Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 304 (2003) 65–75
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*☆

Marcelo U. Ferreira[a,*], Weber L. Ribeiro[b], Angela P. Tonon[a,b], Fumihiko Kawamoto[c], Stephen M. Rich[d]

[a]*Department of Parasitology, Institute for Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes 1374, 05508-900 São Paulo (SP), Brazil*
[b]*Laboratory of Molecular Parasitology, Faculty of Medicine of São José do Rio Preto, 15090-000 São José do Rio Preto (SP), Brazil*
[c]*Department of International Health, Nagoya University School of Medicine, Nagoya 466-8550, Japan*
[d]*Division of Infectious Diseases, Tufts University School of Veterinary Medicine, North Grafton, MA 01536, USA*

Received 15 May 2002; received in revised form 29 September 2002; accepted 29 November 2002
Received by F.G. Alvarez-Valin

Abstract

The merozoite surface protein-1 (MSP-1) of the malaria parasite *Plasmodium falciparum* is a major blood-stage antigen containing highly polymorphic tripeptide repeats in the domain known as block 2 and several non-repetitive domains that are essentially dimorphic. We have analyzed sequence variation in block 2 repeats and in non-repetitive block 17, as well as other polymorphisms within the *MSP-1* gene, in clinical isolates of *P. falciparum*. Repeat haplotypes were defined as unique combinations of repeat motifs within block 2, whereas block 17 haplotypes were defined as unique combinations of single nucleotide replacements in this domain. A new block 17 haplotype, E-TNG-L, was found in one isolate from Vietnam. *MSP-1* alleles, defined as unique combinations of haplotypes in blocks 2 and 17 and other polymorphisms within the molecule, were characterized in 60 isolates from hypoendemic Brazil and 37 isolates from mesoendemic Vietnam. Extensive diversity has been created in block 2 and elsewhere in the molecule, while maintaining significant linkage disequilibrium between polymorphisms across the non-telomeric *MSP-1* locus separated by a map distance of more than 4 kb, suggesting that low meiotic recombination rates occur in both parasite populations. These results indicate a role for non-homologous recombination, such as strand-slippage mispairing during mitosis and gene conversion, in creating variation in a malarial antigen under strong diversifying selection.

Keywords: Merozoite surface protein-1; *Plasmodium falciparum*; Malaria vaccine; Evolution; Population genetics

1. Introduction

Plasmodium falciparum is a major human pathogen associated with 200–300 million clinical cases of malaria and 1–3 million deaths each year worldwide. Several lines of evidence suggest that all living members of this species originated from a single progenitor, between 9500 and 57,500 years ago. Silent nucleotide replacements, thought to be selectively neutral, are very rare or absent in nuclear genes (Rich et al., 2000), as well as in the 6-kb mitochondrial genome of *P. falciparum* (Conway et al., 2000b). Moreover, single nucleotide polymorphisms (SNPs) are rarely found in *P. falciparum* introns (Volkman et al., 2001). This contrasts with an extensive polymorphism in most genes coding for surface antigens of *P. falciparum*, that are under selection to escape memory immune responses elicited by previous exposure to variant forms of the same antigen. B- and T-cell epitopes are often located on variable domains on surface antigens of malaria parasites, several of them comprising tandem repeats (Rich et al., 2000).

The merozoite surface protein-1 (MSP-1) is a major malarial blood-stage antigen containing highly variable

Abbreviations: MSP-1, merozoite surface protein-1; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; PCR, polymerase chain reaction; d.f., degrees of freedom; CSP, circumsporozoite protein; MSP-2, merozoite surface protein-2.

☆ Nucleotide sequence data reported in this paper are available in the GenBank™ database under the accession numbers AF509630–AF509719.

* Corresponding author. Tel.: +55-11-3091-7273; fax: +55-11-3091-7417.
E-mail address: muferrei@usp.br (M.U. Ferreira).

0378-1119/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0378-1119(02)01180-0

repeat arrays. Genetic variation in the *P. falciparum MSP-1* gene has been extensively investigated, since it may represent an obstacle for the development of vaccines based on this molecule. The gene may be divided into 17 blocks according to levels of inter-allele sequence divergence (Fig. 1), and most variation is dimorphic: sequences may be grouped into one of two allelic groups, known as K1 and MAD20. Block 2 represents an exception to dimorphism, since in addition to K1-type and MAD20-type alleles, that contain degenerate tripeptide repeats, an apparently non-repetitive allele known as RO33 is frequently found. RO33-type alleles have been suggested to display ancestral repeats that have been blurred by several random mutations (Jongwutiwes et al., 1992). Block 2 of MSP-1 has recently been defined as a principal target of antibodies associated with clinical immunity in African children (Conway et al., 2000a). Block 17 codes for MSP-1_{19}, the 19-kDa C-terminal product of enzymatic processing of MSP-1 that remains anchored to the merozoite surface at the time of erythrocyte invasion. MSP-1_{19} is a major target of naturally acquired antimalarial immunity (O'Donnell et al., 2001), but the impact on antibody recognition of the limited variation (five SNPs) found in this molecule remains unknown.

The *MSP-1* gene provides an example of how interhelical exchanges of blocks of dimorphic sequences may generate new alleles during sexual (meiotic) recombination (Kerr et al., 1994). However, rates of effective meiotic recombination differ across natural *P. falciparum* populations; a wide variation in inbreeding coefficients, for example, have been estimated for different endemic areas (Babiker and Walliker, 1997). In African isolates, frequent recombination seems to have allowed little two-site linkage disequilibrium (LD) to occur over distances of more than one kb at the non-telomeric, 5 kb-long *MSP-1* locus (Conway et al., 1999). In contrast, strong LD may be found between *MSP-1* polymorphisms situated more than 4 kb apart in isolates from areas of less intense malaria transmission in Brazil, South-East Asia and Southwestern Pacific (Sakihama et al., 1999, 2001; Tanabe et al., 2000; Silveira et al., 2001). Here we sought to determine the relative importance of sexual (meiotic) and mitotic mechanisms in generating diversity at this locus in a hypoendemic region of Brazil and a mesoendemic region of Vietnam.

2. Materials and methods

2.1. Plasmodium falciparum *isolates and DNA isolation*

Field isolates (samples of parasites derived from a single patient on a single occasion) may contain either genetically homogeneous or heterogeneous parasite populations. To analyze associations between sites on polymerase chain reaction (PCR) products obtained in separate experiments, we selected isolates with apparently single-clone infections, as judged by typing several polymorphisms within the single-copy *MSP-1* gene (see below).

The 60 Brazilian isolates available for this study were obtained from febrile patients in Rondônia between 1985 and 1998. The sites and dates of collection were as follows: Porto Velho, Ariquemes, and Guajará-Mirim, 1985 ($n = 24$), Porto Velho, 1995 ($n = 15$), Porto Velho and Ariquemes, 1997 ($n = 12$) and Porto Velho, 1998 ($n = 9$). The state of Rondônia, situated in the southwestern part of the Amazon Basin of Brazil, is characterized by year-round hypoendemic transmission of both *P. falciparum* and *Plasmodium vivax*. Heavily exposed individuals are usually non-immune migrants from malaria-free areas working in close contact with the rain forest. Parasite rates typically range between 0.5 and 2%, and most malaria infections are symptomatic (Ferreira et al., 1998a). Rondônia accounted for 30–50% of all malaria infections recorded in Brazil in the 1980s and 1990s.

