Malaria Journal

Research

Genetic relatedness of *Plasmodium falciparum* isolates and the origin of allelic diversity at the merozoite surface protein-1 (MSP-1) locus in Brazil and Vietnam

Erika HE Hoffmann[1], Paulo EM Ribolla[2] and Marcelo U Ferreira*[1]

Address: [1]Departamento de Parasitologia, Instituto de Ciências Biomédicas da Universidade de São Paulo, Av. Prof. Lineu Prestes 1374, 05508-900 São Paulo (SP), Brazil and [2]Departamento de Parasitologia, Instituto de Biociências da Universidade Estadual Paulista, Caixa Postal 510, 18618-000 Botucatu (SP), Brazil

Email: Erika HE Hoffmann - hoffmann@usp.br; Paulo EM Ribolla - pribolla@ibb.unesp.br; Marcelo U Ferreira* - muferrei@usp.br

* Corresponding author

Abstract

Background: Despite the extensive polymorphism at the merozoite surface protein-1 (MSP-1) locus of *Plasmodium falciparum*, that encodes a major repetitive malaria vaccine candidate antigen, identical and nearly identical alleles frequently occur in sympatric parasites. Here we used microsatellite haplotyping to estimate the genetic distance between isolates carrying identical and nearly identical MSP-1 alleles.

Methods: We analyzed 28 isolates from hypoendemic areas in north-western Brazil, collected between 1985 and 1998, and 23 isolates obtained in mesoendemic southern Vietnam in 1996. MSP-1 alleles were characterized by combining PCR typing with allele-specific primers and partial DNA sequencing. The following single-copy microsatellite markers were typed : Polyα, TA42 (only for Brazilian samples), TA81, TA1, TA87, TA109 (only for Brazilian samples), 2490, ARAII, PfG377, PfPK2, and TA60.

Results: The low pair-wise average genetic distance between microsatellite haplotypes of isolates sharing identical MSP-1 alleles indicates that epidemic propagation of discrete parasite clones originated most identical MSP-1 alleles in parasite populations from Brazil and Vietnam. At least one epidemic clone propagating in Brazil remained relatively unchanged over more than one decade. Moreover, we found no evidence that rearrangements of MSP-1 repeats, putatively created by mitotic recombination events, generated new alleles within clonal lineages of parasites in either country.

Conclusion: Identical MSP-1 alleles originated from co-ancestry in both populations, whereas nearly identical MSP-1 alleles have probably appeared independently in unrelated parasite lineages.

Background

Despite its relatively recent expansion in human populations [1], *Plasmodium falciparum* displays extensive genetic variation in most surface antigens, affecting the development of effective immune responses [2]. Understanding the patterns and mechanisms of DNA sequence variation in major *P. falciparum* surface antigens is important for predicting the efficacy of immunization strategies [3].

The merozoite surface protein-1 (MSP-1) of *P. falciparum* is a prime malaria vaccine candidate antigen. Its coding sequence may be divided into 17 blocks, according to the levels of inter-allele divergence (Figure 1A). Most variation is dimorphic: sequences may be grouped into one of two allelic families (K1 and MAD20). Block 2 represents an exception to dimorphism, since in addition to K1-type and MAD20-type sequences, that contain degenerate tripeptide repeats, an apparently non-repetitive allele known as RO33 is commonly found. Genetic diversity at the MSP-1 locus may be generated by exchanging blocks of sequences during sexual (meiotic) recombination [4] and by putative strand-slippage events during the asexual (mitotic) replication of parasites leading to rearrangements of block 2 tripeptide repeats [5], but the relative contribution of each recombination mechanism remains unknown. High meiotic recombination rates within MSP-1 have been estimated for parasites in areas of intense malaria transmission in Africa, where most human infections consist of mixtures of genetically distinct clones [6]; most new MSP-1 alleles, therefore, originate from cross-mating followed by meiotic recombination. Mitotic recombination events may be important in generating new MSP-1 alleles in areas of low to intermediate levels of malaria transmission outside Africa, where meiotic recombination rates at the MSP-1 locus are substantially lower [5,7].

