the specific erythrocyte binding site displays higher diversity (Ampudia *et al*. 1996; Tsuboi *et al*. 1994; Xainli, Adams & King, 2000).

All changes observed in region I corresponded to parsimonious positions, when all sets of sequences are considered, or when sorted by geographical region. Several variant positions showed co-variation, exhibiting the same amino acid substitutions (ALIGN_000584). This might indicate that the changes in this region could be the result of similar selective pressure. Moreover, most of these changes were conservative. This pattern of higher conservative than non-conservative substitutions was shared by all regions in *dbp*.

Papuan and Korean samples showed the highest proportion of variant sites in Region II; however, the nature of the variation differed amongst them. Papuan samples showed a higher number of singletons than parsimony positions, whereas Korean samples showed the opposite tendency. Colombian samples showed the lowest proportion of variant sites in this region, also showing more parsimonious than singleton sites, like the Korean samples. A high proportion of singletons indicated a disperse substitution pattern, possibly searching for greater sequence diversity (for immune evasion or intra-specific competence). Thus, diversity generation may play a more major role in Papuan samples than in Colombian and Korean samples. On the other hand a more localized variation pattern (estimated by the proportion of parsimonious sites) might be a response to a different type of pressure, acting in a more focused way on specific sectors of the molecule, as observed in Colombian and Korean samples.

Region III showed the lowest difference in variant positions amongst geographical samples; this homogeneity extended to singletons and parsimonious positions; more singleton positions were detected than parsimonious. This region's variability was noticeable for Papuan and Korean samples, both showing the same behaviour as Region II in Papuan samples. Papuan and Colombian Regions IV and V showed a high proportion of singletons and low parsimonious positions. This behaviour was also similar to that observed for region II in Papuan samples, but with less sequence variability in Colombian samples.

Region VI showed different behaviour; the number of singletons and parsimonious informative positions was very similar; Colombian and Papuan samples showed very similar values for variant sites and sequence diversity at nucleotide level, but at amino acid level, Colombian sample values were higher. This indicates that Colombian samples have a more intense process of protein diversification than Papuan samples, although nucleotide diversity is equivalent in both geographical regions.

The overall calculations of the variant sites for each protein region show that their values were always higher than the values from each geographical location in every case. This indicates that the majority of changes were present in positions that did not co-vary and each isolate had a unique mode of variation.

In total, 27 co-variant residues among Colombian, PNG and Korean samples were found when compared to Sal-1 (ALIGN_000584: amino acid position 46, in region I; positions 149, 226, 239–41, 245, 272, 279, 292 and 358 in region II; 476, 509, 572 and 586–94 in region III; 774 in region V; and 834, 888 and 893 in region VI).

The diversity pattern was consistent with observations regarding transmission rates seen in the different geographical regions. In other words, Papua regions suffer stronger rates of transmission and prevalence of *P. vivax* infection than Colombia, and this behaviour was related to high sequence diversity (Cui *et al*. 2003). This means that Colombia has low transmission rates and, therefore, Colombian samples show lower sequence diversity than Papuan isolates.

Tendency towards a higher number of transversion changes rather than transitions, previously described in genes encoding *P. falciparum* membrane proteins due to the A+T composition bias, was observed in all *P. vivax dbp* gene regions; transition

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

Table 2. Nei–Gojobori's and Tajimas's tests for *dpb* gene

(n, number of sequences sampled; Sites, number of sites included; dS, synonymous substitution; dN, non-synonymous substitution. For dN–dS in Nei–Gojobori's test and Tajima's test, positive values correspond to positive selection pressure, and negative values correspond to negative selection pressure. Bold represents significant values $P<0{\cdot}05$. Italic and underlined letters represent significant values $P<0{\cdot}01$.)

