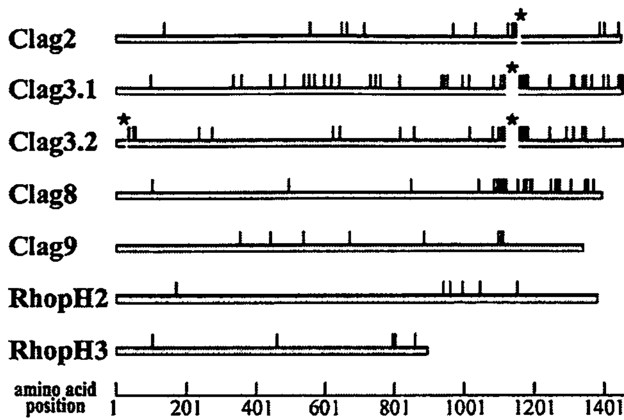

Fig. 1. Locations of amino acid polymorphism of seven components of the *Pf*RhopH complex among five *P. falciparum* parasite lines (3D7, HB3, Dd2, FVO, and 7G8). Indels are shown as gaps with asterisks (aa position 1145 for Clag2, aa positions 1113–1149 for Clag3.1, and aa positions 30–33 and 1115–1148 for Clag3.2). Numbers are those of 3D7 line sequences.

(Fig. 1). In addition, numerous polymorphic sites in this region have more than one amino acid substitution, whereas most polymorphisms in the other regions are dimorphic (at both nucleotide and amino acid levels). Most indels are also located in this region (Fig. 1, asterisks). Thus, the region at aa positions 1000–1200 of RhopH1/Clag is the most highly polymorphic region of the *Pf*RhopH complex.

3.2. Gene conversion between clag3.1 and 3.2

Of interest, *clag3.1* and *3.2* share some polymorphic sites. Because *clag3.1* and *3.2* have 96.7% nucleotide identity (3D7 parasite line) and are located on chr 3 and separated by only 10 kb harboring one putative ORF (PFC0115c) (Fig. 2A), we assessed gene conversion between these two loci. Using an algorithm by Betrán et al. [34], we identified multiple gene conversion tracts located at nt positions 1314–1353, 1447–1452, 1612–1659, 1702–1785, 1852–1983, and 2148–2208 in 3D7 *clag3.1*; nt positions 3824–4240 in HB3 *clag3.1*; nt positions 189–247 in 7G8 *clag3.1*; nt positions 813–817 and 3821–4182 in 3D7 *clag3.2*; nt positions 88–151 in HB3 *clag3.2*; and nt positions 3320–3755 in 7G8 *clag3.2* (Fig. 3). The detected conversion tracts had less than 5% informative nucleotides showing a mosaic origin, indicating that the probability of these tracts being involved in a recombination event more than once is negligible [34]. No gene conversion was detected between the other *rhoph1/clag* genes.

Because gene conversion potentially accelerates nucleotide diversity, we evaluated the evolutionary rates of *clag3.1* and *3.2*. Results showed that *clag2*, *3.1*, and *3.2* form a single clade and *clag8* another (Fig. 4); thus we performed Tajima's relative rate test using *clag8* as an outgroup and found that the evolutionary rates between *clag3.1* and 2 and between *clag3.2* and 2 were significantly different for all combinations of the sequences from five parasite lines. Because *clag3.1* and *3.2* were more diverse than *clag2*, *clag3.1* and *3.2* appear to have evolved more rapidly than *clag2*.

3.3. Amino acid polymorphism of the region around aa positions 1000–1200 of Clag2, 3.1, 3.2, and 8

Because extensive polymorphisms were observed around aa positions 1000–1200 in Clag2, 3.1, 3.2, and 8, we further analyzed polymorphism in this region with additional sequences from parasite lines originating worldwide. Alignment of Clag2 sequences showed multiple amino acid substitutions per site at multiple sites, e.g., five amino acids at aa position 1139 (K, R, S, G, and I). Indels were also observed (Supplemental Fig. S3). Clag8 has even higher levels of amino acid substitutions at between 1077 and 1136; five different amino acids (I, S, R, G, and N) at 1100, seven at 1101 (D, S, T, E, N, I, and K), six at 1104 (S, N, I, K, R, and T), and five at 1105 (G, D, T, S, and N) (Supplemental Fig. S4). Clag3.1 and 3.2 are also highly polymorphic (Fig. 5), which will be discussed later.