Isolates from Vietnam were collected between January and December 1996 from 55 patients attending the Lam Dong Provincial Hospital no. 2, in the town of Bao Loc. Nearly all of them were outpatients with symptomatic but uncomplicated *P. falciparum* malaria. Bao Loc is situated 150 km northeast of Ho Chi Minh City, in a mesoendemic area on the southern highlands of Vietnam. Typical parasite rates are about 30%, and *P. falciparum* accounts for three fourths of all symptomatic malaria infections (Ferreira et al.,

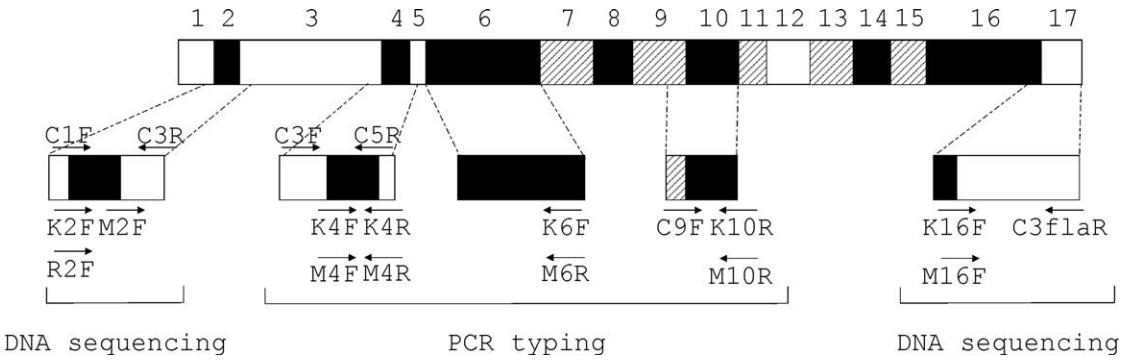


Fig. 1. Schematic representation of the *MSP-1* gene of *P. falciparum*, showing the 17 blocks defined by Tanabe and colleagues (apud Miller et al., 1993). Conserved, semiconserved, and variable blocks are shown as open, hatched and filled boxes, respectively. The location and orientation of primers used for typing and sequencing different fragments of *MSP-1* are also shown; see text for details.

1998c). The population served by the hospital (140,000 people) includes Ethnic-Vietnamese and ethnic minorities who inhabit remote hill areas.

At both study sites, venous blood samples were collected, after informed consent, and parasite DNA was isolated from 100–200-μl aliquots of whole blood using either proteinase K digestion followed by phenol-chloroform extraction and ethanol precipitation (Ferreira et al., 1998a) or digestion with 8 M guanidine isothiocyanate followed by isopropanol precipitation (Kaneko et al., 1997; Ferreira et al., 1998c). After isolation, DNA was resuspended in 100 μl of sterile TE buffer (10 mM Tris base and 1 mM EDTA) or distilled water and stored at −20°C.

2.2. Typing blocks 2, 4, 6, 10, and 16 of MSP-1

A set of oligonucleotide primers pairs corresponding to variable and conserved sequences was applied to PCR to type the variable blocks 2, 4, 6, and 10 of *MSP-1* (Fig. 1) (Kaneko et al., 1997). The typing process consisted of three steps. In the first, gene fragments between blocks 9 and 10 were amplified in two separate reactions containing the forward primer C9F and either K10R or M10R as the reverse primer. In the second step, segments between blocks 2 and 6 were amplified in three separate reactions, with different combinations of the forward primers K2F, M2F or R2F and the forward primers K6R or M6R. Given the reported absence of recombination between blocks 6 and 10 (Kaneko et al., 1997), the allelic type determined for block 10 (either K1 or MAD20) was assumed to be the same for block 6, and this information was used to select the reverse primer (either K6R or M6R) used in the second amplification step. In the third step, block 4 was typed by nested PCR using as template the product amplified in the second step. Alternatively, we used as template the product amplified with primers C3F and C5R. Block 4 was typed in separate reactions, each containing one of four different combinations of forward (K4F or M4F) and reverse primers (K4R or M6R), to detect hybrid sequences (K1:MAD20 or MAD20:K1) generated by recombination within this domain (Kaneko et al., 1997; Ferreira et al., 1998a). Primer sequences and amplification conditions were described elsewhere (Kaneko et al., 1997; Ferreira et al. 1998a,b). Block 16 allelic type was determined by sequencing its 3′ end, as described below. Here we have further confirmed the absence of recombination between blocks 10 and 16 (Kaneko et al., 1997), as judged by the complete LD observed in this segment of *MSP-1*. Full typing results for blocks 2, 4, 6, 10, and 16 of all Brazilian isolates analyzed here have been previously published (Silveira et al., 2001), as did partial typing results of 38 Vietnamese isolates (Ferreira et al., 1998c). Here we determined block 16 allelic types of these previously analyzed parasites and fully typed 17 additional isolates from Vietnam.

2.3. DNA sequencing of block 2

An *MSP-1* gene fragment of about 300 bp, comprising the 3′ end of block 1, full-length block 2 and the 5′ end of block 3, was PCR-amplified using the oligonucleotide primers C1F and C3R (Fig. 1) as described (Ferreira et al., 1998a). PCR products were directly sequenced with the primer C1F using the ThermoSequenase Cy5.5 Dye Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech, Piscataway, NJ). Sequencing reaction products were analyzed in a 4 × 4 Personal Sequencing System (Amersham Pharmacia Biotech). An additional PCR product, obtained independently from a new DNA template, was sequenced whenever new nucleotide substitutions were found. A total of 60 block 2 sequences from Brazilian isolates and 52 sequences from Vietnamese isolates were obtained and analyzed.

2.4. DNA sequencing of block 17

A 390-bp fragment comprising the 3′ end of block 16 and full-length block 17 was amplified, using either K1F or M6F as a forward primer and C3flaR as a reverse primer (Fig. 1), as described elsewhere (Kaneko et al., 1997; Silveira et al., 2001). PCR products were purified and directly sequenced, on a single strand, using the primer K1F or M16F. Sequence analysis was done as above. Block 17 sequences for the 60 Brazilian isolates analyzed here had been previously published (Silveira et al., 2001); here we have obtained and analyzed block 17 sequences from 46 Vietnamese isolates.

2.5. Source of other partial MSP-1 sequences analyzed

Partially sequenced *MSP-1* alleles from mesoendemic Thailand (Jongwutiwes et al., 1992) and holoendemic Tanzania (Jiang et al., 2000) were also analyzed. The first set of DNA sequences was derived from 19 *P. falciparum* isolates collected in Mae Sod, northern Thailand, between November 1988 and January 1989; they comprised blocks 1, 2, and 3 of *MSP-1*. The second set of *MSP-1* alleles was derived from 25 isolates collected in Ifakara, southern Tanzania, in 1989; sequences comprised blocks 1–4. These sequence data were used in the analysis of geographical distribution of block 2 repeat haplotypes and to test for LD between SNPs flanking block 2 repeats in areas with different levels of malaria transmission. Block 17 sequences from 31 isolates collected between 1994 and 1995 in the towns of Bao Loc and Phu Rieng, about 100 km apart (Kaneko et al., 1997), were combined with those from isolates collected in Bao Loc 12–18 months later in the analysis of the temporal distribution of block 17 haplotypes in southern Vietnam. Additional published sequences of parasites from Thailand (72 alleles, Sakihama et al., 1999) and Kenya, East Africa (46 alleles, Qari et al., 1998) were

used to compare the geographical distribution of block 17 haplotypes.

2.6. Data analysis

DNA sequences were aligned using ClustalX 1.8 software, with manual editing. To align repeat arrays of block 2, five K1-type and eight MAD20-type repeat motifs or allotypes were characterized (Fig. 2); repeat motifs, rather than nucleotides, were aligned. Block 2 repeat haplotypes were defined as unique combinations of repeat motifs, whereas block 17 haplotypes were defined as unique combinations of SNPs in positions 1643, 1691, 1700, 1701, and 1716, according to the sequence alignment by Miller et al. (1993). For analytical purposes, *MSP-1* alleles from isolates from Brazil and Vietnam were defined as unique combinations of: (a) allelic types (either K1, MAD20 or RO33) in the non-repetitive 5′ and 3′ parts of blocks 2, 4 and 6–16, as determined by PCR typing; (b) repeat haplotypes in K1-type or MAD20-type block 2 (no variation was found in RO33-type block 2 sequences); and (c) SNPs in the 5′ part of block 1, in the 3′ part of block 3, and within block 17 (Fig. 3). *MSP-1* gene diversity (H) in different populations was calculated as $H = [n/(n-1)]\ [1 - \sum p_i^2]$, where n is the number of isolates sampled and p_i is the frequency of each *MSP-1* allele. Significance of LD was assessed: (a) by applying standard χ^2 tests to 2×2 or 2×3 contingency tables; and (b) by using a Markov chain to explore all theoretically possible contingency tables. P values obtained by this last procedure correspond to the proportion of analyzed contingency tables associated with probabilities smaller or equal to the observed table. LD analyses were performed using Arlequin 2.000 software. Comparisons between nucleotide distances and proportions of sites showing significant LD were done by using χ^2 statistics with Yates correction in 2×2 tables. Wright's F_{ST} statistics was used to analyze temporal variation in block 17 haplotype frequencies in Vietnam, with P values obtained by permutation of haplotypes between populations using Arlequin 2.000. Significance was set at the 5% level.

