

Plasmodium falciparum: Worldwide sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-2 (MSP-2)

Marcelo U. Ferreira^{a,b,*}, Daniel L. Hartl^b

^a Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes 1374, 05508-900 São Paulo (SP), Brazil

^b Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 01238, USA

Received 2 March 2006; received in revised form 6 April 2006; accepted 11 May 2006

Available online 21 June 2006

Abstract

We examined patterns and putative mechanisms of sequence diversification in the merozoite surface protein-2 (MSP-2) of *Plasmodium falciparum*, a major dimorphic malaria vaccine candidate antigen, by analyzing 448 *m*sp-2 alleles from all continents. We describe several nucleotide replacements, insertion and deletion events, frameshift mutations, and proliferations of repeat units that generate the extraordinary diversity found in *m*sp-2 alleles. We discuss the role of positive selection exerted by naturally acquired type- and variant-specific immunity in maintaining the observed levels of polymorphism and suggest that this is the most likely explanation for the significant excess of nonsynonymous nucleotide replacements found in dimorphic *m*sp-2 domains. Hybrid sequences created by meiotic recombination between alleles of different dimorphic types were observed in few (3.1%) isolates, mostly from Africa. We found no evidence for an extremely ancient origin of allelic dimorphism at the *m*sp-2 locus, predating *P. falciparum* speciation, in contrast with recent findings for other surface malarial antigens.

© 2006 Elsevier Inc. All rights reserved.

Index Descriptors and Abbreviations: Malaria; MSP-2; Evolution; Recombination; Allelic dimorphism; Antigenic diversity; Vaccine development; bp, base pairs; MSP-1, merozoite surface protein-1; MSP-2, merozoite surface protein-2; Myr, million years; RHR, repeat homology region; SD, standard deviation; SE, standard error of the mean; TMRCA, time to the most recent common ancestor

1. Introduction

Comparative sequence analyses are essential to infer evolutionary patterns and elucidate mechanisms that generate polymorphism in malarial surface antigens. The merozoite surface protein-2 (MSP-2), a leading candidate antigen for subunit malaria vaccines, comprises highly polymorphic central repeats flanked by unique variable domains and conserved N- and C-terminal domains (Fig. 1). Two features of MSP-2 are particularly relevant for evolutionary studies: (a) it is encoded by highly divergent alleles grouped into two dimorphic families or lineages, FC27 and 3D7 (Snewin

et al., 1991) and (b) sequence diversity in MSP-2 clearly hampers its recognition by naturally acquired (Taylor et al., 1995) and vaccine-induced (Flück et al., 2004) antibodies. Naturally acquired antibodies to MSP-2 have recently been associated with clinical immunity to malaria in Africa (Metzger et al., 2003; Polley et al., 2006).

Due to its polymorphism, the *m*sp-2 gene has been extensively used to type natural isolates of *P. falciparum* (Prescott et al., 1994; Franks et al., 2001). As a result, hundreds of coding sequences for the *m*sp-2 gene of parasites from geographically diverse locations have been logged in the GenBank database over the past 15 years. Although the molecular evolution of *m*sp-2 repeats (Fenton et al., 1991; Felger et al., 1997; Rich and Ayala, 2000) and the signatures of natural selection on this antigen (Hughes and Hughes, 1995; Escalante et al., 1998b) attracted some

* Corresponding author. Fax: +55 11 30917417.

E-mail addresses: muferre@usp.br, mferreir@fas.harvard.edu (M.U. Ferreira).

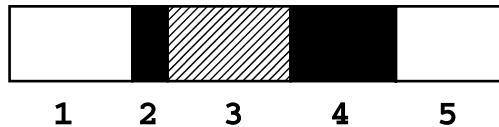


Fig. 1. Schematic representation of the merozoite surface protein-2 of *Plasmodium falciparum*, showing the conserved domains (blocks 1 and 5) as open boxes, the non-repetitive dimorphic domains (blocks 2 and 4) as black boxes, and the central polymorphic repeats (block 3) as a striped box. Domains are defined after Snewin et al. (1991); their borders, in the amino acid alignment shown in Supplementary Fig. 1, are as follows: block 1 (both types), positions 1–43; block 2, FC27-type, 44–59; block 2, 3D7-type, 44–50; block 3, FC27-type, 60–320; block 3, 3D7-type, 51–235; block 4, FC27-type, 321–368; block 4, 3D7-type, 321–410; and block 5 (both types), 411–484.

interest in the 1990s, the ever-increasing *mSP-2* sequence database remains largely unexplored. Here, we have analyzed all *mSP-2* gene sequences currently available in public databases to investigate: (a) geographical patterns of *mSP-2* diversity, (b) putative mechanisms of sequence divergence, (c) evidence for positive selection, and (d) the evolution of allelic dimorphism at the *mSP-2* locus.

2. Materials and methods

2.1. DNA sequences

Partial and complete coding sequences of *P. falciparum* *mSP-2* alleles available in the GenBank database by November 2005 were located by searching several alternative names for MSP-2 that appeared in the early literature (reviewed by Topolska et al., 2004). References cited in papers that described new *mSP-2* sequence data were also examined. We next sought to identify duplicate GenBank entries (two or more sequences derived from the same isolate) by using BLASTN sequence homology searches (Altschul et al., 1990). Three pairs of duplicate sequences were found. The first pair, differing by a single nucleotide, was derived from the reference clone 3D7 from West Africa (Accession numbers M28891 [Smythe et al., 1990] and AE001385 [Gardner et al., 1998]); here we used the most recent sequence, derived from the *P. falciparum* genome sequencing project. The two other pairs comprised full-length sequences identical to each other: (a) isolate FCC1/HN, from Hainan, China (Accession numbers AF334034 [Zheng et al., 2003] and AF217617 [Ma et al., unpublished, 1999]) and (b) isolate FCR-3, from Gambia (M60188 [Thomas et al., 1990] and M28890 [Elliott et al., 1990]). The duplicate GenBank entries for FCR-3 used different designations (FCR-3 and FMG) for the same parasite line maintained in continuous in vitro culture (Jensen and Trager, 1978). All other identical sequences identified in the database had been derived from independent parasite lines or isolates, and were thus maintained in the dataset.