3.4. Copy number polymorphism of rhoph1/clag genes on chr 3

Notably, when PCR amplification was performed to obtain DNA fragments of the entire ORFs of *clag3.1* or *3.2*, 17 parasite

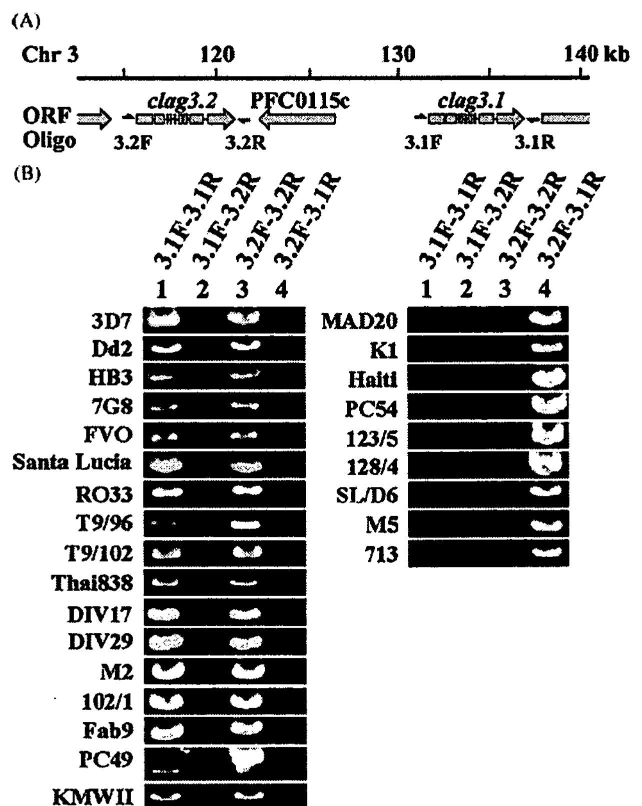

Fig. 2. Copy number polymorphism of *rhoph1/clag* genes on chr 3. (A) Genome organization around *clag3.2* and *3.1* gene loci on chr 3. The locations of the oligonucleotide primers are indicated. Oligonucleotide 3.2F and 3.1F were designed on the 5′ UTR of *clag3.2* and *3.1*, respectively. Oligonucleotide 3.2R and 3.1R were designed on 3′ UTR of *clag3.2* and *3.1*, respectively. (B) PCR-amplified DNA fragments of 26 *P. falciparum* lines with different combinations of oligonucleotides.