Origin	Region	n	Sites	dS	dN	dN–dS	Tajima test
Colombia	RI	15	233	0·000	0·195	**0·195**	1·892
	RII	15	914	0·000	0·024	**0·024**	1·559
	RIII	19	498	0·011	0·038	**0·028**	−1·612
	RIV	19	336	0·009	0·023	0·013	−1·705
	RV	19	303	0·017	0·021	0·004	−*2·046*
	RVI	19	279	0·000	0·065	**0·065**	−1·395
PNG	RI	17	312	0·000	0·016	0·016	−0·491
	RII	71	975	0·004	0·032	**0·028**	−*2·036*
	RIII	12	498	0·071	0·085	0·014	−0·851
	RIV	12	236	0·029	0·048	0·018	−**2·065**
	RV	12	303	0·053	0·060	0·007	−*1·885*
	RVI	12	279	0·034	0·047	0·013	−1·673
Korea	RI	8	194	0·000	0·032	0·032	0·334
	RII	14	975	0·018	0·043	**0·025**	−*1·954*
	RIII	10	498	0·026	0·062	−0·005	−0·770
	RIV	10	122	0·000	0·022	0·022	ND
All	RI	42	312	0·000	0·123	**0·123**	0·282
	RII	102	975	0·006	0·037	**0·031**	−**2·199**
	RIII	42	498	0·042	0·062	0·020	−1·682
	RIV	42	336	0·020	0·030	0·010	−**2·245**
	RV	32	303	0·030	0·042	0·012	−**2·353**
	RVI	32	279	0·013	0·085	**0·072**	−1·485

value was 4·97 and transversion value 5·39. This tendency was maintained throughout geographical locations for all regions. Although transitional substitutions are more common as transversions (Collins & Jukes, 1994), *P. vivax* genome exhibiting a strong codon bias may affect the transition/transversion ratio and estimates of synonymous and non-synonymous substitutions (Escalante, Lal & Ayala, 1998).

The *dbp* gene has lower nucleotide diversity values ($\pi=0{\cdot}01224$) than other *P. vivax* surface proteins, such as AMA1 ($\pi=0{\cdot}0174$), MSP1 ($\pi=0{\cdot}0451$) (Figtree *et al*. 2000), but higher than TRAP ($\pi=0{\cdot}00594$) (Putaporntip *et al*. 2001). This may reflect a major functional restriction in DBP protein, due to its crucial role in the invasion process.

These values are considerably lower than those for *P. falciparum* surface proteins like MSP-1 ($\pi=0{\cdot}08792$), MSP-2 ($\pi=0{\cdot}04409$) and AMA-1 ($\pi=0{\cdot}01635$) (Escalante *et al*. 1998). *P. vivax dbp* is also more diverse than EBA-175 ($\pi=0{\cdot}00366$), its homologue in *P. falciparum* (Escalante *et al*. 1998), a value that might explain *P. vivax*'s dependence on genetic variability for immune response evasion, as opposed to *P. falciparum*'s reliance on alternative invasion pathways. The differences in the diversity pattern between *P. vivax dbp* and other non-homologous proteins in *P. falciparum* might reflect differences in invasion roles, virulence and number of clonal generations.

Evidence of selection in P. vivax *DBP regions*

Several tests were performed to determine the type of selective pressure on *dbp*: Nei–Gojobori's (Nei & Gojobori, 1986), and Tajima's (Tajima, 1989). Nei–Gojobori's test was computed by considering either all positions or just variable positions. All positions were used for Tajima's test. All *dbp* sequences used were those reported in the GenBank database, together with our *dbp* sequences for regions III–VI. Table 2 shows the analysis for all the studied regions (I–VI). Fu-Li's (Fu & Li, 1993) test was also performed, producing similar results to Tajima's test (data not shown).