The dataset comprised 448 sequences (43.5% FC27-type, 53.3% 3D7-type, and 3.1% 3D7/FC27 hybrids); the complete alignment of deduced amino acid sequences is shown

in Supplementary Fig. 1. GenBank Accession numbers and the names and geographic origins of all isolates from which sequences were derived have been entered in an Excel file, available from the authors upon request. Only 36 (8.0%) sequences, 18 FC27-type, 17 3D7-type and one hybrid, were full-length (~800 base pairs [bp]); 173 (38.6%) had not been previously described in formal publications. The geographic distribution of the isolates from which sequences were derived is shown in Table 1.

2.2. Data analysis

Nucleotide sequences were aligned using Clustal X (Thompson et al., 1997) with extensive manual editing. Dimorphic and polymorphic domains (blocks 2–4) were aligned within, but not between, allelic families, because their sequences are not homologous except for a few short segments (Rich and Ayala, 2000). Therefore, these domains were analyzed separately for each allelic family (FC27 and 3D7), while conserved domains (blocks 1 and 5) from isolates belonging to both allelic families were analyzed together. We searched for singleton polymorphisms in all domains under analysis. Separate analyses were made with: (a) a dataset comprising all full-length coding sequences available for a particular domain under analysis and (b) a restricted dataset from which sequences with singleton

Table 1

Geographic origin of *Plasmodium falciparum* isolates from which merozoite surface protein-2 gene sequences analyzed in this study were derived

Continent	Country	Dimorphic type			
		FC27	3D7	Hybrid	Total
Africa		65	65	11	141
	Ghana	12	29	1	42
	Gambia	9	16	2	27
	Gabon	24	0	0	24
	Tanzania	13	1	3	17
	Senegal	0	9	0	9
	Nigeria	2	3	4	9
	Other	5	7	1	13
South America		49	40	1	90
	Brazil	48	40	0	88
	Colombia	1	0	1	2
Asia		44	38	0	82
	Vietnam	18	25	0	43
	India	17	0	0	17
	China	6	3	0	9
	Thailand	1	5	0	6
	Iran	1	3	0	4
	Other	1	2	0	3
Oceania		33	79	0	112
	Papua New Guinea	18	56	0	74
	Indonesia (Irian Jaya) ^a	7	10	0	17
	Solomon Islands	3	9	0	12
	Vanuatu	5	4	0	9
Unknown		4	17	2	23
Total		195	239	14	448

^a For the purposes of this analysis, Irian Jaya was included in Oceania. No sequences from other parts of Indonesia were available.

polymorphisms in that given domain were excluded. Results described in the main text were obtained with the singleton-free dataset, from which most sequencing errors must have been removed, but some results obtained with the complete dataset are provided as supplementary information online, for comparison.

Estimates of π (average number of nucleotide substitutions per site between pairs of sequences in the dataset) and their standard deviations (SD) were obtained, for each *msp-2* domain, with DnaSP 4.0 (Rozas et al., 2003). Sites at which the alignment postulated a gap in any of the sequences, as well as those in the 18-bp frameshift identified in FC27-type alleles (Tonon et al., 2004; see Section 3) were excluded from the analysis. π values in tables are given for 1000 sites. DnaSP 4.0 was also used to measure haplotype diversity (Hd), defined as $Hd = 1 - \sum f_i^2$, where f_i is the frequency of the i -th haplotype in the sample. We used MEGA 3.0 software (Kumar et al., 2004) to estimate the average number of synonymous nucleotide substitutions per synonymous site (d_S) and the average number of nonsynonymous substitutions per nonsynonymous site (d_N), with their respective standard errors (SE) obtained by bootstrapping (Kumar et al., 2004), using the method of Nei and Gojobori (1986) with Jukes and Cantor (1969) correction. Again, d_S and d_N values in tables are given for 1000 sites. To ensure the comparability between our data and those previously reported in the literature (Rich et al., 1998; Tanabe et al., 2004), we used the same method for counting substitutions, despite the fact that the Jukes–Cantor correction may have undercorrected for overlapping substitutions when assuming that all changes are equally probable. Therefore, although absolute d_S and d_N values may have been somewhat underestimated in these studies, comparisons between estimates for different domains of the *msp-2* locus (this paper) and for other genes (reported in the literature) remain valid. The statistical significance of the difference between d_S and d_N was evaluated with one-tailed Z -tests (Nei and Kumar, 2000) with 1000 pseudoreplicates; when d_N significantly exceeded d_S , at the 5% confidence level, in a particular domain, we

took this as an evidence of positive selection favoring amino acid replacements (Hughes and Nei, 1988).

To examine the mode of evolution of FC27-type repeats, we aligned individual 96-bp (here termed R1) and 36-bp (here termed R2) repeat units and analyzed the distribution of the pairwise proportion of nucleotide differences among individual repeat units (Sherry et al., 1994). We hypothesize that recent duplication events would allow little sequence divergence among repeats, while ancient duplications would result in substantially higher numbers of pairwise mismatches.

We estimated the mean time to the most recent common ancestor (TMRCA) of pairs of *msp-2* alleles as $TMRCA = d_S/2\mu_s$, where μ_s is the rate of synonymous substitution per site per year (Hughes and Verra, 2001). We estimated μ_s for the *msp-2* locus as described (Tanabe et al., 2004), based on the number of synonymous substitutions between the orthologous genes in *P. falciparum* and the chimpanzee malaria parasite *P. reichenowi* (Accession number Y14731 [Dubbed et al., 1998]), which most probably diverged along with their hosts between 5 and 7 million years (Myr) ago (Escalante et al., 1998a), and within *P. falciparum*. The sequences compared to estimate μ_s comprised the conserved domains (blocks 1 and 5, 117 codons) of *P. falciparum* and *P. reichenowi msp-2*. To reduce the effect on TMRCA estimates of sequence errors in the GenBank database (Barry et al., 2003), analyses were done with the restricted dataset (32 sequences analyzed), from which singleton polymorphisms were excluded.