Despite the huge potential for variation at the MSP-1 locus, identical alleles often occur in areas of low and moderate malaria transmission [5,7]. The epidemic propagation of a few discrete *P. falciparum* clones might explain these findings. This hypothesis is consistent with the epidemic structure (defined by Maynard Smith and colleagues [8] as the result of short-term expansions of highly successful clonal lineages in an otherwise panmictic population) that characterize most *P. falciparum* populations outside Africa, including those in Brazil [9] and Vietnam [10]. Since identical MSP-1 alleles have been found in the same area in Brazil over more than one decade [5], some epidemic clones could have remained relatively unchanged over several generations. In addition, groups of nearly identical MSP-1 alleles, differing only in the number and arrangement of block 2 repeats, but identical elsewhere in the gene, are also prevalent in *P. falciparum* from Brazil and Vietnam [5]. The hypothesis that nearly identical alleles have originated from each other by strand-slippage events within repeat arrays during the mitotic propagation of parasites [5] remains to be proved. Hypotheses regarding the origin of identical and nearly identical MSP-1 alleles were addressed here by microsatellite characterization of the overall genetic relatedness of parasites carrying known variants of this gene.

Subjects and Methods
Parasite samples and MSP-1 typing
We analysed two sets of human blood samples with apparently single-clone *P. falciparum* infections, as judged by typing MSP-1 polymorphisms and single-copy DNA microsatellite markers. The first set comprised 28 isolates from hypoendemic areas in the towns of Porto Velho, Ariquemes and Guajará-Mirim, all in the state of Rondônia, north-western Brazil, collected in 1985, 1997 and 1998, while the second set comprised 23 isolates obtained in the mesoendemic town of Bao Loc, southern Vietnam, in January-December 1996. Blood samples were collected, after informed consent, at malaria treatment facilities, and all patients had symptomatic *P. falciparum* infections when enrolled. The average proportions of multiple-clone *P. falciparum* infections, as estimated by microsatellite typing, were 11.5% in Rondônia, Brazil [9], and 39.6% in Bao Loc, Vietnam [10]. MSP-1 alleles had been previously characterized in these isolates by combining PCR typing with allele-specific primers and partial DNA sequencing (Figure 1A); each allele was defined as a unique combination of block 2 and 17 haplotypes and other polymorphisms across the molecule (Figure 1B) [5].

DNA microsatellite typing
The following single-copy microsatellite markers were typed (chromosome location in parentheses): Polyα (4), TA42 (only for Brazilian samples) (5), TA81 (5), TA1 (6), TA87 (6), TA109 (only for Brazilian samples) (6), 2490(10), ARAII (11), PfG377 (12), PfPK2 (12), and TA60 (13). We used no marker located on chromosome 9, where the MSP-1 locus is found, to avoid possible biases in overall distance estimates originated from hitchhiking due to selection on this locus. Microsatellite loci were PCR-amplified according to previously described protocols [11]. Length variation of PCR products (a single PCR product was obtained for each locus in each isolate) was measured on either ABI377 (all Brazilian samples) or ABI3100 (all Vietnamese samples) automated DNA sequencers (Applied Biosystems). Part of these microsatellite data had been published elsewhere [9,10].

Data analysis
The levels of genetic diversity at each microsatellite locus were estimated for both parasite samples by calculating the theoretical expected heterozygosity H as follows:

$$H = [n/(n-1)][1 - \sum p_i^2]$$, where n is the number of isolates sampled and p_i is the frequency of each allele at a given locus [9]. Estimates of overall pair-wise genetic distances between isolates from either country, obtained from microsatellite data (1 – proportion of alleles shared between haplotypes), were used to construct Fitch-Margoliash trees using the PHYLIP, version 3.5c, software package, distributed by its author (J. Felsenstein) at http://
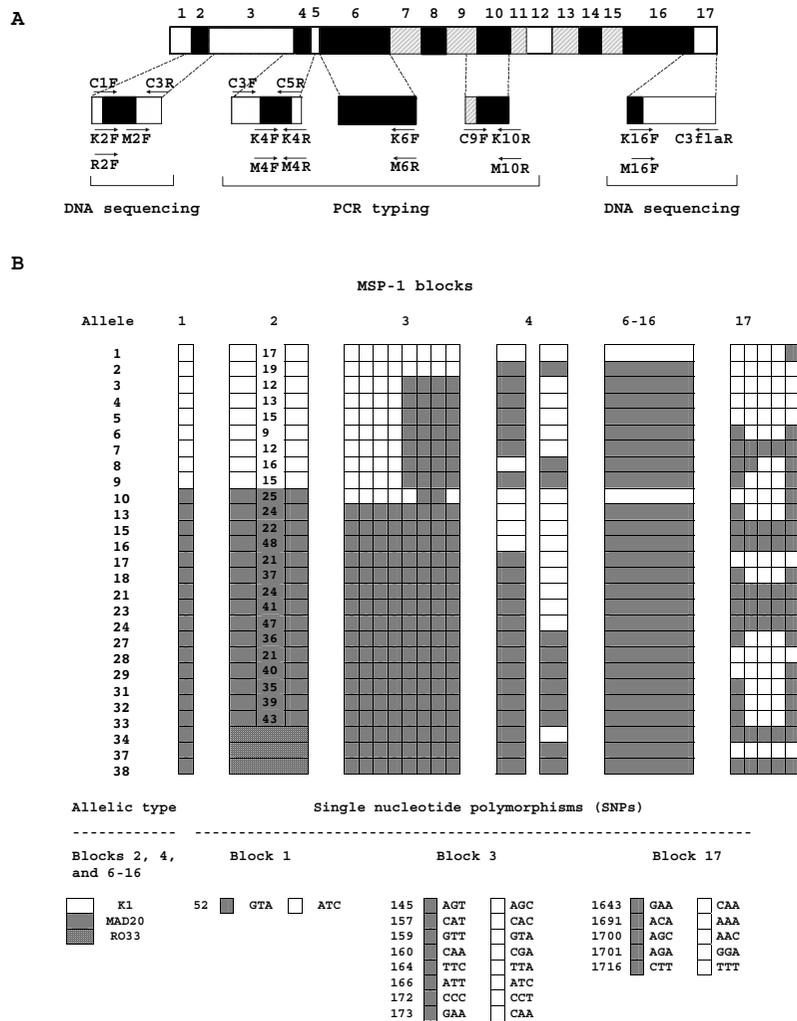
Figure 1
Diversity in the *MSP-1* gene of *Plasmodium falciparum*. In **A**, conserved, semiconserved and variable blocks of the *MSP-1* gene are represented as open, hatched and filled boxes, respectively. The location and orientation of primers used for typing and sequencing this locus are also shown. Procedures are described elsewhere [5]. In **B**, the 27 different *MSP-1* alleles found in the analysed sample set from Brazil and Vietnam are represented. Alleles were defined as unique combinations 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, as determined by DNA sequencing; and (c) nucleotide polymorphisms in blocks 1, 3 and 17, as determined by DNA sequencing. Alleles and repeat haplotypes were numbered according to Ferreira and colleagues [5]. K1-type repeat haplotypes differ in the number and arrangement of SGT and SGP motifs, while MAD20-type repeat haplotypes differ in the number and arrangement of SGG, SVA, SVT and SKG motifs. Codon numbers are given according to Miller and colleagues [13].

evolution.genetics.washington.edu. To test whether identical MSP-1 alleles within each country occurred predominantly in genetically related parasites, we first compared the overall genetic relatedness (based on pair-wise genetic distances defined as above) between isolates from each country sharing or not sharing MSP-1 alleles, by using two-sample randomization tests implemented in version 2.5 of the PopTools software (written by G. Hood and available at: http://www.cse.csiro.au/poptools). Monte-Carlo simulations (10,000 iterations) were used to estimate two-tailed *P* values. This approach was also used to test whether nearly identical MSP-1 alleles (defined as above) tended to occur in genetically related isolates. The second approach to examine the genetic relatedness of isolates sharing or not sharing identical MSP-1 alleles investigated the correlation between similarity matrices obtained with microsatellite data and MSP-1 typing data. We created two model matrices to describe the MSP-1 data set. In the first model matrix (hereafter Model I), values of 0 and 1 were assigned to pair-wise comparisons of isolates with identical and different MSP-1 alleles, respectively. In Model II, three levels of similarity were considered: values of 0, 0.5 and 1 were assigned to pair-wise comparisons of isolates with identical, nearly identical and different MSP-1 alleles, respectively. Mantel tests were used to assess the correlation between the microsatellite similarity matrix and each of the MSP-1 model matrices [12]. Coefficients of determination (r^2) were calculated and two-tailed *P* values were obtained by Monte Carlo simulation with 6,000 permutations performed with the PopTools software. Finally, we compared the overall genetic relatedness of isolates collected in Brazil in the same or different decades.