Tajima and Nei–Gojobori's tests showed distinct tendencies. Nei–Gojobori's test for all positions showed mainly positive selective pressure both when each geographical region was considered individually and also when all sequences were analysed altogether; this positive selection was statistically significant only in region II, except for Colombian samples (data not shown). This tendency to present more non-synonymous than synonymous substitutions has been previously observed in other reports (Ampudia *et al*. 1996; Tsuboi *et al*. 1994; Xainli, Adams & King, 2000). If only variant positions are used, more regions than just region II show positive selection for all sets of sequences (Table 2). Tajima's test showed mainly negative selection pressure in the *dbp* regions considered, this being significant for

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

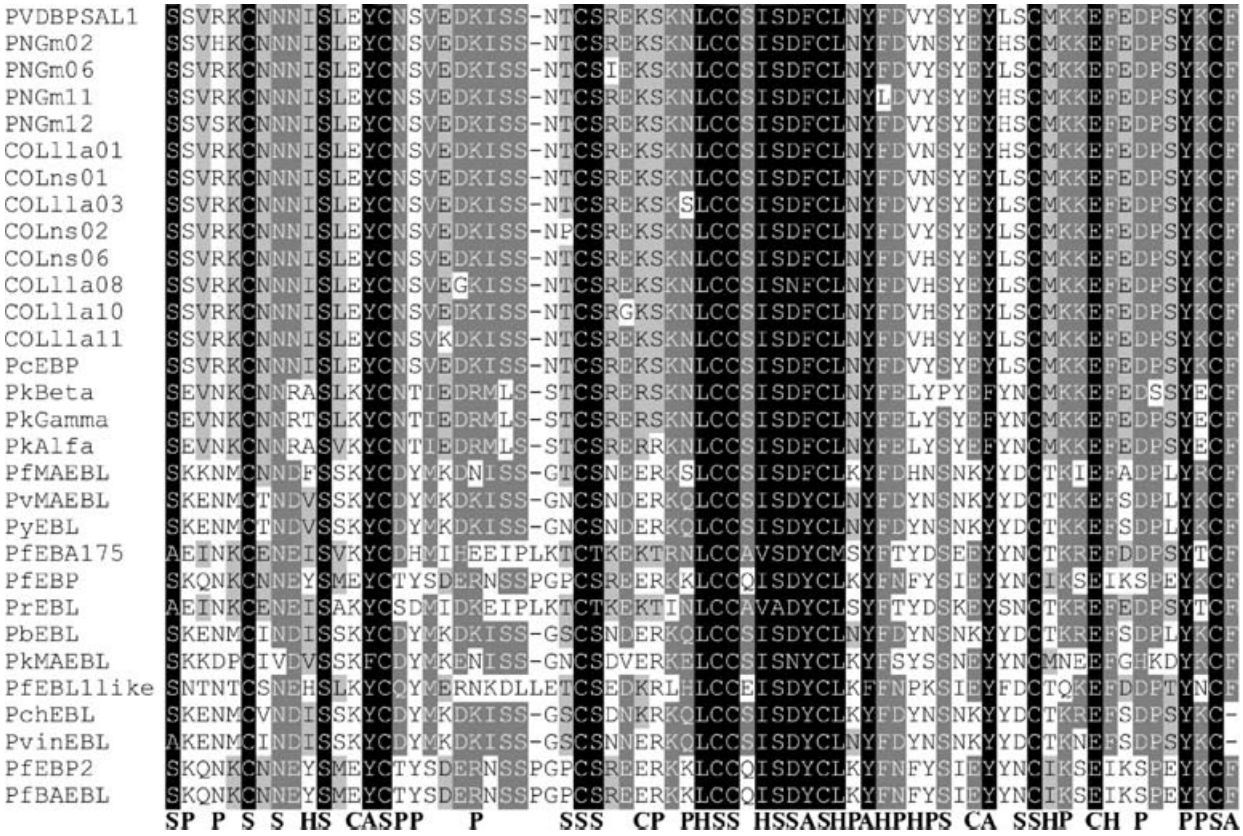

Fig. 4. Conserved carboxyl cysteine-rich alignment of the erythrocyte binding proteins. The shading indicates levels of identity among sequences. White letters on a black background represent amino acid residues that were 100% identical; white letters on a grey background represent amino acid residues that were 80% identical; and black letters on a grey background represent amino acid residues that were 60% identical. Bold letters represent conserved physicochemical properties: Small, S; Polar, P; Hydrophobicity, H; Charged, C; Aromatic, A; *P. vivax* Duffy Binding Protein, PVDBP; Madang Papua New Guinea, PNGm; Colombia, COL; Norte de Santander, ns; Llanos Orientales, lla; erythrocyte binding protein, EBP; erythrocyte binding antigen, EBA; erythrocyte binding like, EBL; *P. cynomolgy*, Pc; *P. knowlesi*, Pk; *P. vivax*, Pv; *P. falciparum*, Pf; *P. yoelii*, Py; *P. reichenowi*, Pr; *P. berghei*, Pb; *P. chabaudi*, Pch; *P. vinckei*, Pvin. Dashes represent gaps introduced to maintain alignment.