3. Results and discussion

3.1. Polymorphism and natural selection in non-repetitive domains of *msp-2*

Table 2 gives estimates of nucleotide polymorphism in non-repetitive domains of *msp-2* (blocks 1, 2, 4, and 5 in Fig. 1) obtained with the dataset from which singleton substitutions were excluded. Corresponding results for the complete dataset are provided in the Supplementary Table

Table 2
Nucleotide polymorphism and positive selection in non-repetitive domains of the merozoite surface protein-2 gene of *Plasmodium falciparum*

Domain	Length	n	S	H	Hd (SD)	π (SD)	d_S (SE)	d_N (SE)	P
Block 1	129	38	1	2	0.149 (0.074)	1.16 (0.59)	0	1.52 (1.49)	0.162
Block 2 FC27	48	167	11	121	0.881 (0.014)	46.22 (1.25)	2.18 (1.96)	67.50 (22.63)	0.002
Block 2 3D7	27	239	16	46	0.950 (0.005)	119.42 (3.73)	4.04 (3.17)	199.05 (53.28)	<0.001
Block 4 FC27	141 ^a	77	3	4	0.149 (0.054)	1.97 (0.75)	0	2.73 (1.72)	0.058
Block 4 3D3	270 ^b	105	14	22	0.838 (0.024)	7.29 (0.58)	1.17 (0.81)	9.74 (3.43)	0.012
Block 5	225	33	4	4	0.636 (0.066)	5.15 (0.86)	4.42 (4.19)	5.58 (3.52)	0.396

Length of each domain given in base pairs (bp); n = number of sequences analyzed; S = number of segregating sites; H = number of haplotypes; Hd = haplotype diversity (and its standard deviation, SD); π = average number of nucleotide substitutions per 1000 sites between pairs of sequences (and SD); d_S = average number of synonymous nucleotide substitutions per 1000 synonymous sites (and its standard error, SE), d_N = average number of nonsynonymous substitutions per 1000 nonsynonymous sites (and SE). P values are given for the comparison between d_N and d_S with one-tailed Z -tests (Nei and Kumar, 2000).

^a 123 bp analyzed, 18-bp region with frameshift excluded.

^b 237 bp analyzed; indels excluded.

1. No attempt was made to align sequences of blocks 2 and 4 between dimorphic families; therefore, π , d_S , and d_N estimates given here for these two domains refer to comparisons within dimorphic groups, either FC27 or 3D7. The 27-bp long block 2 of 3D7-type alleles was found to be the most polymorphic domain ($\pi = 0.11942$, $SD = 0.00373$), with 46 different haplotypes.

Nonsynonymous nucleotide replacements (d_N) significantly exceeded synonymous replacements (d_S) in 3D7-type block 2, FC27-type block 2, and 3D7-type block 4 (Table 2). The d_N excess in FC27-type block 4 sequences was of borderline statistical significance when the singleton-free dataset was analyzed (Table 2), but a significant d_N excess was found in the complete dataset (Supplementary Table 1). These data suggest a role for naturally acquired immunity in maintaining sequence divergence in non-repetitive domains within *mSP-2* dimorphic groups. Antibody recognition of MSP-2 comprises a type-specific component, which discriminates between dimorphic types, and a variant-specific component, which discriminates among variants within each dimorphic group (Franks et al., 2003; Felger et al., 2003b; Kanunfre et al., 2003; Tonon et al., 2004). Therefore, both variant-specific and type-specific immunity may be major selective forces driving the evolution of *mSP-2*, with clear implications for designing adequate MSP-2-based immunization strategies (Felger et al., 2003a). As expected, no significant excess of nonsynonymous nucleotide replacements was found in blocks 1 and 5 sequences (Table 2), which encode conserved domains poorly recognized by naturally acquired (Taylor et al., 1995; Metzger et al., 2003) and vaccine-elicited (Flück et al., 2004) antibodies.

Additional polymorphism in FC27-type sequences was generated by two indels resulting in a frameshift that involved six codons at the 5' end of block 4 (positions 324–329 in Supplementary Fig. 1). Of 202 sequences analyzed, 38 (18.8%) had the variant form AAG TTC TGG CAA TCG ACA (encoding KFWQCT), instead of the wild-type sequence AGT TCT GGC AAT GCA CCA (encoding SSGNAP). These sequences differ by the insertion of one base (underlined in the variant sequence), that changed the reading frame, and the deletion of one base (underlined in the wild-type sequence), reestablishing the reading frame. This frameshift was found in isolates of all continents but Africa, being particularly common (46.9%) in Brazil (Tonon et al., 2004).

The 11-amino acid motif PKGKGVEVQKPN, at the 5' end of 3D7-type block 4 sequences (positions 343–353 in Supplementary Fig. 1), was deleted in 15 (9.0%) of 166 alleles analyzed, originating from all continents.

3.2. Sequence diversity and evolution of FC27-type repeats

The 209 FC27-type or hybrid (3D7/FC27-type) block 3 sequences analyzed (positions 60–320 in Supplementary Fig. 1) comprised 1–4 copies of the 96-bp repeat unit (R1), followed by a unique, conserved 21-bp motif (NR1;

positions 254–260, Supplementary Fig. 1) and 0–5 copies of the 36-bp repeat unit (R2). Possibly due to size constraints, the numbers of R1-type and R2-type repeat units in each allele were negatively correlated ($r = -0.757$, $P < 0.01$; Supplementary Table 2). A single polymorphic site (nonsynonymous nucleotide replacement) was found in the 207 NR1 sequences analyzed ($\pi = 0.00137$, $SD = 0.00077$).

Alignment of 284 R1-type repeat units revealed 15 distinct repeat haplotypes (Supplementary Fig. 2; Table 3). Eleven of 96 sites (11.5%) were polymorphic, and 8 polymorphic sites were located in the 5' half of the R1 repeat unit, the only R1 segment with unequivocal homology with the orthologous sequence of *P. reichenowi* (Rich and Ayala, 2000). Interestingly, the first copy of R1-type motifs is less polymorphic than block 2 sequences of FC27-type alleles ($\pi = 0.01973$, $SD = 0.00090$ vs. $\pi = 0.04622$, $SD = 0.00125$). This difference suggests that block 2 epitopes are the major target of diversifying selection due to variant-specific immunity, among FC27-type variants, arguing for the inclusion of several block 2 variants in future MSP-2-based multivalent vaccine prototypes.