Results and Discussion
Distribution of MSP-1 alleles
Twenty-seven unique MSP-1 alleles were found, 10 in Brazil (alleles # 2, 3, 4, 5, 6, 7, 17, 28, 34 and 37) and 17 in Vietnam (alleles # 1, 8, 9, 10, 13, 15, 16, 18, 21, 23, 24, 27, 29, 31, 32, 33 and 38). Five MSP-1 alleles in Brazil and four in Vietnam were shared by two or more isolates, but no allele was shared by parasites from both countries (Figure 1B). Moreover, four groups of nearly identical MSP-1 alleles were found in parasites from the same country (one in Brazil and three in Vietnam). Alleles within each of these groups differed in the number and arrangement of tripeptide repeats in block 2, but were identical elsewhere in the MSP-1 gene [5].

Genetic relatedness of Brazilian isolates
The number of alleles at each microsatellite locus in Brazilian isolates ranged between 2 and 5 (average, 2.64), with relative low estimates of expected heterozygosity (0.14–0.49; average, 0.38) (Table 1). Most Brazilian isolates sharing identical MSP-1 alleles tended to cluster together in the Fitch-Margoliash tree based on pair-wise genetic distances of 11-locus microsatellite haplotypes (Figure 2). Three out of four groups of Brazilian isolates with identical multilocus haplotypes (as indicated by the zero length of terminal branches in the tree) comprised parasites sharing identical MSP-1 alleles collected during the same decade. The average genetic distance of isolates sharing identical MSP-1 alleles (0.304; 50 pair-wise comparisons) was significantly lower (*P* < 0.0001) than that estimated for isolates with different MSP-1 alleles (0.431; 328 pair-wise comparisons). Microsatellite data thus indicate that most identical MSP-1 alleles in this population occurred in genetically related parasites, possibly due to the epidemic propagation of a few discrete parasite lineages. Moreover, they suggest that at least one of these epidemic lineages changed little over more than one decade, since isolates sharing the MSP-1 allele #34 which were collected both in 1985 (RO37 and RO41) and 1998 (13OC) clustered together in the tree (Figure 2).

We next examined the overall genetic relatedness of Brazilian isolates with nearly identical MSP-1 alleles (alleles #3, #4 and #5), which differed only in the number of SGT repeats in their K1-type block 2 sequences (GenBank accession numbers: AF509630, AF509632-3 and AF509705) (Figure 1B). The average genetic distance of Brazilian isolates with nearly identical MSP-1 alleles (0.436; 5 pair-wise comparisons) was similar to that calculated for local isolates with different MSP-1 alleles (0.431; 323 pair-wise comparisons) (*P* = 0.810). The lack of evidence for co-ancestry of these isolates argues against the hypothesis that nearly identical MSP-1 alleles in Brazil were generated by mitotic recombination events in block 2 repeats within clonal lineages of parasites [5].

Figure 2 shows some temporal clustering of microsatellite haplotypes in Brazil. In fact, the average pair-wise genetic distance of isolates collected in the same decade (0.398; 218 pair-wise comparisons) was significantly lower (*P* = 0.010) than that estimated for isolates collected in different decades (0.438; 160 pair-wise comparisons). To test whether the date of sample collection confounded the previous analyses of genetic distance, we analysed separately 12 isolates collected in 1985 and 16 isolates collected in 1997–98. Isolates sharing identical MSP-1 alleles collected in 1985 (0.364; 9 pair-wise comparisons) and 1997–98 (0.282; 31 pair-wise comparisons) were again more closely related to each other than those with different MSP-1 alleles collected during the corresponding decade (0.405 [57 pair-wise comparisons] and 0.428 [89 pair-wise comparisons], respectively). Differences in genetic distance were significant for comparisons within 1997–98 (*P* < 0.0001) but not for those within 1985 (*P* = 0.441), probably due to the smaller sample size.