3. Results

3.1. Sequence variation in block 2 of MSP-1 in Brazil, Vietnam, Thailand, and Tanzania

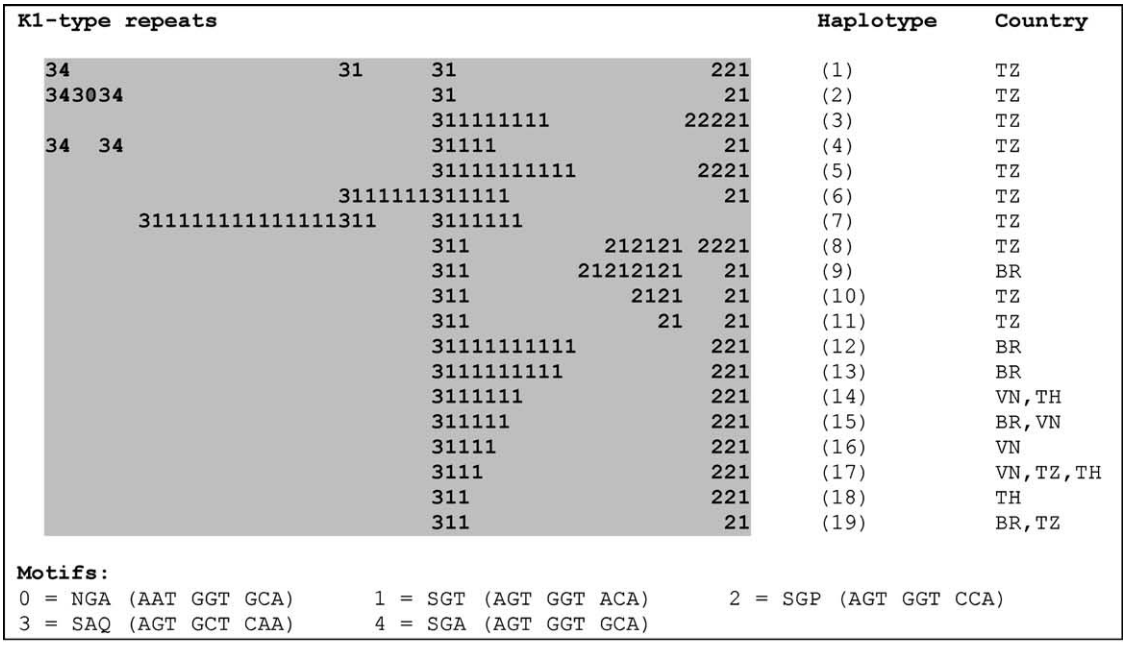
We characterized 19 different K1-type repeat haplotypes by analyzing 52 K1-type block 2 sequences from Brazil ($n = 22$), Vietnam ($n = 7$), Thailand ($n = 6$), and Tanzania ($n = 17$). The highest levels of diversity in K1-type repeats were found in Tanzania, where 12 different haplotypes were found. The overall structure of repeat arrays was remarkably well conserved. Most haplotypes were combinations of the repeat motifs SGT, SGP, and SAQ (Fig. 2). Nearly all K1-type repeats started with the hexapeptide SAQSGT, with three exceptions found in Tanzania, and ended with the hexapeptide SGPSGT, with a single exception in Tanzania. Most diversity was due to duplications and deletions of the repeat motifs SGT and SGP; no synonymous nucleotide replacements were found in repeat motifs. Further variation was created, in a single allele from Tanzania (IFA17.25), by the duplication of the sequence SPPADA in the non-repetitive 3′ part of block 2 (Jiang et al., 2000). The geographical distribution of K1-type repeat haplotypes was highly biased; only four of them (#14, #15, #17, and #19) were shared by parasites from different countries (Fig. 2).

Among 78 MAD20-type block 2 sequences from Brazil ($n = 17$), Vietnam ($n = 39$), Thailand ($n = 16$), and Tanzania ($n = 6$), 30 different repeat haplotypes were found; 15 of them were present in Vietnam (Fig. 2). MAD20-type repeat haplotypes represented essentially different combinations of the motifs SGG, SVA, SVT, and SKG, starting alternatively with SKG or SGGSVT (with a single exception in Thailand) and usually ending with SVASGG (with four exceptions in Thailand and Vietnam). Most diversity can be explained by duplications and deletions of the motifs SGG, SVA and SVT, frequently combined in hexapepides. Synonymous nucleotide replacements were found in the repeat motifs SGG and SVA. However, as previously described for repeats of the circumsporozoite protein (*CSP*) gene of *P. falciparum* (Rich et al., 1997), the distribution of silent substitutions across MAD20-type repeat arrays was far from random, being restricted to the central region black-shaded in Fig. 2. TCAGGTGG*C* (silent substitution italicized), an alternative repeat allotype coding for SGG, was found in the central region of 19 MAD20-type haplotypes, as a component of a repetitive unit TCAGTTGCT[TCAGGTGGT]TCAGGTGGC (sequence in square brackets may be either present or absent). Only three MAD20-type repeat haplotypes (#22, #37, and #41) were shared by parasites from different countries (Fig. 2).

The 32 RO33-type block 2 sequences analyzed (21 from Brazil, four from Vietnam, three from Thailand, and four from Tanzania) were nearly identical. A single non-synonymous nucleotide replacement was frequently found in codon 64 (position according to the alignment by Miller et al., 1993), resulting in D (as in Ghanian isolate RO33) being replaced by G (as in Thai isolate CSL2) (Miller et al., 1993). The geographical distribution of this dimorphic variation was biased: all Brazilian and Vietnamese isolates had CSL2-type sequences, as did one out of three Thai isolates, but all isolates in Tanzania and two in Thailand were RO33-type. Another G → D substitution, in residue 104, was found in the RO33-type sequence of isolate IFA7, from Tanzania (Jiang et al., 2000).

3.2. Distribution of block 17 haplotypes

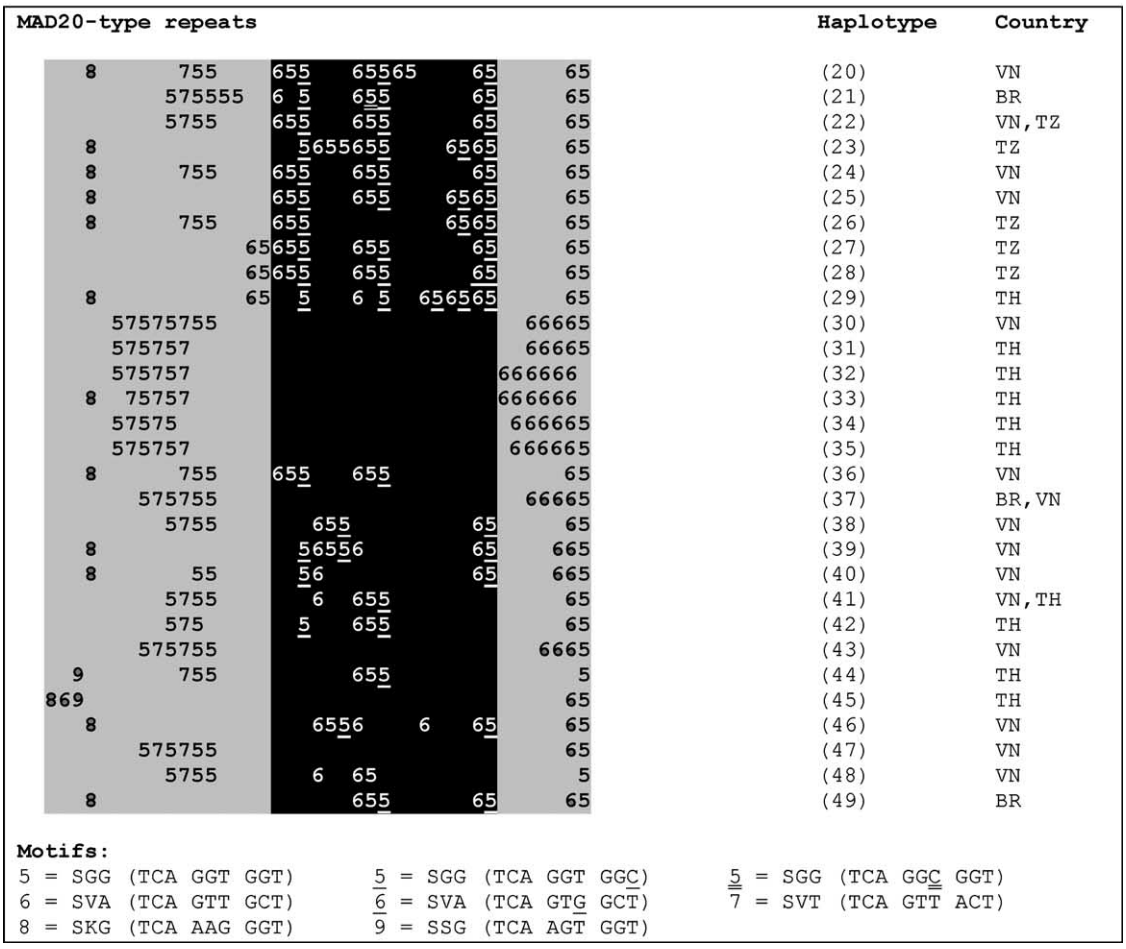
We have compared the distribution of block 17 haplotypes in two samples of parasites collected in southern