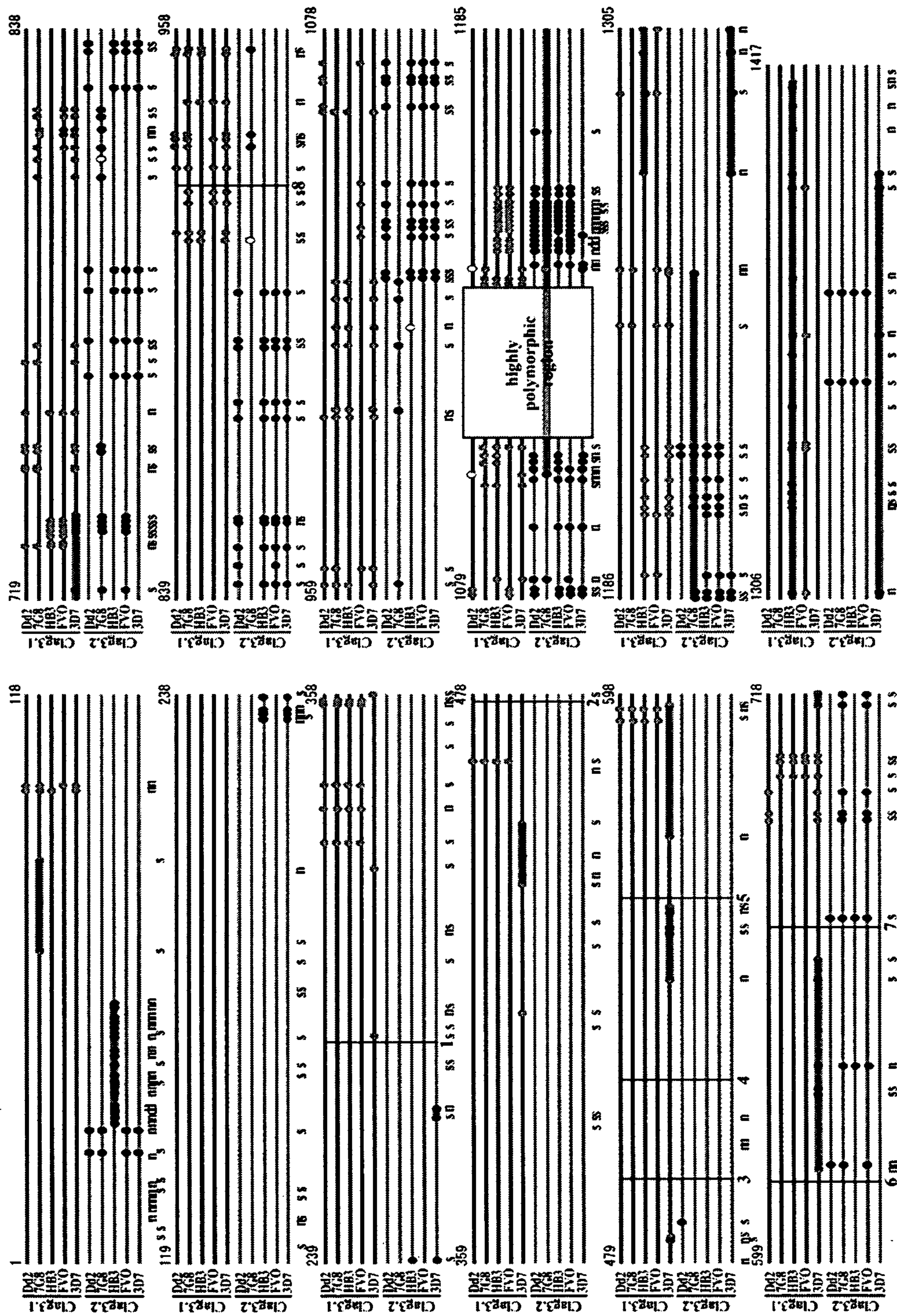

Fig. 3. Gene conversion tracts in *clag3.1* and *clag3.2*. Polymorphic codons (circles) in the coding sequences of Clag3.1 and 3.2 were compared in five *P. falciparum* lines. Clag3.1, black bar; Clag3.2, gray bar. Polymorphisms matching the paralogous sequence are shown in gray or black circles, respectively, and rare polymorphisms by an open circle. Exons are separated by vertical bar with the intron number at the bottom. Polymorphic sites that differ between consensus sequences are shown below the line classified as nonsynonymous (n), synonymous (s), and deletion (d). Gene conversion tracts identified using algorithm by Betrán et al. [34], wide gray bars.

lines showed the 2 expected positive bands with the primer sets 3.1F–3.1R and 3.2F–3.2R, whereas 9 parasite lines showed a positive band only with the primer set 3.2F–3.1R, which suggests that these 9 parasite lines possessed a hybrid gene with *clag3.2* sequence at the 5′ UTR and *clag3.1* sequence at the 3′ UTR (Fig. 2B). DNA fragments were not amplified with other primer combinations, indicating that artificial amplification due to primer mispairing was negligible. This is consistent with a

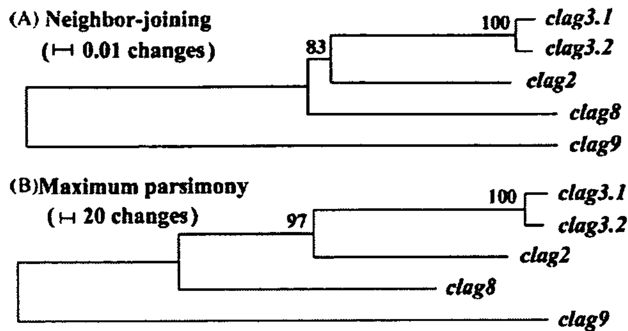

Fig. 4. Unrooted dendrograms of *pfrhophl/clag* genes using nucleotide sequences from the 3D7 parasite line. The trees were constructed by the neighbor-joining and maximum parsimony methods using MEGA 3.1. Numbers on branches indicate bootstrap values (500 pseudoreplicates).