MSP-1 allele
and year of
collection

3	1985
7	1985
34	1985
4	1985
28	1998
17	1998
6	1998
6	1997
2	1997
6	1998
6	1997
6	1997
6	1998
6	1997
34	1985
28	1997
3	1998
34	1985
34	1998
34	1985
28	1985
28	1997
28	1997
28	1998
5	1985
37	1985
37	1985
37	1985

AM159
AM82
AM100
AM170
10OC
D
31OC
62
49B
G
32A
4
H
27
B36
42
11OC
RO37
13OC
RO41
AM89
34
45
22OC
AM209
AM84
AM160
AM255

0.1
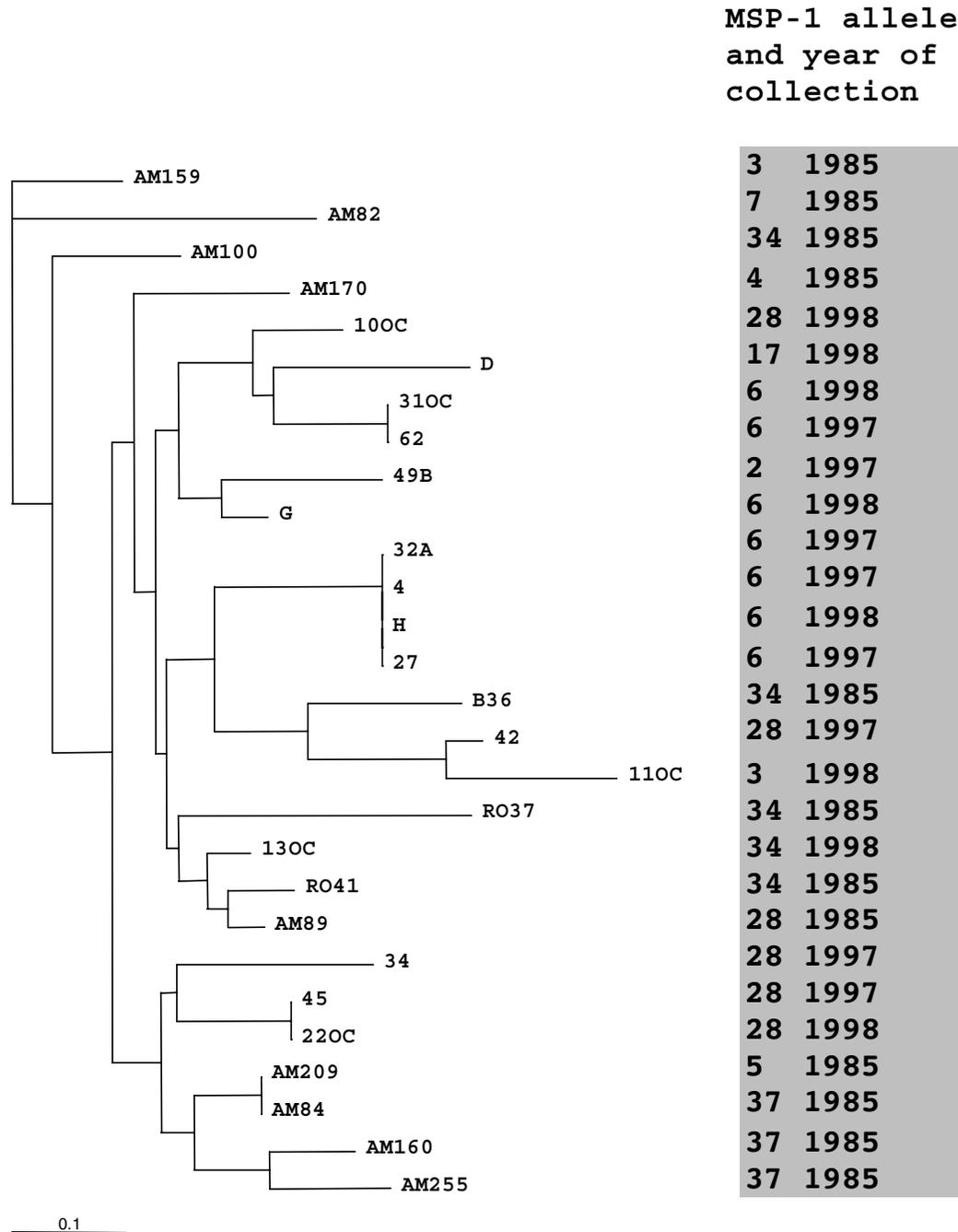
Figure 2
Fitch-Margoliash tree showing the relationships between 11-locus microsatellite haplotypes in 28 *Plasmodium falciparum* isolates from Brazil. *MSP-1* alleles (numbered according to Fig. 1B) and dates of collection are also shown.