K1-type repeats	Haplotype	Country
34 31 31 221	(1)	TZ
343034 31 21	(2)	TZ
311111111 22221	(3)	TZ
34 34 31111 21	(4)	TZ
31111111111 2221	(5)	TZ
3111111311111 21	(6)	TZ
311111111111111311 3111111	(7)	TZ
311 212121 2221	(8)	TZ
311 21212121 21	(9)	BR
311 2121 21	(10)	TZ
311 21 21	(11)	TZ
31111111111 221	(12)	BR
3111111111 221	(13)	BR
3111111 221	(14)	VN,TH
311111 221	(15)	BR,VN
31111 221	(16)	VN
3111 221	(17)	VN,TZ,TH
311 221	(18)	TH
311 21	(19)	BR,TZ

Motifs:

0 = NGA (AAT GGT GCA) 1 = SGT (AGT GGT ACA) 2 = SGP (AGT GGT CCA)
3 = SAQ (AGT GCT CAA) 4 = SGA (AGT GGT GCA)



MAD20-type repeats	Haplotype	Country
8 755 655 65565 65 65	(20)	VN
575555 6 5 655 65 65	(21)	BR
5755 655 655 65 65	(22)	VN,TZ
8 5655655 6565 65	(23)	TZ
8 755 655 655 65 65	(24)	VN
8 655 655 6565 65	(25)	VN
8 755 655 6565 65	(26)	TZ
65655 655 65 65	(27)	TZ
65655 655 65 65	(28)	TZ
8 65 5 6 5 656565 65	(29)	TH
57575755 66665	(30)	VN
575757 66665	(31)	TH
575757 666666	(32)	TH
8 75757 666666	(33)	TH
57575 666665	(34)	TH
575757 666665	(35)	TH
8 755 655 655 65	(36)	VN
575755 66665	(37)	BR,VN
5755 655 65 65	(38)	VN
8 56556 65 665	(39)	VN
8 55 56 65 665	(40)	VN
5755 6 655 65	(41)	VN,TH
575 5 655 65	(42)	TH
575755 6665	(43)	VN
9 755 655 5	(44)	TH
869 65	(45)	TH
8 6556 6 65 65	(46)	VN
575755 65	(47)	VN
5755 6 65 5	(48)	VN
8 655 65 65	(49)	BR

Motifs:

5 = SGG (TCA GGT GGT) 5 = SGG (TCA GGT GGC) 5 = SGG (TCA GGC GGT)
6 = SVA (TCA GTT GCT) 6 = SVA (TCA GTG GCT) 7 = SVT (TCA GTT ACT)
8 = SKG (TCA AAG GGT) 9 = SSG (TCA AGT GGT)

Fig. 2. Alignment of the K1-type and MAD20-type tripeptide repeats in block 2 of the *MSP-2* gene and definition of repeat haplotypes. The black shading in MAD20-type repeats indicates the region where synonymous nucleotide replacements are found in repeat motifs. The origin of isolates displaying each repeat haplotype is indicated as follows: BR, Brazil, VN, Vietnam, TH, Thailand, and TZ, Tanzania. Sequence data for isolates from Thailand and Tanzania are from Jongwutiwes et al. (1992) and Jiang et al. (2000), respectively.