all sets of sequences in regions II, IV and V. Cole-Tobian *et al*. (2003) showed the same tendencies for region II, where opposite results were found according to the test used (Table 2). The contrasting detection of both negative and positive selection pressures in several of the molecule's regions (especially in regions II and VI) may imply a balance between a high functional restriction for maintaining structural constraint due to DBP's key role in the erythrocyte invasion process and the search for diversity as a response against host immune system pressure (as occurs in other *Plasmodium* surface antigens (Escalante *et al*. 1998; Putaporntip *et al*. 2001; Verra & Hughes, 2000)).

The differences between the selective pressures in samples from geographical regions showed that Colombian isolates possess a diversifying selection in more molecule regions than Papuan and Korean isolates. This might reflect a true variation between isolates, in which case, Colombian parasites suffer a stronger selective stress than Papuan and Korean parasites. Although these differences between geographical samples were also sustained by the patterns of diversity observed, these results might reflect a sample size effect and the method used for isolating *dbp* genes (PCR or DNA library). More sampling will be necessary for clarifying this issue.

Analysis of the 3′ cysteine region

A special approach was used for region VI to define conserved residues and physicochemical properties, due to this region's particular characteristics suggesting its possible relevance in vaccine design: the interspecies conserved cysteine arrangement; the presence of one high-activity reticulocyte binding peptide (1660) (Fig. 2) (Ocampo *et al*. 2002) and its diversifying and selection pattern, suggesting its exposure to the host immune system.

A clear conservation tendency could be seen in the general alignment of Colombian samples, particularly in cysteines and aromatic amino acids in region VI (Fig. 2).