recent report by Chung et al. [36], who found that some parasite lines possess only a single *rhophl/clag* on chr 3 by Southern blot hybridization. We here designate this *clag3* gene as *clag3h* (*clag3 hybrid*; Clag3H for protein). In addition, we obtained two distinct sequences for *clag3.1* from the KMWII parasite line (Fig. 5), using several cloned plasmids after experiencing difficulty in direct sequencing of PCR products. Sequences for *clag2*, *3.2*, and *8* were easily obtained from the KMWII line by direct sequencing of the PCR products, supporting the assumption that this line was a clone. Thus, the KMWII line appears to possess at least three *clag3*-related sequences in the genome. This data suggests that the number of *clag3*-related sequences in *P. falciparum* varies from one to at least three.

To deduce the direction of the one-gene to two-gene (or vice versa) change, we searched *P. reichenowi* orthologs in the genome database and found one sequence read (reich908g11.plk) showing high similarity with the sequence around the start codon of *clag3.1* and *3.2*. We also found two reads (reich289f06.qlk and reich1194c08.plk) showing strong homology with the sequence around the stop codons of *clag3.1* and *3.2*. Comparison of the nucleotide sequences at the UTR revealed that reich908g11.plk and reich289f06.qlk were similar to the *pfclag3.2* sequence and that reich1194c08.plk was similar to the *pfclag3.1* sequence (Fig. 6). Thus, duplication of *clag3.1* and *3.2* gene loci appears to predate the divergence of *P. falciparum* and *P. reichenowi*, suggesting that a single *rhophl/clag* (*clag3h*) found in some *P. falciparum* lines is likely a result of an unequal crossover between two closely related genes. Notably, Clag3H had characteristic amino acids that were not observed in Clag3.1 and 3.2. For example, Ala at 1116 was found in three of nine Clag3H (30%). If Clag3H originated recently, for example during culture, the amino acid allele observed in Clag3H would also exist in Clag3.1 or 3.2; however, Ala at 1116 was not found in a total of 36 sequences of non-Clag3H protein sequences. Three in nine Clag3Hs is a significant excess compared to zero Ala at 1116 in 36 non-Clag3H sequences by Fisher's exact test ($P = 0.013$). This suggests that at least some Clag3H have accumulated some unique amino acid substitutions since their creation.

3.5. Selection on the PfRhopH complex

Positive selection was evaluated by comparing synonymous and nonsynonymous substitutions (Table 1). A significant excess of d_N over d_S was observed for *clag9* (entire ORF of five parasite lines) and for *clag8* (highly polymorphic region at nt positions 3022–3606 of 26 parasite lines), suggesting positive selection acting on these genes. A sliding window plot of d_N:d_S ratios revealed that *clag2* and *8* had the highest peaks, around nt positions 3000–3600 (Fig. 7). It should be noted that the corresponding regions of *clag3.1* and *3.2* are the regions showing highly extensive polymorphism with indels (asterisks in Fig. 7), thereby preventing evaluation of d_N:d_S ratios in this region. The peak at the N-terminus of *clag3.2* is due to introduction of part of the *clag3.1* sequence into the HB3 line *clag3.2* by gene conversion (see Fig. 3).

Positive selection was further evaluated by the McDonald–Kreitman test using *P. reichenowi* orthologs for *clag2* and *8*. Significant excess of intraspecific nonsynonymous substitutions over synonymous substitutions was observed in *clag8* as compared with interspecies fixed differences of nonsynonymous and synonymous changes, suggesting positive selection (Table 2).