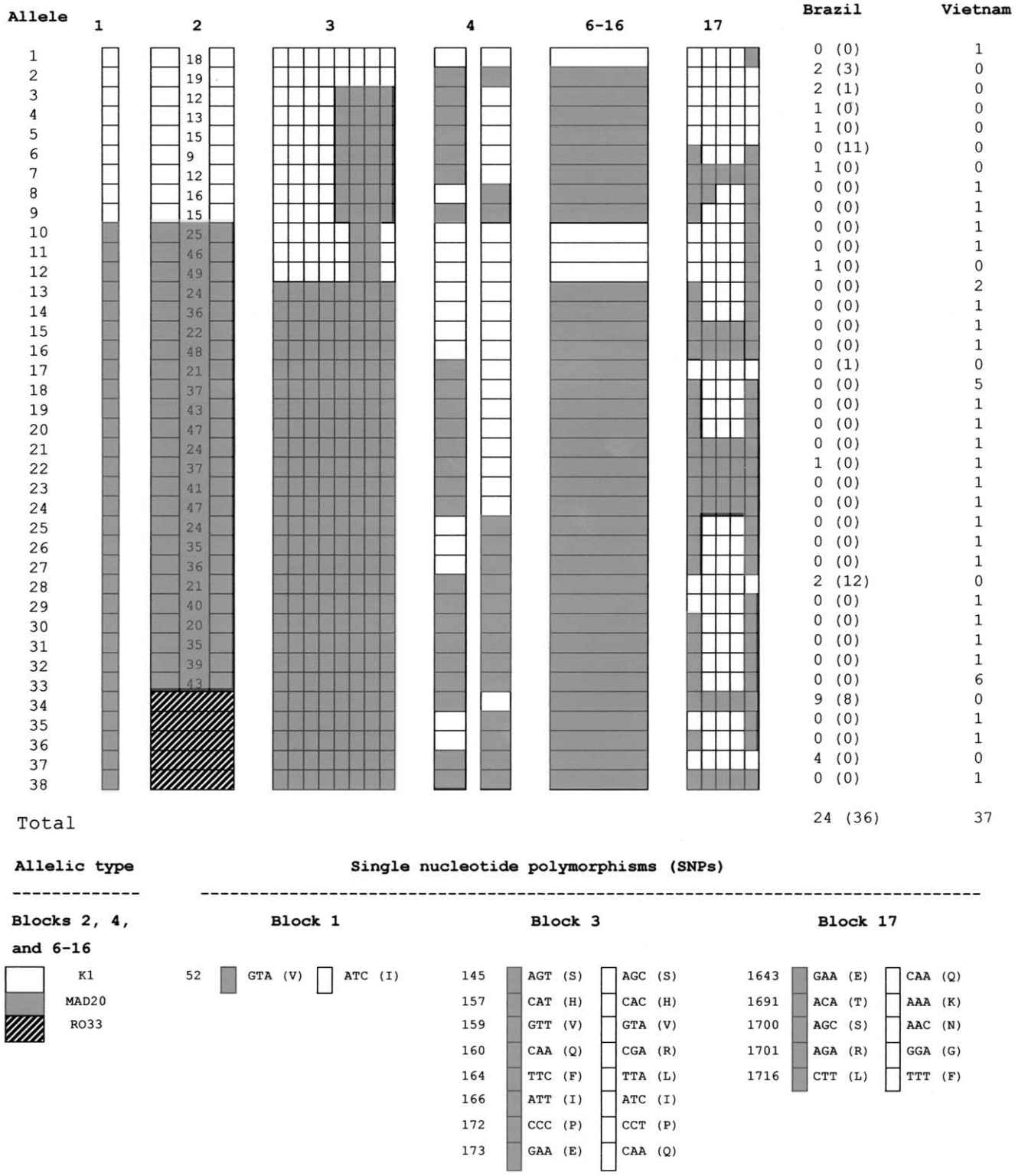


Fig. 3. The 38 *MSP-1* alleles found in parasite populations from Brazil and Vietnam. Each allele was defined as a unique combination of: (a) allelic types in non-repetitive parts of blocks 2, 4 and 6–16, as determined by PCR typing; (b) repeat haplotypes in K1-type and MAD20-type block 2 (according to numbering shown in Fig. 2), as determined by DNA sequencing; and (c) single nucleotide polymorphisms (SNPs) in blocks 1, 3, and 17, as determined by DNA sequencing. Codon numbers are given according to Miller et al. (1993). Note that, strictly speaking, variation found in codon 52 does not represent a SNP, since two nucleotides are dimorphic. Allele frequencies found in Brazil are shown separately for 24 isolates collected in 1985 (numbers outside parentheses) and 36 isolates collected between 1995 and 1998 (numbers in parentheses).

Vietnam: 31 isolates collected in 1994 and 1995 (Kaneko et al., 1997) and 46 isolates collected in 1996. The haplotype E-KNG-L (also called Uganda [Qari et al., 1998]) predominated on both occasions (58% in 1994–1995 and 65% in 1996); no temporal variation was found in haplotype frequencies in parasite populations sampled 12–18 months apart ($\chi^2 = 3.42$, 4 degrees of freedom [d.f.], $P = 0.508$). Other block 17 haplotypes found in 1996 were E-TSR-L (20% of the samples), Q-KNG-L (13%) and E-TNG-L (2%). The haplotype E-TNG-L, found in a single isolate collected in 1996 (BL154), represents a new unique combination of SNPs in block 17 that has been confirmed on three independent sequencing experiments performed with different PCR products. This new haplotype, hereafter called 'Vietnam', may be interpreted as a result of either a point mutation or a single crossover event between haplotypes E-KNG-L (Uganda) and E-TSR-L (Indo) (Qari et al., 1998), both of them commonly found in Vietnam.

The haplotype E-KNG-L, that predominated in Vietnam, is widespread in three continents: in recent studies, it was the most frequent haplotype found in Thailand (Sakihama et al., 1999) and Kenya (Qari et al., 1998), and was the second most frequent haplotype found in Brazil during the 1990s (Silveira et al., 2001). However, significant between-population variation in block 17 haplotype frequencies, as estimated by Wright's F_{ST} statistics, was found in five out of six pair-wise comparisons: Vietnam and Brazil, $F_{ST} = 0.145$ ($P < 0.0001$); Vietnam and Thailand, $F_{ST} = 0.024$ ($P = 0.045$); Vietnam and Kenya, $F_{ST} = 0.039$ ($P = 0.027$); Brazil and Thailand, $F_{ST} = 0.056$ ($P = 0.009$); Brazil and Kenya, $F_{ST} = 0.070$ ($P < 0.0001$); and Thailand and Kenya, $F_{ST} = 0.011$ ($P = 0.72$). Since temporal variation in block 17 haplotype frequencies occurred in Brazil (Silveira et al., 2001), this comparison was restricted to isolates collected in the 1990s.

3.3. Generation of new MSP-1 alleles in Brazil and Vietnam

The 60 fully typed isolates from Brazil were grouped into 12 different alleles, giving an estimate of gene diversity (H) of 0.83 for isolates collected in the 1980s and of 0.76 for isolates collected in the 1990s. The 37 fully typed isolates from Vietnam belonged to 27 different alleles, giving a substantially higher H estimate (0.97). Only one allele (#22) was shared by both parasite populations (Fig. 3). Instances of possible epidemic propagation of parasites carrying particular *MSP-1* alleles were found in Brazil: 13 out of 24 isolates collected in the 1980s belonged to alleles #34 or #37, whereas two thirds of the isolates collected in the 1990s belonged to alleles #6 or #28.

Data shown in Fig. 3 suggest that extensive genetic variation in *MSP-1* may have been created by exchanging blocks of alternate dimorphic sequences during meiosis. However, this data set also comprises examples of new *MSP-1* alleles generated by rearrangements in block 2 repeats, that can result from strand-slippage events during asexual (mitotic) propagation of parasites. K1-type alleles #3, #4, and #5, for instance, differed only in the number of SGT repeats but were identical in all dimorphic sites analyzed across the molecule. There were seven groups of MAD20-type alleles that differed in block 2 repeats but were identical elsewhere: (i) #10, #11, and #12; (ii) #13 and #14; (iii) #15 and #16; (iv) #18, #19, and #20; (v) #21, #22, #23, and #24; (vi) ##25, #26, and #27; and (vii) #30, #31, #32, and #33. Note that alleles from groups (i) and (ii) displayed different numbers and arranges of SGG and SVA repeats in the central part of MAD20-type repeats (black-shaded in Fig. 2), whereas alleles from group (iv) differed only in the number of SVA copies in the 3′ part of the repeat array. Most of these clusters of nearly identical alleles comprised sympatric isolates.

Five additional pairs of alleles (#19 and #33; #26 and #31; #17 and #28, #14 and #27, and #13 and #25) were also nearly identical: they shared the same block 2 repeat haplotype and dimorphic substitutions in all other sites analyzed except for the 3′ part of block 4 (also called block 4b [Ferreira et al., 1998a]). Four of them were pairs of alleles found in Vietnam and one pair was found in Brazil (Fig. 3). Alleles within each group may have been derived from each other by introducing an alternate block 4b sequence without changing the dimorphic sequence types on either side of this segment of *MSP-1*.

3.4. Linkage disequilibrium at the MSP-1 locus

The analysis of LD between pairs of dimorphic sites (SNPs and variable blocks shown in Fig. 3) within *MSP-1* alleles of Brazil and Vietnam is summarized in Fig. 4. Note that this pair-wise analysis does not include sites within the 3.7-kb-long stretch between blocks 6 and 16, where full LD has been observed in all alleles so far typed or sequenced. In both endemic areas, significant LD was observed between neighboring SNPs within blocks 3 and 17. Moreover, we have confirmed that block 2 repeats are a cold spot for meiotic recombination, since significant LD is maintained between sites flanking the repeats (blocks 1, 2, and 3). Strong LD between SNPs within blocks 1–3 of *MSP-1* was also detected in partial *MSP-1* sequences from Thailand and Tanzania (data not shown), as well as in the data set analyzed by Conway et al. (1999). Significant LD was more frequently found between relatively close (3–978 bp) than distant (3.75–4.99 kb) sites in isolates from both Brazil (70 versus 26% of pairwise comparisons being significant, $\chi^2 = 27.76$, 1 d.f., $P < 0.0001$) and Vietnam (64 versus 33%, $\chi^2 = 11.20$, 1 d.f., $P = 0.0012$). The LD patterns shown in Fig. 4, however, contrast with those reported for African isolates by Conway et al. (1999) in that we have found several instances of significant LD between 5′ sites (blocks 2 and 3) and 3′ sites (blocks 6–16 and 17) (Fig. 4), which are separated by a map distance between 3.75 and 4.99 kb, suggesting that meiotic recombination in local parasites was unable to disrupt linkage between distant