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

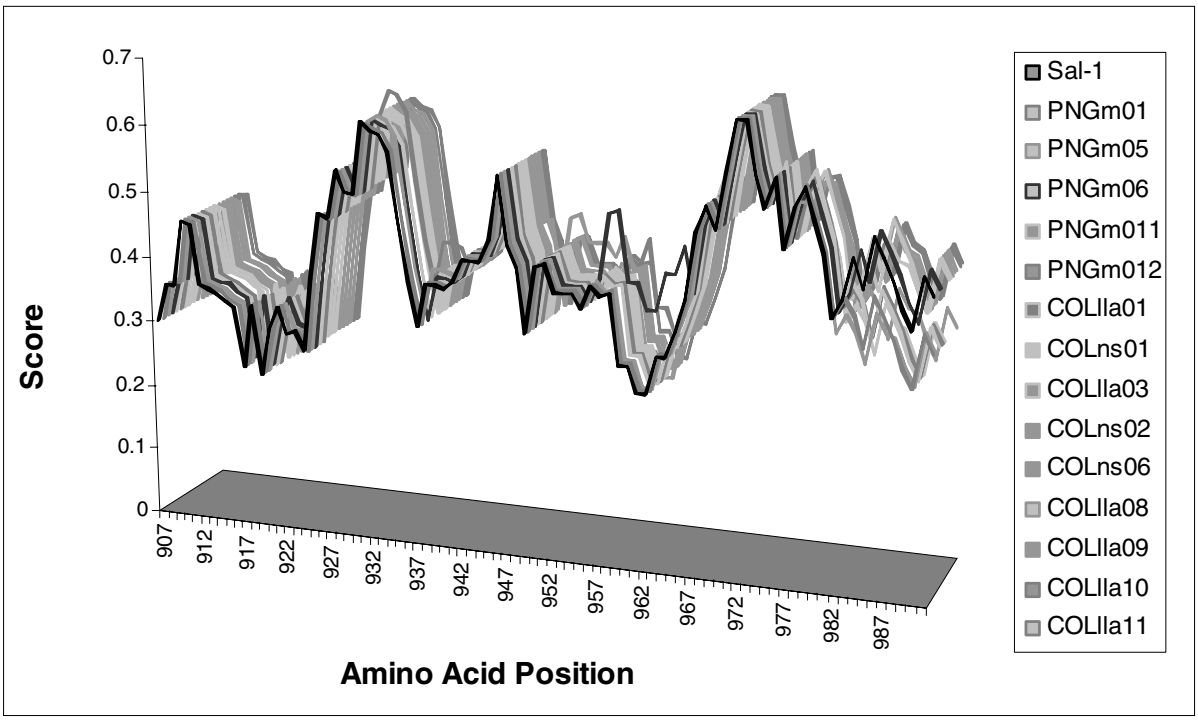

Fig. 5. Kyte–Doolitle Hydropathy Plot of region VI (3 cys) for variant sequences from Papua New Guinea and Colombia. Score corresponds to normalized (to 1) values used in the computation. Position corresponds to amino acid numeration according to Sal-1 sequence (Fang *et al*. 1991).

Region VI presented the highest geographical diversity (Table 1), although it possessed the lowest number of variable sites, suggesting that this region was highly conserved. This conservation pattern might indicate a key role for these residues in maintaining protein tertiary structure or other undetermined functions (Michon *et al*. 2002).

Colombian and Papua New Guinean sequences were chosen for specifically determining which residues were conserved and some of their physicochemical properties. EBL family proteins from different species were also included in this analysis (see Materials and Methods section).

Fig. 4 shows region VI sequence's protein alignment. High conservation levels (around 67% or higher) were observed, as well as the previously described 8 cysteine residues and hydrophobic amino acids, like tyrosine, leucine, phenylalanine and isoleucine (Adams *et al*. 1992).

Five physicochemical properties proved to be conserved following analysis with GeneDoc software. Size was the most conserved factor (17/70 residues), followed by polarity (14/70), hydrophobicity (5/70) and aromatic residues (6/70); overall, physicochemical properties were conserved in 68% of total residues.

Kyte–Doolittle hydrophobicity variation analysis was done for every position in region VI (Fig. 5) to see the effect of amino acid substitution on the region's hydrophobic characteristics (Kyte & Doolittle, 1982). The tendency towards maintaining hydrophobic characteristics that could be observed in this region was probably related to its function.

The results from hydrophobicity analysis of variant sequences led to the conclusion that region VI was a diverse region as presented in Table 1, having a high number of non-synonymous substitutions that did not alter its general hydrophobic characteristics.

The presence of conserved hydrophobic residues makes this group's recombinant expression very difficult due to its insolubility (Fraser *et al*. 1997), a fact which has paralysed functional assays aimed at describing region VI's role in the *P. vivax* invasion process.

The role of *P. falciparum* EBA-175 recombinant protein excluding the carboxy-terminal cysteine-rich region (C-cys), transmembranal and cytoplasmic domains, has been indirectly evaluated in preceding studies. These studies determined that the whole protein is not necessary during the invasion process or for parasite growth, but that this domain might be involved in intracellular mobilization processes and the use of alternative invasion pathways. Such domains are probably indispensable for the protein's structural conformation, facilitating its binding to erythrocytes. This preliminary evidence suggests that these C-cys domains are involved in the invasion process, not directly related to adhesion but being capable of transmitting or receiving signals from the interaction of the binding regions (EBL) with receptors on the target cell (Reed *et al*. 2000).