3.6. Early origin of the clag2 and 8 polymorphism

We estimated the TMRCA for *clag2* and *8* polymorphism using aligned regions. Distances of synonymous single-nucleotide polymorphisms are 0.0139 ± 0.0031 for *clag2* and 0.0106 ± 0.0030 for *clag8*. Distances between *P. falciparum* and *P. reichenowi* are 0.0455 ± 0.0082 and 0.0748 ± 0.0120 for *clag2* and *8*, respectively. Assuming that the divergence time of *P. falciparum* and *P. reichenowi* was 6 million years ago (mya) [37,38], the estimated TMRCA of the polymorphism of *clag2*

Table 2
The McDonald–Kreitman test of selection for *Plasmodium falciparum clag2* and *8*

Locus	n[a]	No. of sites	Fixed differences between species		Polymorphic sites within *P. falciparum*		
			Syn	Nsyn[b]	Syn	Nsyn	P[c]
clag2	5	3273	23	47	18	12	(0.011)
clag8	5	2715	36	55	11	38	0.030

[a] *n*, Number of *P. falciparum* lines used.
[b] Syn, synonymous; Nsyn, nonsynonymous substitutions.
[c] Fisher's exact test (one-tailed) was used. *P*-value indicates that Nsyn are significantly greater than Syn. Value in parenthesis indicates that Syn are significantly greater than Nsyn.