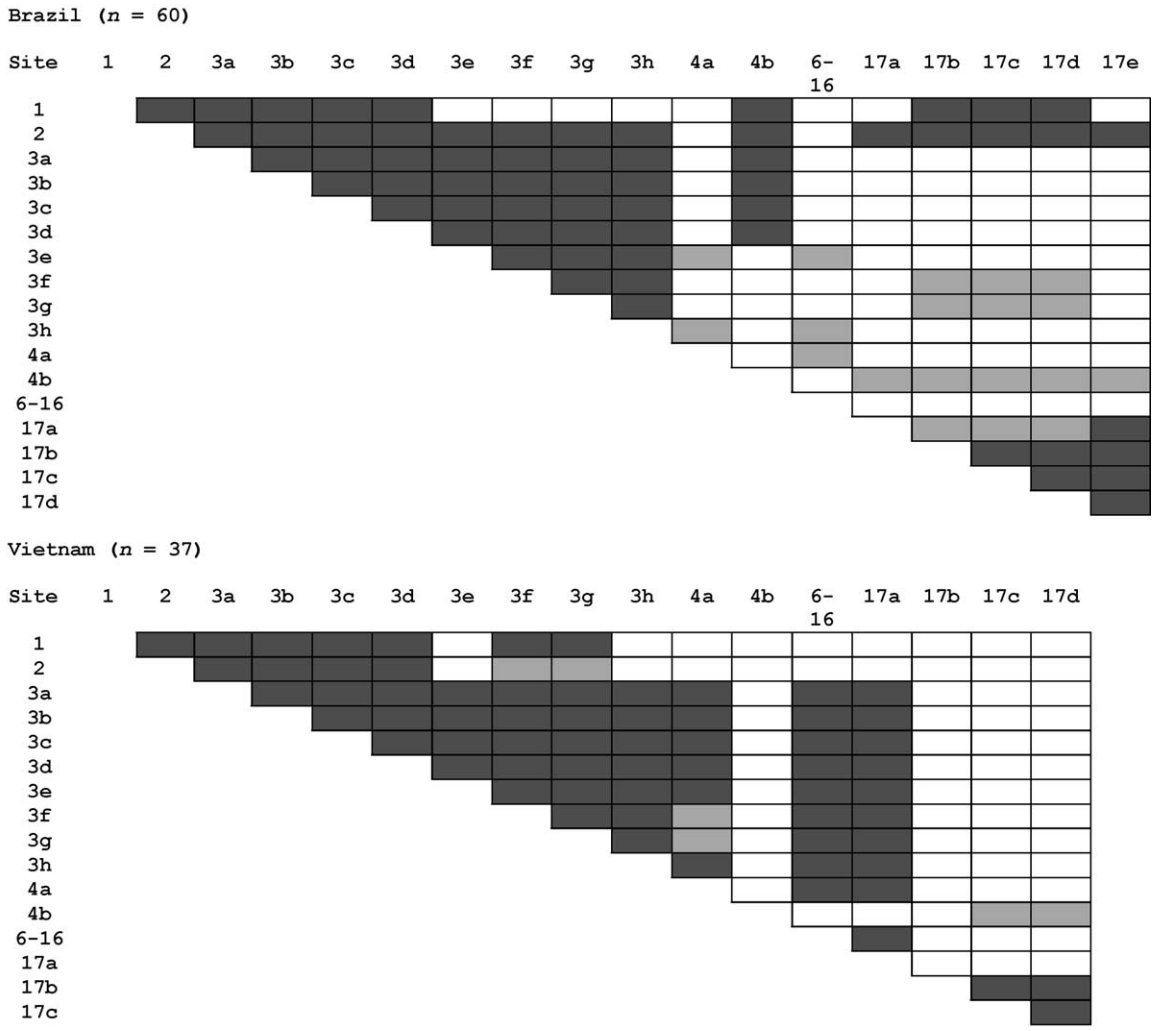


Fig. 4. Linkage disequilibrium (LD) between pairs of polymorphic sites (SNPs and variable blocks shown in Fig. 3) within *MSP-1* alleles from Brazil and Vietnam. The eight SNPs in block 3 and the five SNPs in block 17 are represented as 3a–h and 17a–e, respectively. Blocks 4a and 4b are respectively the 5′ and 3′ part of block 4, that are separated by a recombination site (Kaneko et al., 1997; Ferreira et al., 1998a). Two tests were used to test the significance of LD, standard χ^2 statistics and Markov chains (see text for details). Dark-shaded and light-shaded boxes indicate statistical significance ($P < 0.05$) detected by both methods and only one of these methods, respectively; the absence of significant LD is represented with open boxes. The site 17e was monomorphic in isolates from Vietnam, and thus excluded from pair-wise analyses.

polymorphisms. Moreover, no negative correlation between levels of LD (as measured by the *D* statistic [Lewontin and Kojima, 1960]) and genetic map distance between polymorphic sites was apparent in our data sets for Brazil and Vietnam (data not shown).

4. Discussion

Here we have shown that variation in the number, sequence, and arrangement of repeat units within block 2 defines several new *MSP-1* alleles in two parasite populations with putatively low effective meiotic recombination rates, as judged by the high levels of LD maintained across this locus. Most sequence diversity in other major surface malaria antigens, such as CSP and merozoite surface protein-2 (MSP-2), is associated with variation in immunodominant repeat arrays, that may be generated by non-meiotic recombination, such as strand-slippage events and gene conversion, in addition to classical homologous recombination during meiosis (Rich et al., 1997, 2000). However, the putative impact on *MSP-1* evolution of intrahelical exchanges of block 2 repeats followed by immune-mediated selection of new variants has been overlooked, with emphasis being placed on interhelical exchanges of blocks of dimorphic non-repetitive sequences during meiosis. For comparison, both recombination mechanisms are shown in Fig. 5. Direct repeats are restricted to a short segment (block 2) of the gene, but other cryptic repetitive sequences that have been detected elsewhere at the *MSP-1* locus putatively undergo mitotic recombination as well (Rich et al. 2000). Accordingly, several repetitive elements are scattered across the variable regions in the *MSP-1* genes from the rodent parasites