Table 1: Number of different alleles and expected heterozygozity at microsatellite loci in *Plasmodium falciparum* isolates from Brazil (n = 28) and Vietnam (n = 23).

Locus	No. of alleles		Expected heterozygosity	
	Brazil	Vietnam	Brazil	Vietnam
Poly - α	4	11	0.48	0.91
TA42	3	Not tested	0.32	Not tested
TA81	3	6	0.39	0.74
TA1	5	4	0.28	0.71
TA87	3	4	0.43	0.69
TA109	3	Not tested	0.14	Not tested
2490	2	4	0.48	0.52
ARAII	2	7	0.39	0.82
PfG377	2	3	0.49	0.69
PfPk2	2	5	0.39	0.79
TA60	3	5	0.37	0.65
Average	2.64	5.44	0.38	0.72

Genetic relatedness of Vietnamese isolates

The levels of genetic diversity of Vietnamese isolates, based on 9 microsatellite loci, were about two times higher than those observed in Brazil: the number of alleles at each locus ranged between 3 and 11 (average, 5.44), with estimates of expected heterozygosity ranging between 0.52 and 0.91 (average, 0.72) (Table 1). The overall genetic relatedness of Vietnamese isolates is represented in Figure 3. Both pairs of isolates with identical haplotypes (V57-V10 and BL199-BL185) also shared identical MSP-1 alleles (respectively #13 and #33). The average genetic distance of isolates sharing identical MSP-1 alleles (0.568; 9 pair-wise comparisons) was significantly lower (P = 0.008) than that estimated for isolates with different MSP-1 alleles (0.715; 244 pair-wise comparisons). Thus, microsatellite data also provided evidence that identical MSP-1 alleles in Vietnam tend to occur in genetically related parasites.

We also examined the genetic relatedness within three groups of Vietnamese isolates with nearly identical MSP-1 alleles: (a) alleles #15 and #16, (b) alleles #21, #23 and #24, and (c) alleles #31, #32 and #33. Alleles within each group only differed in the number of SGG, SGA and SKG repeats in their MAD20-type block 2 sequences (GenBank accession numbers: AF509642, AF509645-7, AF509654, AF509670, AF509679, AF509686, AF509691-2 and AF509702) (Fig. 1B). No significant difference (P = 0.615) was found when the average genetic distance of Vietnamese isolates with nearly identical MSP-1 alleles (0.710; 13 pair-wise comparisons) was compared with that of isolates with quite different MSP-1 alleles (0.715, 231 pair-wise comparisons). This result supports the conclusion that nearly identical MSP-1 alleles in Vietnam have not

been recently created by variation in block 2 repeats within clonal parasite lineages.

Matrix correlation analysis

The association between MSP-1 allele sharing and low overall genetic distance between isolates was further confirmed by the significant correlation (r^2 = 0.082, P < 0.00001 for Brazil and r^2 = 0.025, P = 0.014 for Vietnam) between the microsatellite-derived similarity matrices and Model I MSP-1 matrices. The magnitude of the coefficients of determination remained nearly unchanged when the microsatellite-derived similarity matrices were compared to Model II MSP-1 matrices (r^2 = 0.079, P < 0.00001 for Brazil and r^2 = 0.020, P = 0.025 for Vietnam), again indicating that pairs of isolates with nearly identical MSP-1 alleles could not be placed at intermediate levels of overall genetic distance, when compared with local isolates with either identical or different MSP-1 alleles.