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press

The presence of DBL and C-cys domains in several different *Plasmodium* species' molecules suggests that they are performing a diverse role in receptor-ligand interactions. The residues conserved within these domains can be useful in identifying crucial determinants for cell–cell interaction during malaria.

Concluding remarks

Three forces in conflict (functional restriction, adaptive response to immune pressure, and contingency phenomena), represented in processes like diversifying selection, purifying selection, recombination and functional restriction, drive the evolution of DBP. These forces operate with different intensity depending on the considered molecule regions. Thus, two main region classes (or modules) can be delimited. The first type showed relatively low variation amongst sequence groups, showing mainly purifying selection and did not display a specific geographical clustering pattern (regions III, IV, V). The second type showed relatively high variation amongst sequence groups, diversifying selection when just variant positions were considered and purifying selection when all positions were analysed; it also displayed a geographical clustering pattern (regions II and VI belong to this second type).

Regions II and VI were associated with receptor-binding functions, a key role in the invasion process. This has been proved in region II and multiple evidence has implied a similar function in region VI. Simultaneously, exposure to the host's immune system causes selective pressure leading to an evasive response. The specific nature of this adaptation to the host drives these sectors' evolution towards geographical clustering patterns; these patterns are absent in other molecule sectors where low exposure to the host immune system and/or a strongest functional restriction (like structural constraints) produce high conservation.

Multiple and opposite selective forces acting in a relatively small region of a protein have been previously described in HLA (Suzuki & Gojobori, 1999), indicating that selective forces act on specific residues rather than whole regions; similar behaviour could be occurring within *dpb*. Although this might imply the existence of different functional subdomains within these sectors, more information is necessary to clarify this issue.

Recombination forces appear to be acting in all DBP and, therefore, their influence may be less important. Recognizing these different modules and understanding the factors that drive their evolution will allow the design of strategies for elucidating host–parasite interactions, in turn leading to new vaccine targets.

This work was supported by the President of Colombia's office and the Ministry of Public Health. We would like to thank Josefa Rodríguez and Andrómeda Gomez for their technical support, Stella Buitrago for supplying the *P. vivax*-infected blood samples and Jason Garry for patiently reviewing the manuscript. We would also like to thank Professor Manuel Elkin Patarroyo for his invaluable comments and suggestions. This article is in memory of Professor Fernando Chalem Benattar.

REFERENCES

ADAMS, J. H., BLAIR, P. L., KANEKO, O. & PETERSON, D. S. (2001). An expanding ebl family of *Plasmodium falciparum*. *Trends in Parasitology* **17**, 297–299.

ADAMS, J. H., SIM, B. K., DOLAN, S. A., FANG, X., KASLOW, D. C. & MILLER, L. H. (1992). A family of erythrocyte binding proteins of malaria parasites. *Proceedings of the National Academy of Sciences, USA* **89**, 7085–7089.

AMPUDIA, E., PATARROYO, M. A., PATARROYO, M. E. & MURILLO, L. A. (1996). Genetic polymorphism of the Duffy receptor binding domain of *Plasmodium vivax* in Colombian wild isolates. *Molecular and Biochemical Parasitology* **78**, 269–272.

BARNWELL, J. W. & GALINSKI, M. R. (1995). *Plasmodium vivax*: a glimpse into the unique and shared biology of the merozoite. *Annals of Tropical Medicine and Parasitology* **89**, 113–120.

BARNWELL, J. W., NICHOLS, M. E. & 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.

CHITNIS, C. E. (2001). Molecular insights into receptors used by malaria parasites for erythrocyte invasion. *Current Opinion in Hematology* **8**, 85–91.

CHITNIS, C. E. & MILLER, L. H. (1994). Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. *Journal of Experimental Medicine* **180**, 497–506.

COLE-TOBIAN, J. & KING, C. L. (2003). Diversity and natural selection in *Plasmodium vivax* Duffy binding protein gene. *Molecular and Biochemical Parasitology* **127**, 121–132.