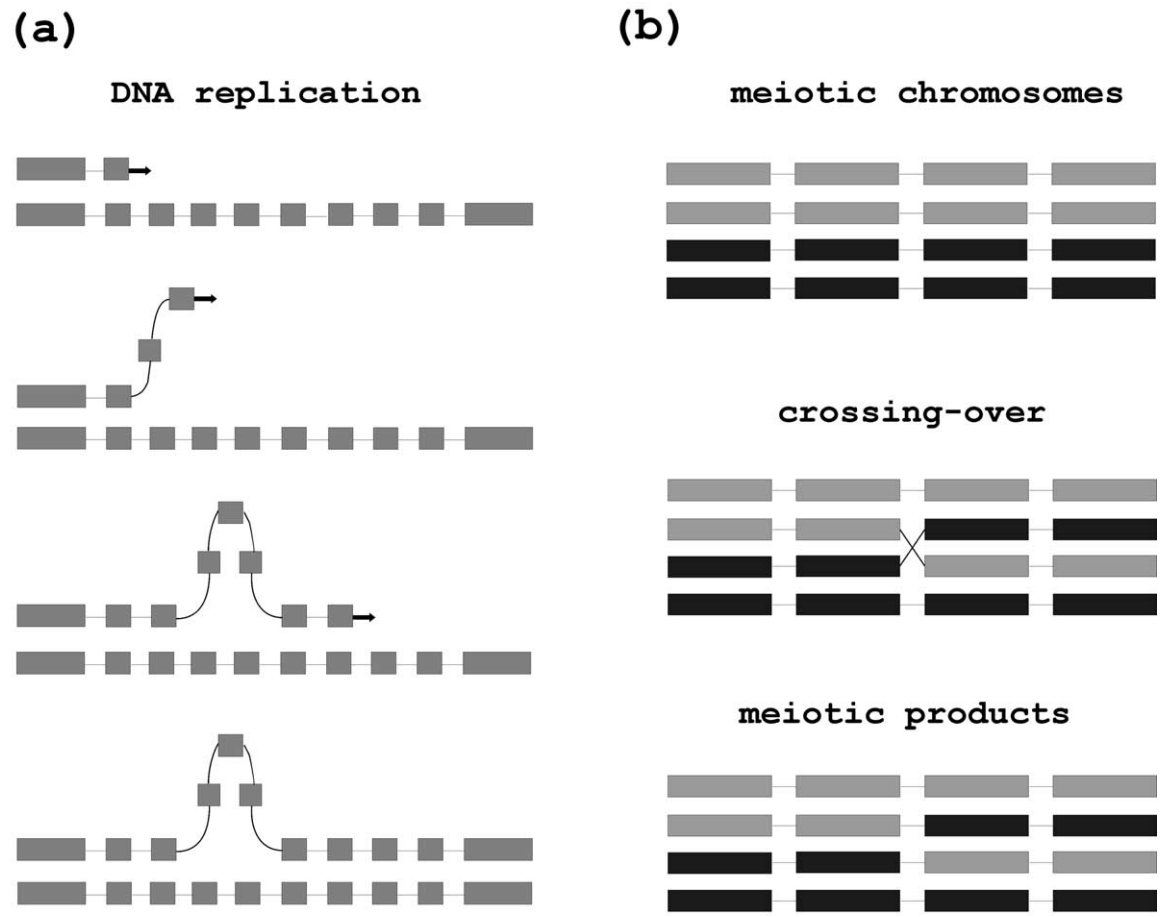


Fig. 5. Major recombination mechanisms putatively involved in *MSP-1* evolution: (a) strand slippage during DNA replication generating variation in repeat motifs within the same allelic family; and (b) crossing-over between blocks of sequences of different allelic families during meiosis. In the left panel (a), the nascent (upper) strand becomes displaced from the parental (lower) strand during DNA replication and reanneles out of phase with it, generating an insertion of a repeat motif (square) without changing non-repetitive flanking sequences (rectangles). A similar slippage event in the parental strand would result in the deletion of one repeat motif from the nascent strand. In the right panel (b), four blocks of non-repetitive sequences belonging to each allelic family (either black or gray rectangles) are represented for two duplicated homologous chromosomes during meiosis. One crossing-over event leads to a new combination of dimorphic sequence blocks in two of the meiotic products.

Plasmodium chabaudi and *Plasmodium yoelii* (Jennings et al., 1998), suggesting that variation in repeats may have a strong impact on the overall structure and evolution of this locus in different species. In analogy with repeats in *CSP* and *MSP-2*, which are bracketed between non-repetitive sequences related to each other (Rich et al., 2000), variation in block 2 is generated while maintaining strong LD in non-repetitive flanking sequences in blocks 1, 2 and 3 (Conway et al., 1999, and Fig. 4). This indicates that repeats are cold spots for homologous recombination in these malarial genes. The absence of homologous recombination in this region may be a consequence of the polymorphism, that makes less possible helical pairing and crossing over because of the several mismatches. Alternatively, recombinants may be negatively selected, during the parasite cycle in the human host, due to functional constraints. Since the function of MSP-1 is unknown, the nature of these putative functional constraints remains elusive.

A role for block 2 repeats in naturally acquired immunity has recently been determined (Conway et al., 2000a). If block 2 repeats are targeted by protective antibodies, as further suggested by in vitro studies (Locher et al., 1996), does variation in repeat arrays constitute a possible immune evasion mechanism? Available data indicate that the answer is yes. Narrowly specific antibody recognition of repetitive block 2 variants within the K1 and MAD20 group of alleles has recently been described for malaria-exposed Brazilians (Silveira et al., 1999) and Africans (Jouin et al., 2001). Moreover, both murine (Locher et al., 1996) and human (Sowa et al., 2001) monoclonal antibodies are able to discriminate between repeat variants within the same group of alleles. The monoclonal antibody isolated by Sowa et al. (2001) from a malaria patient-derived phage display antibody library, for example, discriminates between the MAD20 and HB3 variants of block 2, although differences between them are restricted to two additional copies of the hexapeptide SGGSVA present in MAD20, a variation that can readily be generated by mitotic recombination events. Similar phenomena have been recently described in the

antibody recognition of MSP-2 repeat variants (Tonhosolo et al., 2001).

The variant-specificity of naturally acquired antibodies to block 2 repeats supports the notion that parasites carrying new repeat variants stochastically generated during mitosis may evade immune responses and be positively selected (Rich et al., 2000). However, block 2 is a major, but not the only putative target of protective antibodies within MSP-1 (O'Donnell et al., 2001), and several other surface antigens are involved in immunity against *P. falciparum* blood stages. Moreover, 45% of clinical isolates of *P. falciparum* so far examined in Brazil (Silveira et al., 1999), but only 11% of those in Vietnam (Ferreira et al., 1998c), carry the non-repetitive RO33-type version of block 2, suggesting that the impact of diversity in block 2 repeats on the acquisition of antimalarial immunity and immune evasion may vary, at the population level, in different areas of endemicity.

The relation between LD and genetic map distance between SNPs may be used to estimate effective recombination rates in parasite populations, and thus to determine the relative role of meiotic recombination in generating new alleles. When meiotic recombination is frequent, a significant negative correlation between nucleotide distance and LD is expected. Rich et al. (1997) detected strong LD between pairs of SNPs flanking the central repetitive region, independent of the distance of the intervening makers (up to 1100 bp), in 25 complete sequences of the *CSP* gene of *P. falciparum*. This supports the notion that mechanisms other than homologous recombination play a role in generating most variation in this major vaccine-candidate antigen (Rich et al., 1997). In contrast, the failure to detect LD between dimorphic sites of *MSP-1* separated by more than 1 kb in five of six African *P. falciparum* populations was interpreted as evidence that these parasites frequently undergo meiotic recombination (Conway et al., 1999). Nevertheless, this analysis disregarded the full linkage between polymorphisms situated across the 3.7-kb-long stretch between blocks 6 and 16 in all *MSP-1* alleles so far analyzed, regardless of their geographical origin (Miller et al., 1993; Kaneko et al., 1997; Tanabe et al., 2000, and this paper). Therefore, LD may be maintained over long genetic map distances within the *MSP-1* locus in worldwide parasite populations, perhaps as a result of natural selection favoring a few possible structural conformations of this part of the molecule.

Variable levels of LD were found when polymorphisms within 5′ and 3′ sites flanking blocks 6–16 (instead of those within blocks 6–16) were analyzed. Interestingly, some LD between 5′ and 3′ sites within *MSP-1* was detected in all non-African parasite samples so far examined (Sakihama et al., 1999, 2001; Tanabe et al., 2000; Silveira et al., 2001, and Fig. 4). These findings have been suggested to reflect possible epistatic interactions between epitopes situated at the N-terminal and C-terminal regions of *MSP*-1 (Sakihama et al., 1999; Tanabe et al., 2000). It has recently been suggested that the folding of the native molecule on merozoite's surface creates structural interactions that can explain the finding that antibodies recognizing N-terminal epitopes are able to abrogate binding of protective antibodies to the C-terminal region of this antigen (Patiño et al., 1997). Although significant LD occurred in the *MSP-1* gene of *P. falciparum* populations from mesoendemic southern Vietnam (Tanabe et al., 2000, and Fig. 4b), this locus displayed more diversity in local parasites than in those from holoendemic northern Tanzania, as judged by the number of different combinations of dimorphic blocks observed in each population (Ferreira et al., 1998b,c). Thus, strong LD, a surrogate marker of low meiotic recombination rates, may co-exist with extensive antigenic diversity in MSP-1.

The same pair of dimorphic sites within *MSP-1* may be either strongly linked or independent in different endemic areas (Tanabe et al., 2000). The contrasting pattern of LD between block 4b and other 5′ and 3′ sites in Brazil and Vietnam (Figs. 4a,b) illustrates this situation. However, it remains unclear why five pairs of *MSP-1* alleles (four of them from Vietnam) analyzed here differed only in block 4b, but were otherwise identical, including the highly variable block 2 repeats. If we consider that these pairs of alleles, from sympatric parasites, did not arise independently but were derived from each other, the alternative block 4b sequence has not been introduced by a single crossover during homologous recombination. This may result from double crossover events during homologous recombination. However, it seems simpler to think that gene conversion, that may be either interhelical or intrahelical, mitotic or meiotic, has created this and similar 'patchwork' patterns of genetic variation in blocks 3 and 4 of *MSP-1* (Miller et al., 1993) and elsewhere in the genome of *P. falciparum* (Su et al., 1999).