R1 units were duplicated in 33.0% FC27-type alleles (Supplementary Table 2). The first copy of R1 showed extensive polymorphism, but copies in the subsequent positions in the array, when present, were clearly less variable when pairwise comparisons between alleles were made (Table 3). In fact, only a small part of the repertoire of R1-type repeat haplotypes was duplicated. Alleles with two or more R1 copies displayed either haplotypes #1, 2 or 3 in the first position of the array (Supplementary Table 3), while haplotypes #4–15 were only found in alleles with a single copy of R1. Therefore, 53 of the 96 alleles (54.5%) with haplotypes #1, 2 or 3 in the first position of the R1-type repeat array, but none of the 80 alleles with other haplotypes in this position, had two or more R1 copies.

The first and second copies of R1-type repeat units differed by either one or four nucleotides in 41.5% of the repeat arrays (Supplementary Fig. 2 and Supplementary Table 3). However, the other adjacent repeat units in the array (second, third, and fourth copy), when present, were always identical to each other in the same FC27-type allele. These data suggest an ancient origin for the first R1 duplication event, with a more recent acquisition of additional (third and fourth) R1 copies (Fig. 2). Since increasing numbers of R1-type repeat copies appear to increase the natural immunogenicity of MSP-2 variants (Franks et al., 2003), we can hypothesize that new units added by gene conversion or unequal crossing-over are likely to be subsequently removed under immune pressure, shortening their lifespan in the array.

R2 repeats were absent in 32 (15.3%) FC27-type alleles (Supplementary Table 2), 27 of them from either Africa or Oceania, possibly impairing the recognition of the corresponding antigens by naturally acquired antibodies (Ranford-Cartwright et al., 1996; Felger et al., 2003b;

Table 3

Nucleotide polymorphism in FC27-type repeat units of *merozoite surface protein-2* alleles of *Plasmodium falciparum* in relation to their position in the repeat array

Repeat type	Length	Position in the array	<i>n</i>	<i>S</i>	<i>H</i>	Hd (SD)	π (SD)
R1	96 bp	1	176	11	15	0.862 (0.011)	19.73 (0.90)
		2	64	1	2	0.062 (0.041)	0.64 (0.43)
		3	37	1	2	0.105 (0.066)	1.09 (0.69)
		4	7	1	2	0.476 (0.171)	4.96 (1.78)
R2	36 bp	1	172	7	12	0.710 (0.027)	26.09 (1.68)
		2	96	3	4	0.636 (0.029)	21.30 (1.58)
		3	93	4	5	0.684 (0.022)	25.87 (1.78)
		4	39	3	4	0.675 (0.039)	30.06 (1.79)
		5	8	2	2	0.428 (0.169)	23.81 (9.37)

bp = base pairs; *n* = number of sequences analyzed; *S* = number of segregating sites; *H* = number of haplotypes; Hd = haplotype diversity (and its standard deviation, SD); π = average number of nucleotide substitutions per 1000 sites between pairs of sequences (and SD).



Fig. 2. Hypothetical model for the evolution of R1-type (96-bp) and R2-type (36-bp) repeats in FC27-type alleles of the *Plasmodium falciparum* *merozoite surface protein-2* gene. After the first (ancient) duplication event, sequences of both copies diverged as a result of mutations, followed by a more recent proliferation of the second copy. This model could explain the nucleotide mismatches in pairwise comparisons of sequences of the first and second repeat units in the array and the complete (or nearly so) identity among the sequences of second, third, and fourth repeat units.

Franks et al., 2003). Alignment of 408 R2-type repeat units revealed 13 distinct repeat haplotypes (Supplementary Fig. 3; Table 3); 7 of 32 (21.9%) sites were found to be polymorphic. No association was found between levels of nucleotide polymorphism in R2-type motifs and either their position in the array (Table 3) or the distance between pairs of repeat units in the array (data not shown). Most haplotypes (#4–13) were found exclusively in alleles with two or more copies of R2.

Pairwise comparisons of the first and second copies of R2 revealed one or two nucleotide mismatches in 71.0% of the arrays (Supplementary Table 4), but the adjacent copies (second, third, fourth, and fifth), when present, were identical to each other (in the same 3D7-type allele) in 97.6% comparisons. As above, these results could be explained by an ancient duplication event (generating the second R2 copy) followed by a more recent acquisition of additional copies (Fig. 2). Again, the hypothesis of an increased turnover rate of repeat units under immune-mediated selection may be suggested for the evolution of R2-type arrays. However, it remains unclear why most alleles with haplotype #1 in the first position of the R2 array had haplotype #2 in the second position and vice versa (Supplementary Table 4), suggesting that sequence divergence between the first and the second copies of R2-type motifs was not a random process.

As noted elsewhere (Irion et al., 1997), proliferation of PNA or PKA motifs at the 3' end of block 2 (positions 58–126, Supplementary Fig. 1) created an additional repeat

in nine FC27-type alleles (6 from Tanzania, 2 from Vietnam, and one from Vanuatu). The PNA motif has been deleted in a single hybrid 3D7/FC27 allele shared by isolates from Nigeria (AF148225) and Ghana (AF217008) (Supplementary Table 5).

BLASTN searches identified nine FC27-type alleles that were shared by two or more isolates from different countries (Supplementary Table 5). Identity among alleles was defined when no nucleotide mismatch was found across at least 300 bp of aligned sequence (corresponding to the total length of the shorter sequence in pairwise comparisons), comprising the most polymorphic domains (blocks 2 and 3) of the gene.