COLE-TOBIAN, J. L., CORTES, A., BAISOR, M., KASTENS, W., XAINLI, J., BOCKARIE, M., ADAMS, J. H. & KING, C. L. (2002). Age-acquired immunity to a *Plasmodium vivax* invasion ligand, the duffy binding protein. *Journal of Infectious Diseases* **186**, 531–539.

COLLINS, D. W. & JUKES, T. H. (1994). Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics* **20**, 386–396.

CUI, L., ESCALANTE, A. A., IMWONG, M. & SNOUNOU, G. (2003). The genetic diversity of *Plasmodium vivax* populations. *Trends in Parasitology* **19**, 220–226.

DAYHOFF, M. O., SCHWARTZ, R. M. & ORCUTT, B. C. (1978). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (ed. Dayhoff, M. O.). National Biomedical Research Foundation. **5** (Suppl. 3e), 345–352.

ESCALANTE, A. A., GREBERT, H. M., ISEA, R., GOLDMAN, I. F., BASCO, L., MAGRIS, M., BISWAS, S., KARIUKI, S. & LAL, A. A. (2002). A study of genetic diversity in the gene encoding the circumsporozoite protein (CSP) of *Plasmodium falciparum* from different transmission areas-XVI. Asembo Bay Cohort Project. *Molecular and Biochemical Parasitology* **125**, 83–90.

ESCALANTE, A. A., LAL, A. A. & AYALA, F. J. (1998). Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* **149**, 189–202.

FANG, X. D., KASLOW, D. C., ADAMS, J. H. & MILLER, L. H. (1991). Cloning of the *Plasmodium vivax* Duffy receptor. *Molecular and Biochemical Parasitology* **44**, 125–132.

FIGTREE, M., PASAY, C. J., SLADE, R., CHENG, Q., CLOONAN, N., WALKER, J. & SAUL, A. (2000). *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. *Molecular and Biochemical Parasitology* **108**, 53–66.

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

FRASER, T., MICHON, P., BARNWELL, J. W., NOE, A. R., AL-YAMAN, F., KASLOW, D. C. & ADAMS, J. H. (1997). Expression and serologic activity of a soluble recombinant *Plasmodium vivax* Duffy binding protein. *Infection and Immunity* **65**, 2772–2777.

FU, Y. X. & LI, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

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

HUDSON, R. R. & 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.

KHO, W. G., CHUNG, J. Y., SIM, E. J., KIM, D. W. & CHUNG, W. C. (2001). Analysis of polymorphic regions of *Plasmodium vivax* Duffy binding protein of Korean isolates. *The Korean Journal of Parasitology* **39**, 143–150.

KIM, T., KIM, Y. J., SONG, K. J., SONG, J. W., CHA, S. H., KIM, Y. K., SHIN, Y. K., SUH, I. B. & LIM, C. S. (2002). The molecular characteristics of circumsporozoite protein gene subtypes from *Plasmodium vivax* isolates in Republic of Korea. *Parasitology Research* **88**, 1051–1054.

KUMAR, S., TAMURA, K., JAKOBSEN, I. B. & NEI, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.

KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.

LOMBARD, V., CAMON, E. B., PARKINSON, H. E., HINGAMP, P., STOESSER, G. & REDASCHI, N. (2002). EMBL-Align: a new public nucleotide and amino acid multiple sequence alignment database. *Bioinformatics* **18**, 763–764.

MANN, V. H., HUANG, T., CHENG, Q. & SAUL, A. (1994). Sequence variation in the circumsporozoite protein gene of *Plasmodium vivax* appears to be regionally biased. *Molecular and Biochemical Parasitology* **68**, 45–52.

MICHON, P., FRASER, T. & ADAMS, J. H. (2000). Naturally acquired and vaccine-elicited antibodies block erythrocyte cytoadherence of the *Plasmodium vivax* Duffy binding protein. *Infection and Immunity* **68**, 3164–3171.