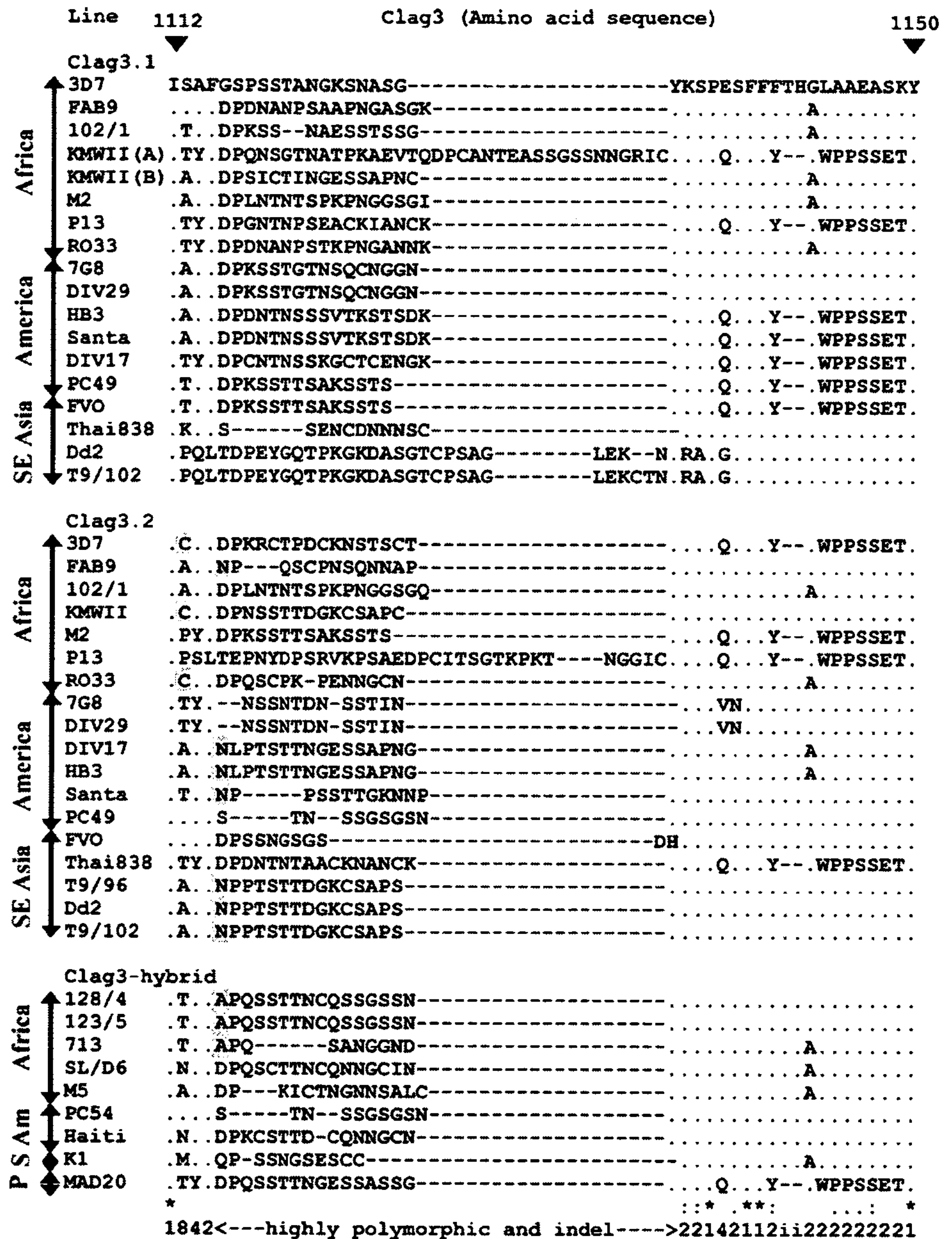

Fig. 5. Polymorphism of Clag3. An amino acid region 1112–1150 (after 3D7 line Clag3.1 sequence) of *P. falciparum* (27 lines) was aligned. Geographic origins are shown at left: SE Asia or S, Southeast Asia; Am, America; P, Papua New Guinea; Santa, Santa Lucia cloned line. Identical, conserved, or semiconserved residues in the alignment are indicated with asterisk, colon, or period, respectively. The number of amino acid replacements at each position and the region with indels are shown at the bottom. Cys residue at aa position 1113 and Asn and Ala residues at aa position 1116 are masked.

and *8* are 1.89 (95% CI, 1.02–3.18) and 0.87 (95% CI, 0.42–1.54) mya, respectively.

4. Discussion

4.1. Diversifying selection on the rhophl/clag gene loci

The present study revealed that the RhopH1/Clag-encoding genes *clag2*, *3.1*, *3.2*, and *8* contain a highly polymorphic region, particularly at nt positions 3000–3600. Diversifying selection increases nucleotide diversity (π), and an excess of d_N to d_S is indicative of positive selection favoring amino acid replacement [39]. Thus, the observed excess of d_N to d_S at nt positions 3000–3600 of *clag8* suggests that the polymorphism in *clag8* is positively maintained. An excess of d_N to d_S was also observed for *clag9*, indicating that this gene is also under positive selection. The most polymorphic region, in which positive selection was detected for *clag8*, was excluded from *clag2*, *3.1*, and *3.2* due to extensive sequence variation that made sequence alignment unreliable. Further analysis is required to evaluate positive selection on these three *rhoph1/clag* genes.

```
5' UTR
PfClag3.1-5'UTR    caaagtgtacatgcaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaATGGTTTCA
PfClag3.2-5'UTR    caatttgaaaa---aaaatattgtacacgtgaataaagtattagtaaaaaATGGTTTCA
reich908g11.plk    caatttgaaaa---aatatattgtacacgtgaataaagtattagtaaaaaATGGTTTCA
                   ***  ** * *    * * *    * *    ** ***  *  * **************
                                                                 start
3' UTR
PfClag3.1-3'UTR    TTTCACTAAtataaaatataaaaatatatatatatata------------atttgtaca
reich1194c08.plk   TTTCACTAAtataaaatataaatatatatatatatatatatatatatataatttgtaca
PfClag3.2-3'UTR    TTTCATTAAtatatatacataca-tacctatgtatatgtgtaatta--------------
reich289f06.qlk    TTTCACTAAaatatatacatata-tatctatgtatatatatgtgtaatatatatatatat
                   ***** *** *** * * *** *   * *** *****
                         stop
```

Fig. 6. Nucleotide sequence alignment of the 5′ and 3′ UTRs for *P. falciparum clag3.1* and *3.2* genes and *P. reichenowi* orthologous sequences. A *P. reichenowi* sequence possessing homology with *pfclag3.1* was not found in the current database. Nucleotide sequences corresponding to the ORF and the UTR are shown with upper case and lower case letters, respectively. Putative start and stop codons are boxed. Characteristic nucleotides are displayed in reverse (*clag3.1*) or masked with gray (*clag3.2*).