3.3. Sequence diversity and evolution of 3D7-type repeats

For this analysis, the 239 3D7-type block 3 sequences available (positions 51–235 in Supplementary Fig. 1) were further divided into two repetitive domains (R1 and R2) interspersed with two unique sequences (NR1 and NR2; positions 51–58 and 167–194, respectively, in Supplementary Fig. 1). R1 corresponds to the GSA-rich repeat units, while R2 corresponds to the poly-threonine stretch (Felger et al., 1997; Hoffmann et al., 2001; Tonon et al., 2004). R1-type repeats were absent in a single 3D7-type allele (AY378307), of unknown origin (Felger et al., 2003b). The most common and widespread R1 haplotypes were GASGSA (found in 34.9% of 3D7-type alleles), GGSGSA (19.1%), GAVAGSGA (14.8%), GGSA (12.9%), and

GASGNPPAGA (6.7%). The analysis of this large dataset further indicates that R1-type repeat arrays originated from duplications of the GGTGCT hexamer followed by mutations, fusions with nearby motifs, and further duplications and mutations (Fenton et al., 1991; Felger et al., 1997; Dubbeld et al., 1998; Rich and Ayala, 2000). Unequal crossing-over, gene conversion, and replication slippage are the putative mechanisms generating the extensive variation in R1 repeat arrays of 3D7-type alleles (Felger et al., 1997; Rich and Ayala, 2000). Nearly all R1-type haplotypes had even numbers of codons (2, 4, 6, 8 or 10), as expected from the proposed mode of evolution of these repeats (Felger et al., 1997); the only exceptions are 5-mer, truncated motifs sometimes found in a single copy in the repeat array. The two most common R1-type repeat motifs are, in fact, NR1 sequences that have proliferated (Hoffmann et al., 2001). Although the GGSA motif, present in a 3D7-based malaria vaccine prototype used in clinical trials (Felger et al., 2003a), has been found to be relatively common in Africa (13.8% of 3D7-typed alleles), Asia (15.8%), and Oceania (11.5%), it is substantially less prevalent in Brazil (5.0%). Accordingly, recombinant antigens containing GGSA-type repeats are poorly recognized by antibodies of malaria-exposed subjects in Brazil, while locally prevalent 3D7-type variants were readily recognized (Kanunfre et al., 2003; Tonon et al., 2004). Therefore, it remains uncertain whether vaccines containing GGSA-type repeats may induce significant protection in areas where this MSP-2 variant is infrequent.

R2 repeat arrays in the 234 3D7-type alleles analyzed differed in the number of copies (range: none to five) of the nonamer ACTACCACA, followed by a single copy of ACTACT, with a total of 2–17 threonine codons. Only three alleles had imperfect copies (one each) of the nonamer. A single lysine codon (AAA) was found in the R2 array of 16 alleles (13 of them from Africa). Interestingly, all of these had 13 threonine codons arranged in the same way and displayed an identical nucleotide replacement (ACA → AAA) in the second copy of the ACTACCACA nonamer. A few other mutations found in R2 arrays were singletons, which may represent sequencing errors.

An additional repeat was created, in 18 3D7-type alleles (14 of them from Africa), by the proliferation of a TPA motif, present in 2–9 copies at the 3' end of NR2 (positions 192–218 in [Supplementary Fig. 1](#)). This same motif was deleted in two alleles (AY375170, of unknown origin, and U91663, from Gambia). In addition, the PT motif at the 3' end of block 2 (positions 51–52 in [Supplementary Fig. 1](#)) was duplicated in 16 alleles (originating from all continents except Asia).

Despite the huge potential for sequence variation in 3D7-type alleles, BLASTN searches identified five alleles of this dimorphic type that were shared by 2 or 3 isolates from different countries ([Supplementary Table 5](#)). Sequence identity criteria were as described for FC27-type alleles (see Section 3.2).

3.4. Intragenic recombination between *msp-2* alleles of different dimorphic types

Hybrid variants with both FC27-type and 3D7-type sequences appear to be rare. Only 14 (3.1%) of 448 *msp-2* alleles analyzed had dimorphic sequence motifs derived from both types ([Table 1](#)). All hybrid alleles (11 of them from Africa; none from Asia or Oceania) had a 3D7-type block 2 and a FC27-type block 4; no example of FC27/3D7 hybrid was found so far. The same hybrid allele was shared by two isolates from different countries in West Africa (Nigeria [AF148224] and Gambia [U91675]) ([Supplementary Table 5](#)). A single cross-over site within block 3 was identified for all 3D7/FC27 alleles (positions 59–133 in [Supplementary Fig. 1](#), sequences AY375166 through AF148222), as described elsewhere (Irion et al., 1997; Hoffmann et al., 2001). Nevertheless, given the low proportion of hybrid alleles so far described, this region can hardly be defined as a recombination hotspot.

If type-specific immunity represents a major selective force, hybrid variants are expected to be rare, since they would be readily recognized by subjects previously exposed to either FC27 or 3D7-type variants. The hypothesis of immune-mediated negative selection of hybrid variants, however, is at odds with the recent observation that, in Africa, the sequential exposure to different versions of MSP-2 appears to elicit immune responses that cross-recognize the first version of this antigen to which the individual was exposed, limiting the repertoire of antigenic specificities that are subsequently recognized (Franks et al., 2003). More data are required to evaluate the relative importance of this phenomenon (known as clonal imprinting or original antigenic sin) in the immune recognition of MSP-2 by other malaria-exposed populations, especially in areas of lower endemicity. Alternatively, the sequence divergence between FC27-type and 3D7-type alleles in the repetitive domains might inhibit chiasma formation and subsequent meiotic recombination between allelic families. However, the analysis of the putative recombination site reveals a region of clear sequence similarity between 3D7-type R1 repeats and the 5' end of the FC27-type repetitive motif R1 (Irion et al., 1997; Hoffmann et al., 2001; [Supplementary Fig. 1](#)).

3.5. Origin and evolution of dimorphic *msp-2* families

In 1993, Louis Miller and colleagues concluded their analysis of coding sequences of MSP-1, another major dimorphic antigen of *P. falciparum*, by saying that “some speculation on the origins of the dimorphic forms is warranted in that no one has commented on the origin of two such divergent alleles in the absence of intermediates, that is, missing links between the two” (Miller et al., 1993). We now have evidence that the dimorphic *msp-1* types of *P. falciparum* originated about 27 Myr ago, before the split between *P. falciparum* and *P. reichenowi* (Polley et al., 2005); *msp-1* alleles of *P. vivax* are probably nearly

as ancient (Putaporntip et al., 2006). Alleles of another dimorphic antigen-coding gene of *P. falciparum*, *msp-3*, may be even older (Polley et al., 2005). However, little is currently known about the origin of allelic dimorphism itself; comparative analyses of *msp-2* sequences may therefore provide further insights into the origins and maintenance of allelic dimorphism in malarial surface antigens.