Conclusions

The low overall genetic distance among microsatellite haplotypes of isolates sharing identical MSP-1 alleles supported the hypothesis that these identical alleles have originated from recent epidemic propagations of discrete parasite lineages within parasite populations from Brazil and Vietnam. Moreover, we found no evidence that isolates with nearly identical MSP-1 alleles (differing only in block 2 repeats) are genetically related. Although rearrangements in block 2 repeats may represent an important source of sequence variation in MSP-1 [2,5], our results present no evidence for this process during the relatively short time frame for which clonal lineages were maintained within parasite populations in Brazil and Vietnam. One possible explanation for these negative findings is that, in areas of low to moderate levels of malaria trans-
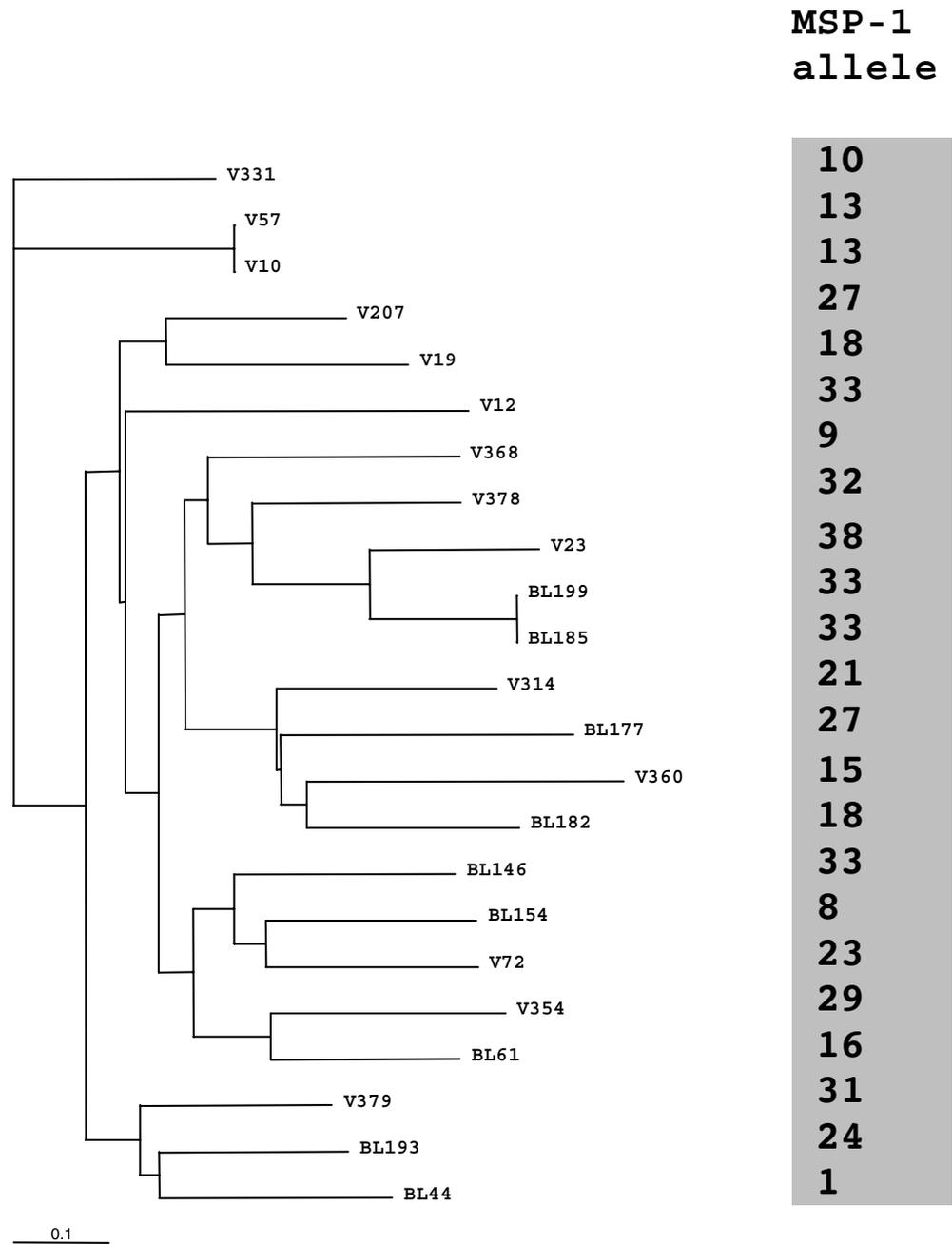
Figure 3
Fitch-Margoliash tree showing the relationships between 9-locus microsatellite haplotypes in 23 *Plasmodium falciparum* isolates from Vietnam. *MSP-1* alleles (numbered according to Fig. 1B) are also shown.