To date, observation of such high levels of polymorphism for *clag2*, *3.1*, *3.2*, and *8* ($\pi = 0.0053$–0.0164; $d_N = 0.0032$–0.0065) has not been reported for other known malaria rhoptry protein genes. The high polymorphism observed in *clag2*, *3.1*, and *3.2* is consistent with the observation by Kidgell et al. [40] based on the hybridization of genomic DNA from a panel of parasite lines to an oligonucleotide array for the *P. falciparum* genome. In addition, the polymorphism levels are comparable to those of the microneme proteins such as *eba-175* ($\pi = 0.0030$; $d_N = 0.0037$) and *ama-1* ($\pi = 0.0166$; $d_N = 0.0207$), which are exposed to host immune responses [17,18]. Rhoptry proteins are released into the PV and are considered to be minimally exposed to host immunity. If RhopH1/Clag polymorphism is generated by host immune pressure, the questions arises as to how RhopH1/Clag is exposed to host immunity. There are a few possible explanations. First, RhopH1/Clag may be released from the merozoites before attachment to the erythrocyte surface, thereby becoming a target of host immunity. Second, the RhopH complex, which is released into PVs, may be leaked to the surface of infected erythrocytes through the junction between invading parasite and the erythrocyte membrane. Leaked RhopH complex, and therefore parasite-infected erythrocytes, are then potential targets of host immunity. Indeed, the *Pf*RhopH complex and rhoptry-associated protein 2 (RAP-2, RSP-2), another malaria rhoptry protein, have been detected on the erythrocyte surface upon parasite attachment to erythrocytes [41,42]. The last possibility is that RhopH1/Clag, after release into the PV, may be distributed to the parasite-derived membranous network (i.e., Maurer's clefts) in the erythrocyte cytosol, where it is exposed to host immunity. RhopH2 and RhopH3 have recently been observed in materials deriving from Maurer's cleft by proteome analyses, consistent with this possibility [43,44].

There are no obvious associations between particular haplotypes and their geographic origins, and most haplotypes co-exist in different geographic areas, similar to other known polymorphic antigens such as MSP-1 [45,46]. RhopH1/Clag polymorphism might be maintained in natural parasite populations to evade host immunity. Using a T-cell epitope prediction algorithm (SYFPEITHI software) [47], we found that binding of a predicted T-cell epitope peptide to a particular HLA allotype was dramatically affected by the RhopH1/Clag polymorphism in silica. For example, aa positions 1094–1108 of 3D7 line Clag8 (KRISTSIDHISGGKW) was predicted as a T-cell epitope of HLA-DRB1*1101 with a score of 22, but the score for the corresponding region of Camp line Clag8 (MRISSTSTYISNNEW) was 0, emphasizing a potential involvement in immune evasion of RhopH1/Clag polymorphism. We consider the algorithm useful because the score of HLA-DRB1*0701 for the *Pf*CSP Th2R domain, a well characterized malaria polymorphic T-cell epitope peptide, is 22 for the K1 parasite line (KIQYSLSTEWSPCSV) but only 12 for that of the 3D7 line (KIQNSLSTEWSPCSV).

4.2. Evolution of *Pf*RhopH1/Clag polymorphism of the extant *P. falciparum* population

Gene conversion has been reported for other *P. falciparum* loci, such as *falcipain 2* [48] and *var* [49], as a source of genetic diversity. In this study, we show that *clag3.1/3.2*, which interchange their sequences by gene conversion, evolved more rapidly than *clag2*. The precise function of the RhopH complex remains unknown, and thus whether the gene conversion observed was functionally advantageous or neutral is also unknown; however, gene conversion can be a mechanism for antigenic variation to evade host immunity. Some examples include the *vsg* gene *of Trypanosoma brucei*, causative agent of African sleeping sickness, the *ves* gene of a cattle parasite *Babesia bovis* [50], and *var* genes in *P. falciparum* [51].

Based on the shared hybridization pattern between *clag3.1* and *3.2*, Chung et al. [36] proposed that these genes are alleles of the same locus; however, because the origin of two *rhoph1/clag* loci on chr 3 appear to predate the *P. falciparum*–*P. reichenowi* divergence, these should be categorized as paralogous genes but not the same gene. Shared features between these loci detected by Southern blot hybridization by Chung et al. can be simply explained by the gene conversion identified in this study. *clag3h* could be generated by an unequal crossover between *clag3.1* and *3.2* of a set of chromosomes during meiosis. If such crossover had occurred, parasite lines possessing three *rhoph1/clag* on chr 3 would be expected, with a molecule having its 5′ end derived from *clag3.1* and its 3′ end from *clag3.2* (Supplemental Fig. S5, model 1); however, this type of *rhoph1/clag* was not detected in this study. KMWII line appears to possess 3 *rhoph1/clag*