BLASTN homology searches with two dimorphic sequences of *P. falciparum* (isolates FC27 [J03828] and 3D7 [AE001385]) and the available *msp-2* sequence of *P. reichenowi* (Y14731) failed to identify orthologous *msp-2* sequences in other malaria parasites. We have examined the published genome of the rodent malaria parasite *P. yoelii* (Carlton et al., 2002) and partial genome sequence data of other rodent (*P. berghei* and *P. chabaudi*), primate (*P. knowlesi*), bird (*P. gallinaceum*) (all available at <http://www.sanger.ac.uk/Projects/Protozoa/>) and human malaria parasites (*P. vivax*, available at http://www.tigr.org/msc/p_vivax/). Therefore, the *msp-2* gene of the chimpanzee malaria parasite *P. reichenowi* (Dubbed et al., 1998) is the only known ortholog of *P. falciparum msp-2* from which evolutionary inferences may be obtained.

Rich and Ayala (2000) identified, in *P. reichenowi msp-2*, three domains (defined as repeat homology regions or RHRs) that are homologous to repeat motifs found in dimorphic types of *P. falciparum msp-2*. They proposed that, in alleles in which the FC27-type RHRs present in the common ancestral of *P. reichenowi* and *P. falciparum* proliferated, 3D7-type precursors were lost due to size constraints on the molecule, and vice versa. The proliferation and deletion of different RHRs could generate extensive sequence divergence between dimorphic types of *P. falciparum msp-2* over a relatively short period of time (5000–50,000 years) (Rich and Ayala, 2000). This model implies a very recent origin of allelic dimorphism at the *msp-2* locus, contrasting with the extremely ancient origin of dimorphic lineages in the *msp-1* locus of both *P. falciparum* (Polley et al., 2005) and *P. vivax* (Putaporntip et al., 2006). The extraordinary plasticity of *msp-2* sequences documented in the present analysis is in accordance with this elegant evolutionary model, but the hypothesis of a recent common ancestry of all *msp-2* alleles in current populations of *P. falciparum* remains to be explicitly tested.

We derived approximate TMRCA estimates for extant *P. falciparum msp-2* alleles by using the d_S value (0.00387, SE = 0.00380) calculated for 32 sequences (singletons excluded) of blocks 1 and 5, the only *msp-2* domains that can be unambiguously aligned between allelic types. Point estimates for TMRCA were 456,537 years, under the assumption that *P. falciparum* and *P. reichenowi* diverged 5 Myr ago ($\mu_S = 4.24 \times 10^{-9}$ synonymous substitutions per site per year), and 639,144 years, assuming that these species diverged 7 Myr ago ($\mu_S = 3.03 \times 10^{-9}$ synonymous substitutions per site per year). The upper 95% confidence limits of these estimates are 1.33 Myr and 1.87 Myr, respectively; because of the large standard errors of our d_S estimates, the lower 95% confidence limit falls in the

5000–50,000-year range. The inclusion of four sequences with singletons in the analysis inflated TMRCA estimates by 31–33% (data not shown). A similar analysis of non-repetitive dimorphic domains (blocks 2 and 4), in the singleton-free dataset, resulted in TMRCA estimates for all FC27-type and 3D7-type alleles of 120,237 years (upper 95% confidence limit, 0.31 Myr) and 243,093 years (upper 95% confidence limit, 0.45 Myr), respectively, under the assumption that *P. falciparum* and *P. reichenowi* diverged 5 Myr ago, and of 168,330 years (upper 95% confidence limit, 0.44 Myr) and 340,326 years (upper 95% confidence limit, 0.64 Myr), respectively, assuming that *P. falciparum* and *P. reichenowi* diverged 7 Myr ago.

In addition to the large stochastic error associated with the relatively small number of codons compared, our TMRCA estimates may have been affected by: (a) extraordinarily high levels of spontaneous mutation in silent sites of *msp-2*, (b) sequencing errors in the DNA sequences (Barry et al., 2003), and (c) selection on silent variation at this locus. Below we consider each of these three hypotheses.

To avoid the first drawback (hypothesis a), we derived *msp-2*-specific rates of synonymous substitutions per site per year (μ_S), which in fact are 17–36% lower than those available for other nuclear genes of *P. falciparum* (Hughes and Verra, 2001; Tanabe et al., 2004). Exclusion of sequences with singleton polymorphism from the dataset under analysis surely eliminated most sequencing errors (as well as some true rare polymorphisms), suggesting that sequence artifacts (hypothesis b) are unlikely to have inflated our d_S estimates; of course, however, some underestimation is possible. It has been suggested that synonymous nucleotide replacements in *P. falciparum* might be under purifying selection, leading to d_S estimates that are somewhat lower than those expected under neutrality (hypothesis c) (Forsdyke, 2002). No solid evidence has so far been published to support this hypothesis, and comparisons of d_N/d_S ratios for malarial genes or domains putatively under purifying selection, diversifying selection or neutrally evolving (Hughes and Hughes, 1995; Escalante et al., 1998b; Hughes and Verra, 2001; Tanabe et al., 2004; Table 2) do not suggest any bias due to d_S underestimation. If hypothesis c is correct, purifying selection on synonymous sites would affect similarly published TMRCA estimates for several malarial genes (Rich et al., 1998; Hughes and Verra, 2001; Tanabe et al., 2004; Putaporntip et al., 2006), including *msp-2* (this study), all of them based on comparisons of synonymous substitution rates. As a consequence, all TMRCA estimates would be still comparable to each other.

In conclusion, we found no support for an extremely ancient age of *msp-2* alleles. Our findings suggest different time scales for the evolution of allelic dimorphism in two major malarial antigens, MSP-1 and MSP-2. Additional high-quality sequences are required for the fine tuning of TMRCA estimates for *msp-2* alleles and further hypothesis formulation and testing.