mission, selective pressure may be too weak to drive the emergence of new MSP-1 variants originated by putative mitotic recombination events during human infections, since most local *P. falciparum*-infected patients have not been able to develop effective variant-specific anti-parasite immunity.

Authors' Contributions

Field work was done at different malaria clinics in Rondônia, north-western Brazil, and at the Lam Dong Provincial Hospital of Bao Loc, southern Vietnam. EHEH and MUF performed the laboratory work and participated in data analysis, with significant help from PEMR in data analysis and interpretation of results. All authors read and approved the manuscript.

References

1. Joy DA, Feng XR, Mu JB, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P and Su X-Z: **Early origin and recent expansion of *Plasmodium falciparum*.** *Science* 2003, **300:**318-21.
2. Rich SM, Ferreira MU and Ayala FJ: **The origin of antigenic diversity in *Plasmodium falciparum*.** *Parasitol Today* 2000, **16:**390-95.
3. Hartl DL, Volkman SK, Nielsen KM, Barry AE, Day KP, Wirth DF and Winzeler EA: **The paradoxical population structure of *Plasmodium falciparum*.** *Trends Parasitol* 2002, **18:**266-72.
4. Kerr PJ, Ranford-Cartwright LC and Walliker D: **Proof of intragenic recombination in *Plasmodium falciparum*.** *Mol Biochem Parasitol* 1994, **66:**241-48.
5. Ferreira MU, Ribeiro WL, Tonon AP, Kawamoto F and Rich SM: **Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*.** *Gene* 2003, **304:**65-75.
6. Conway DJ, Roper C, Oduola AMJ, Arnot DE, Kremsner PG, Grobusch MP, Curtis CF and Greenwood BM: **High recombination rate in natural populations of *Plasmodium falciparum*.** *Proc Natl Acad Sci USA* 1999, **96:**4506-11.
7. Tanabe K, Sakihama N, Nakamura Y, Kaneko O, Kimura M, Ferreira MU and Hirayama K: **Selection and genetic drift of polymorphisms within the merozoite surface protein-1 gene of *Plasmodium falciparum*.** *Gene* 2000, **241:**325-31.
8. Maynard Smith J, Smith NH, O'Rourke M and Spratt BG: **How clonal are bacteria?** *Proc Natl Acad Sci USA* 1993, **90:**4384-8.
9. Anderson TJC, Haubold B, Williams JT, Estrada-Franco JC, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Withworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU and Day KP: **Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*.** *Mol Biol Evol* 2000, **17:**1467-82.
10. Ferreira MU, Nair S, Hyunh TV, Kawamoto F and Anderson TJC: **Microsatellite characterization of *Plasmodium falciparum* from cerebral and uncomplicated malaria patients in southern Vietnam.** *J Clin Microbiol* 2002, **40:**1854-7.
11. Anderson TJC, Su X-Z, Bockarie M, Lagog M and Day KP: **Twelve microsatellite markers for characterisation of *Plasmodium falciparum* from finger-prick blood samples.** *Parasitology* 1999, **119:**113-26.
12. Rodrigues FM, Diniz-Filho JAF, Bataus JAM and Bastos RP: **Hypothesis testing of genetic similarity based on RAPD data using Mantel tests and model matrices.** *Genetics Mol Biol* 2000, **25:**435-9.
13. Miller LH, Roberts T, Shahabuddin M and McCutchan TF: **Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1).** *Mol Biochem Parasitol* 1993, **59:**1-14.