In conclusion, although there is no doubt that meiotic recombination may create extensive genetic variation in *P. falciparum* (Kerr et al., 1994), we suggest that its frequency and role in *MSP-1* evolution in natural parasite populations may have been overestimated. If new alleles are frequently generated during the clonal (mitotic) propagation of parasites, antigenic diversity in *P. falciparum* may be less constrained by certain parameters, such as the intensity of malaria transmission and the abundance of genetically distinct clones available for cross-mating, than it would be if meiotic recombination were the predominant or sole diversifying mechanism. Further population-based studies of sequences of *MSP-1* and other repetitive *P. falciparum* antigens will help address this issue in the near future.

Acknowledgements

This study received financial support from the UNDP/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases,

the Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Toyota Foundation. W.L.R. and A.P.T. were supported by studentships from FAPESP, and M.U.F is a recipient of a research scholarship from CNPq.

References

Babiker, H.A., Walliker, D., 1997. Current views on the population structure of *Plasmodium falciparum*: implications for control. Parasitol. Today 13, 262–267.

Conway, D.J., Roper, C., Oduola, A.M.J., Arnot, D.E., Kremsner, P.G., Grobusch, M.P., Curtis, C.F., Greenwood, B.M., 1999. High recombination rate in natural populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA 96, 4506–4511.

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

Conway, D.J., Fanello, C., Lloyd, J.M., Al-Joubori, B.M.A.S., Baloch, A.H., Somanath, S.D., Roper, C., Oduola, A.M.J., Mulder, B., Póvoa, M.M., Singh, B., Thomas, A.W., 2000b. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. Mol. Biochem. Parasitol. 111, 163–171.

Ferreira, M.U., Liu, Q., Kaneko, O., Kimura, M., Tanabe, K., Kimura, E.A.S., Katzin, A.M., Isomura, S., Kawamoto, F., 1998a. Allelic diversity at the merozoite surface protein-1 locus of *Plasmodium falciparum* in clinical isolates from the southwestern Brazilian Amazon. Am. J. Trop. Med. Hyg. 59, 474–480.

Ferreira, M.U., Liu, Q., Kimura, M., Ndawi, B.T., Tanabe, K., Kawamoto, F., 1998b. Allelic diversity in the merozoite surface protein-1 and epidemiology of multiple-clone *Plasmodium falciparum* infections in northern Tanzania. J. Parasitol. 84, 1286–1289.

Ferreira, M.U., Liu, Q., Zhou, M., Kimura, M., Kaneko, O., Thien, H.V., Isomura, S., Tanabe, K., Kawamoto, F., 1998c. Stable patterns of allelic diversity at the *Merozoite Surface Protein-1* locus of *Plasmodium falciparum* in clinical isolates from southern Vietnam. J. Eukaryot. Microbiol. 45, 131–136.

Jennings, G.J., Toebe, C.S., van Belkum, A., Wiser, M.F., 1998. The complete sequence of *Plasmodium berghei* merozoite surface protein-1 and its inter- and intra-specific variability. Mol. Biochem. Parasitol. 93, 43–53.

Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M., Pluschke, G., 2000. Sequence diversity of the merozoite surface protein 1 of *Plasmodium falciparum* in clinical isolates from the Kilombero District, Tanzania. Acta Trop. 74, 51–61.

Jongwutiwes, S., Tanabe, K., Nakazawa, S., Yanagi, T., Kanbara, H., 1992. Sequence variation in the tripeptide repeats and T cell epitopes in P190 (MSA-1) of *Plasmodium falciparum* from field isolates. Mol. Biochem. Parasitol. 51, 81–90.

Jouin, H., Rogier, C., Trape, J.F., Mercereau-Puijalon, O., 2001. Fixed, epitope-specific, cytophilic antibody response to the polymorphic block 2 domain of the *Plasmodium falciparum* merozoite surface antigen MSP-1 in humans living in a malaria-endemic area. Eur. J. Immunol. 31, 539–550.

Kaneko, O., Kimura, M., Kawamoto, F., Ferreira, M.U., Tanabe, K., 1997. *Plasmodium falciparum*: allelic variation in the merozoite surface protein 1 gene in wild isolates from southern Vietnam. Exp. Parasitol. 86, 45–57.

Kerr, P.J., Ranford-Cartwright, L.C., Walliker, D., 1994. Proof of intragenic recombination in *Plasmodium falciparum*. Mol. Biochem. Parasitol. 66, 241–248.

Lewontin, R.C., Kojima, K., 1960. The evolutionary dynamics of complex polymorphisms. Evolution 14, 450–472.

Locher, C.P., Tam, L.Q., Chang, S.P., McBride, J.S., Siddiqui, W.A., 1996. *Plasmodium falciparum*: gp195 tripeptide repeat-specific monoclonal antibody inhibits parasite growth in vitro. Exp. Parasitol. 84, 74–83.

Miller, L.H., Roberts, T., Shahabuddin, M., McCutchan, T.F., 1993. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol. Biochem. Parasitol. 59, 1–14.

O'Donnell, R.A., de Koning-Ward, T.F., Burt, R.A., Bockarie, M., Reeder, J.C., Cowman, A.F., 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. J. Exp. Med. 193, 1403–1412.

Patiño, J.A., G, ., Holder, A.A., McBride, J.S., Blackman, M.J., 1997. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. J. Exp. Med. 186, 1689–1699.

Qari, S.H., Shi, Y.P., Goldman, I.F., Nahlen, B.L., Tibayrenc, M., Lal, A.A., 1998. Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen. Mol. Biochem. Parasitol. 92, 241–252.

Rich, S.M., Hudson, R.R., Ayala, F.J., 1997. *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. Proc. Natl. Acad. Sci. USA 94, 13040–13045.

Rich, S.M., Ferreira, M.U., Ayala, F.J., 2000. The origin of antigenic diversity in *Plasmodium falciparum*. Parasitol. Today 16, 390–395.

Sakihama, N., Kimura, M., Hirayama, K., Kanda, T., Na-Bangchang, K., Jongwutiwes, S., Conway, D., Tanabe, K., 1999. Allelic recombination and linkage disequilibrium within *Msp-1* of *Plasmodium falciparum*, the malignant human malaria parasite. Gene 230, 47–54.

Sakihama, N., Kaneko, A., Hattori, T., Tanabe, K., 2001. Limited recombination in merozoite surface protein-1 alleles of *Plasmodium falciparum* on islands. Gene 279, 41–48.

Silveira, L.A., Dorta, M.L., Kimura, E.A.S., Katzin, A.M., Kawamoto, F., Tanabe, K., Ferreira, M.U., 1999. Allelic diversity and antibody recognition of *Plasmodium falciparum* merozoite surface protein 1 during hypoendemic malaria transmission in the Brazilian Amazon Region. Infect. Immun. 67, 5906–5916.

Silveira, L.A., Ribeiro, W.L., Kirchgatter, K., Wunderlich, G., Matsuoka, H., Tanabe, K., Ferreira, M.U., 2001. Sequence diversity and linkage disequilibrium within the merozoite surface protein-1 (*Msp-1*) locus of *Plasmodium falciparum*: a longitudinal study in Brazil. J. Eukaryot. Microbiol. 48, 433–439.

Sowa, K.M.P., Cavanagh, D.R., Creasey, A.M., Raats, J., McBride, J., Sauerwein, R., Roeffen, W.F., Arnot, D.E., 2001. Isolation of a monoclonal antibody from a malaria patient-derived phage display library recognizing the block 2 region of *Plasmodium falciparum* merozoite surface protein-1. Mol. Biochem. Parasitol. 112, 143–147.

Su, X.-Z., Ferdig, M.T., Huang, Y., Huynh, C.Q., Liu, A., You, J., Wootton, J.C., Wellems, T.E., 1999. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. Science 286, 1351–1353.

Tanabe, K., Sakihama, N., Nakamura, Y., Kaneko, O., Kimura, M., Ferreira, M.U., Hirayama, K., 2000. Selection and genetic drift of polymorphisms within the merozoite surface protein-1 gene of *Plasmodium falciparum*. Gene 241, 325–331.

Tonhosolo, R., Wunderlich, G., Ferreira, M.U., 2001. Differential antibody recognition of four allelic variants of the merozoite surface protein-2 (MSP-2) of *Plasmodium falciparum*. J. Eukaryot. Microbiol. 48, 556–564.

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., Hartl, D.L., 2001. Recent origin of *Plasmodium falciparum* from a single progenitor. Science 293, 482–484.