Acknowledgment

M.U.F. is a visiting scholar supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (200213/2005-9).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.exppara.2006.05.003](https://doi.org/10.1016/j.exppara.2006.05.003).

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Barry, A.E., Leliwa, A., Choi, M., Nielsen, K.M., Hartl, D.L., Day, K.P., 2003. DNA sequence artifacts and the estimation of time to the most recent common ancestor (TMRCA) of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 31, 143–147.
- Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., Peterson, J.D., Pop, M., Kosack, D.S., Shumway, M.F., Bidwell, S.L., Shallom, S.J., van Aken, S.E., Riedmuller, S.B., Feldblyum, T.V., Cho, J.K., Quackenbush, J., Sedegah, M., Shoabi, A., Cummings, L.M., Florens, L., Yates, J.R., Raine, J.D., Sinden, R.E., Harris, M.A., Cunningham, D.A., Preiser, P.R., Bergman, L.W., Vaidya, A.B., van Lin, L.H., Janse, C.J., Waters, A.P., Smith, H.O., White, O.R., Salzberg, S.L., Venter, J.C., Fraser, C.M., Hoffman, S.L., Gardner, M.J., Carucci, D.J., 2002. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419, 512–519.
- Dubbeld, M.A., Kocken, C.H., Thomas, A.W., 1998. Merozoite surface protein 2 of *Plasmodium reichenowi* is a unique mosaic of *Plasmodium falciparum* allelic forms and species-specific elements. *Molecular and Biochemical Parasitology* 92, 187–192.
- Elliott, J.F., Albrecht, G.R., Gilladoga, A., Handunnetti, S.M., Neequaye, J., Lallinger, G., Minjas, J.N., Howard, R.J., 1990. Genes for *Plasmodium falciparum* surface antigens cloned by expression in COS cells. *Proceedings of the National Academy of Sciences of United States of America* 87, 6363–6367.
- Escalante, A.A., Freeland, D.E., Collins, W.E., Lal, A.A., 1998a. The evolution of primate malaria parasites based on the gene encoding cytochrome *b* from the linear mitochondrial genome. *Proceedings of the National Academy of Sciences of United States of America* 95, 8124–8129.
- Escalante, A.A., Lal, A.A., Ayala, F.J., 1998b. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 149, 189–202.
- Felger, I., Marshal, V.M., Reeder, J.C., Hunt, J.A., Mgone, C.S., Beck, H.-P., 1997. Sequence diversity and molecular evolution of the merozoite surface antigen 2 of *Plasmodium falciparum*. *Journal of Molecular Evolution* 45, 154–160.
- Felger, I., Genton, B., Smith, T., Tanner, M., Beck, H.-P., 2003a. Molecular monitoring in malaria vaccine trials. *Trends in Parasitology* 19, 60–63.
- Felger, I., Steiger, S., Hatz, C., Smith, T., Beck, H.-P., 2003b. Antigenic cross-reactivity between different alleles of the *Plasmodium falciparum* merozoite surface protein 2. *Parasite Immunology* 25, 531–543.
- Fenton, B., Clark, J.T., Khan, C.M., Robinson, J.V., Walliker, D., Ridley, R., Scaife, J.G., McBride, J.S., 1991. Structural and antigenic polymorphism of the 35- to 48-kilodalton merozoite surface antigen (MSA-2) of the malaria parasite *Plasmodium falciparum*. *Molecular and Cell Biology* 11, 963–971.
- Flück, C., Smith, T., Beck, H.-P., Irion, A., Betuela, I., Alpers, M.P., Anders, R., Saul, A., Genton, B., Felger, I., 2004. Strain-specific humoral response to a polymorphic malaria vaccine. *Infection and Immunity* 72, 6300–6305.
- Forsdyke, D.R., 2002. Selective pressures that decrease synonymous mutations in *Plasmodium falciparum*. *Trends in Parasitology* 18, 411–417.
- Franks, S., Baton, L., Tetteh, K., Tongren, E., Dewin, D., Akanmori, B.D., Koram, K.A., Ranford-Cartwright, L., Riley, E.M., 2003. Genetic diversity and antigenic polymorphism in *Plasmodium falciparum*: extensive serological cross-reactivity between allelic variants of merozoite surface protein 2. *Infection and Immunity* 71, 3485–3495.
- Franks, S., Koram, K.A., Wagner, G.E., Tetteh, K., McGuinness, D., Wheeler, J.G., Nkrumah, F., Ranford-Cartwright, L., Riley, E.M., 2001. Frequent and persistent, asymptomatic *Plasmodium falciparum* infections in African infants, characterized by multilocus genotyping. *Journal of Infectious Diseases* 183, 796–804.
- Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C., Hoffman, S.L., 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* 282, 1126–1132.
- Hoffmann, E.H.E., da Silveira, L.A., Tonhosolo, R., Pereira, F.J.T., Ribeiro, W.L., Tonon, A.P., Kawamoto, F., Ferreira, M.U., 2001. Geographical patterns of allelic diversity in the *Plasmodium falciparum* malaria-vaccine candidate, merozoite surface protein-2. *Annals of Tropical Medicine and Parasitology* 95, 117–132.
- Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at MHC class I loci reveals overdominant selection. *Nature* 335, 167–170.
- Hughes, A.L., Verra, F., 2001. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proceedings of the Royal Society of London B* 268, 1855–1860.
- Hughes, M.K., Hughes, A.L., 1995. Natural selection on *Plasmodium* surface proteins. *Molecular and Biochemical Parasitology* 71, 99–113.
- Irion, A., Beck, H.-P., Felger, I., 1997. New repeat unit and hot spot of recombination in FC27-type alleles of the gene coding for *Plasmodium falciparum* merozoite surface protein 2. *Molecular and Biochemical Parasitology* 90, 367–370.
- Jensen, J.B., Trager, W., 1978. *Plasmodium falciparum* in culture: establishment of additional strains. *American Journal of Tropical Medicine and Hygiene* 27, 743–746.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, R.E. (Ed.), *Mammalian Protein Metabolism*. Academic Press, New York, pp. 21–132.
- Kanunfre, K.A., Leoratti, F.M., Hoffmann, E.H.E., Durlacher, R.R., Ferreira, A.W., Moraes-Avila, S.L., Ferreira, M.U., 2003. Differential recognition of *Plasmodium falciparum* merozoite surface protein 2 variants by antibodies from malaria patients in Brazil. *Clinical and Diagnostic Laboratory Immunology* 10, 973–976.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5, 150–163.
- Metzger, W.G., Okenu, D.M., Cavanagh, D.R., Robinson, J.V., Bojang, K.A., Weiss, H.A., McBride, J.S., Greenwood, B.M., Conway, D.J., 2003. Serum IgG3 to the *Plasmodium falciparum* merozoite surface protein 2 is strongly associated with a reduced prospective risk of malaria. *Parasite Immunology* 25, 307–312.
- 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). *Molecular and Biochemical Parasitology* 59, 1–14.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3, 418–426.
- Nei, M., Kumar, S., 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, p. 55.