MICHON, P., STEVENS, J. R., KANEKO, O. & ADAMS, J. H. (2002). Evolutionary relationships of conserved cysteine-rich motifs in adhesive molecules of malaria parasites. *Molecular Biology and Evolution* **19**, 1128–1142.

MILLER, L. H., MASON, S. J., CLYDE, D. F. & 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. & GOJOBORI, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

NICHOLAS, K., NICHOLAS, H. B. Jr. & DEERFIELD, D. W. II (1997). GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW NEWS* **4**, 14.

OCAMPO, M., VERA, R., EDUARDO RODRIGUEZ, L., CURTIDOR, H., URQUIZA, M., SUAREZ, J., GARCIA, J., PUENTES, A., LOPEZ, R., TRUJILLO, M., TORRES, E. & ELKIN PATARROYO, M. (2002). *Plasmodium vivax* Duffy binding protein peptides specifically bind to reticulocytes. *Peptides* **23**, 13–22.

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. *Proceedings of the National Academy of Sciences, USA* **99**, 16348–16353.

PUTAPORNTIP, C., JONGWUTIWES, S., TIA, T., FERREIRA, M. U., KANBARA, H. & TANABE, K. (2001). Diversity in the thrombospondin-related adhesive protein gene (TRAP) of *Plasmodium vivax*. *Gene* **268**, 97–104.

QARI, S. H., GOLDMAN, I. F., POVOA, M. M., DI SANTI, S., ALPERS, M. P. & LAL, A. A. (1992). Polymorphism in the circumsporozoite protein of the human malaria parasite *Plasmodium vivax*. *Molecular and Biochemical Parasitology* **55**, 105–113.

RANJAN, A. & CHITNIS, C. E. (1999). Mapping regions containing binding residues within functional domains of *Plasmodium vivax* and *Plasmodium knowlesi* erythrocyte-binding proteins. *Proceedings of the National Academy of Sciences, USA* **96**, 14067–14072.

RAYNER, J. C., CORREDOR, V., FELDMAN, D., INGRAVALLO, P., IDERABDULLAH, F., GALINSKI, M. R. & 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. & 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. & ROZAS, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

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

SINGH, S. K., SINGH, A. P., PANDEY, S., YAZDANI, S. S., CHITNIS, C. E. & SHARMA, A. (2003). Definition of structural elements in *Plasmodium vivax* and *Plasmodium knowlesi* Duffy binding domains necessary for erythrocyte invasion. *The Biochemical Journal* **374**, 193–198.

SUH, I. B., HOFFMAN, K. J., KIM, S. H., SONG, K. J., SONG, J. W., LEE, J. S. & LIM, C. S. (2001). The analysis of *Plasmodium vivax* Duffy receptor binding domain gene sequence from resurgent Korean isolates. *Parasitology Research* **87**, 1007–1010.

SUZUKI, Y. & GOJOBORI, T. (1999). A method for detecting positive selection at single amino acid sites. *Molecular Biology and Evolution* **16**, 1315–1328.

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

TAMURA, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution* **9**, 678–687.

THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. & 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.

TSUBOI, T., KAPPE, S. H., AL-YAMAN, F., PRICKETT, M. D., ALPERS, M. & ADAMS, J. H. (1994). Natural variation within the principal adhesion domain of the *Plasmodium vivax* duffy binding protein. *Infection and Immunity* **62**, 5581–5586.

VERRA, F. & HUGHES, A. L. (2000). Evidence for ancient balanced polymorphism at the Apical Membrane Antigen-1 (AMA-1) locus of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **105**, 149–153.

WERTHEIMER, S. P. & BARNWELL, J. W. (1989). *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Experimental Parasitology* **69**, 340–350.

XAINLI, J., ADAMS, J. H. & KING, C. L. (2000). The erythrocyte binding motif of *Plasmodium vivax* duffy binding protein is highly polymorphic and functionally conserved in isolates from Papua New Guinea. *Molecular and Biochemical Parasitology* **111**, 253–260.

https://doi.org/10.1017/S0031182003004773 Published online by Cambridge University Press