- Polley, S.D., Conway, D.J., Cavanagh, D.R., McBride, J.S., Lowe, B.S., Williams, T.N., Mwangi, T.W., Marsh, K., 2006. High levels of serum antibodies to merozoite surface protein 2 of *Plasmodium falciparum* are associated with reduced risk of clinical malaria in coastal Kenya. *Vaccine* 24, 4233–4246.
- Polley, S.D., Weedall, G.D., Thomas, A.W., Golightly, L.M., Conway, D.J., 2005. Orthologous gene sequences of merozoite surface protein 1 (MSP1) from *Plasmodium reichenowi* and *P. gallinaceum* confirm an ancient divergence of *P. falciparum* alleles. *Molecular and Biochemical Parasitology* 142, 25–31.
- Prescott, N., Stowers, A.W., Cheng, Q., Bobogare, A., Rzepczyk, C.M., Saul, A., 1994. *Plasmodium falciparum* genetic diversity can be characterised using the polymorphic merozoite surface antigen 2 (MSA-2) gene as a single locus marker. *Molecular and Biochemical Parasitology* 63, 203–212.
- Putaporntip, C., Jongwutiwes, S., Iwasaki, T., Kanbara, H., Hughes, A.L., 2006. Ancient common ancestry of the merozoite surface protein 1 of *Plasmodium vivax* as inferred from its homologue in *Plasmodium knowlesi*. *Molecular and Biochemical Parasitology* 146, 105–108.
- Ranford-Cartwright, L.C., Taylor, R.R., Asgari-Jirhandeh, N., Smith, D.B., Roberts, P.E., Robinson, V.I., Babiker, H.A., Riley, E.M., Walliker, D., McBride, J.S., 1996. Differential antibody recognition of FC27-like *Plasmodium falciparum* merozoite surface protein MSP2 antigens which lack 12 amino acid repeats. *Parasite Immunology* 18, 411–420.
- Rich, S.M., Ayala, F.J., 2000. Population structure and recent evolution of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of United States of America* 97, 6994–7001.
- Rich, S.M., Licht, M.C., Hudson, R.R., Ayala, F.J., 1998. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of United States of America* 95, 4425–4430.
- Rozas, J., Sanchez-DelBarrio, J.C., Messenguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other approaches. *Bioinformatics* 19, 2496–2497.
- Sherry, S.T., Rogers, A.R., Harpending, H., Soodyall, H., Jenkins, T., Stoneking, M., 1994. Mismatch distributions of human mtDNA reveal recent human population expansions. *Human Biology* 66, 761–775.
- Smythe, J.A., Peterson, M.G., Coppel, R.L., Saul, A.J., Kemp, D.J., Anders, R.F., 1990. Structural diversity in the 45-kilodalton merozoite surface antigen of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 39, 227–234.
- Snewin, V.A., Herrera, M., Sanchez, G., Scherf, A., Langsley, G., Herrera, S., 1991. Polymorphism of the alleles of the merozoite surface antigens MSA1 and MSA2 in *Plasmodium falciparum* wild isolates from Colombia. *Molecular and Biochemical Parasitology* 49, 265–275.
- Tanabe, K., Sakihama, N., Hattori, T., Ranford-Cartwright, L., Goldman, I., Escalante, A.A., Lal, A.A., 2004. Genetic distance in housekeeping genes between *Plasmodium falciparum* and *Plasmodium reichenowi* and within *P. falciparum*. *Journal of Molecular Evolution* 59, 687–694.
- Taylor, R.R., Smith, D.B., Robinson, V.J., McBride, J.S., Riley, E.M., 1995. Human antibody response to *Plasmodium falciparum* merozoite surface protein 2 is serogroup specific and predominantly of the immunoglobulin G3 subclass. *Infection and Immunity* 63, 4382–4388.
- Thomas, A.W., Carr, D.A., Carter, J.M., Lyon, J.A., 1990. Sequence comparison of allelic forms of the *Plasmodium falciparum* merozoite surface antigen MSA2. *Molecular and Biochemical Parasitology* 43, 211–220.
- 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* 24, 4876–4882.
- Tonon, A.P., Hoffmann, E.H.E., Silveira, L.A., Ribeiro, A.G., Goncalves, C.R.S., Ribolla, P.E.M., Wunderlich, G., Ferreira, M.U., 2004. *Plasmodium falciparum*: sequence diversity and antibody recognition of the merozoite surface protein-2 (MSP-2) in Brazilian Amazonia. *Experimental Parasitology* 108, 114–125.
- Topolska, A.E., Wang, L., Black, C.G., Coppel, R.L., 2004. Merozoite cell biology. In: Waters, A.P., Jense, C.J. (Eds.), *Malaria Parasites: Genomes and Molecular Biology*. Caister Academic Press, Wymondham, pp. 365–444.
- Zheng, C., Xie, P., Chen, Y., 2003. Molecular cloning and sequencing of the merozoite surface antigen 2 gene from *Plasmodium falciparum* strain FCC-1/HN and expression of the gene in mycobacteria. *Journal of Eukaryotic Microbiology* 50, 140–143.