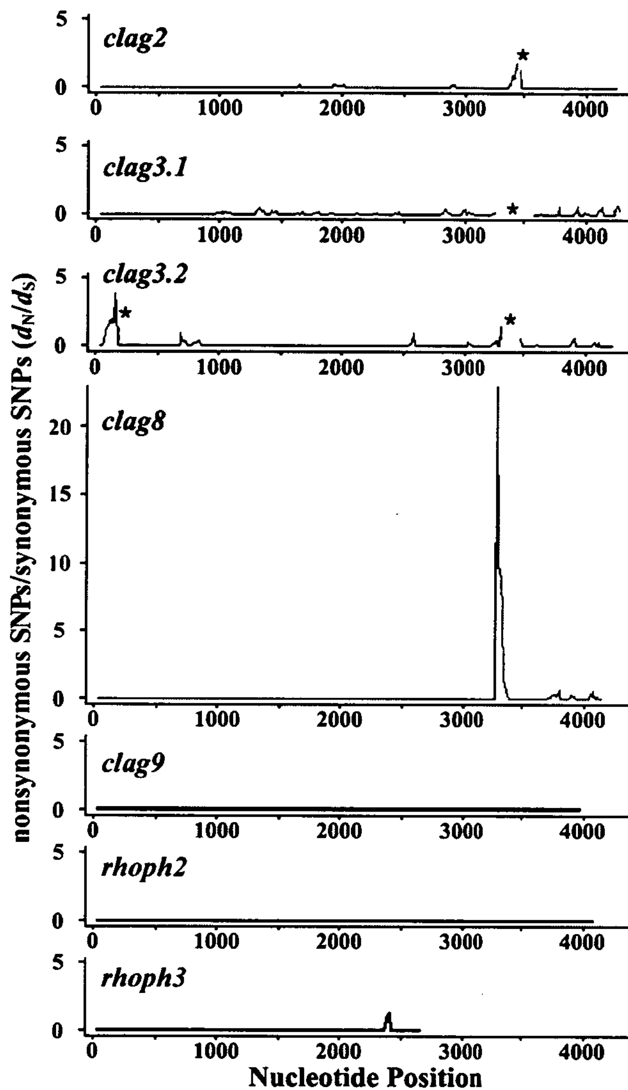

Fig. 7. Sliding window plot of d_N/d_S ratio for seven genes of the *Pf*RhopH complex. For optimum sequence alignment, nt positions 3433–3435 was excluded from *clag2*, nt positions 3337–3447 from *clag3.1*, and nt positions 88–99 and 3343–3444 from *clag3.2* for the analysis (asterisks), because these regions were highly polymorphic with indels. Nucleotide numbers are those of 3D7 line sequences. Window length is 50 bp, and step size is 10 bp.

genes on chr 3, but the third *rhoph1/clag* on chr 3 appears not to be generated by the mechanism described above, because this *rhoph1/clag* was obviously a duplicated *clag3.1* gene amplified with the *clag3.1*-specific primer set. Thus, *clag3h* is more likely a product of a recombination event between *clag3.1* and *3.2* on the same chromosome (Supplemental Fig. S5, model 2). Because a unique amino acid of Clag3H (e.g., Ala at aa position 1116) suggests a relatively old origin of *clag3h*, recombination events between *clag3.1* and *3.2* might be rare in the natural population.

Four highly polymorphic *rhoph1/clag* genes contained unexpectedly large numbers of synonymous substitutions. Based on the ratio of interspecific distance to intraspecific distance, the TMRCA of the polymorphism of *P. falciparum clag2* and *8* were estimated to be 1.89 (95% CI, 1.02–3.18) and 0.87 (95% CI, 0.42–1.54) mya, respectively. Although there is still controversy surrounding its accuracy, TMRCA of the extant *P. falciparum* population was estimated to be approximately 0.1–0.2 mya based on the genetic distance in nuclear genome housekeeping genes between *P. falciparum* and *P. reichenowi* ([52], Tanabe, unpublished data). Thus, polymorphism of *clag2* and *8* appears to be generated between the divergence of *P. falciparum* and *P. reichenowi* and TMRCA of the extant *P. falciparum* populations. Early origins of the polymorphism have been suggested for merozoite surface proteins *Pf*MSP-1 and *Pf*MSP-2, for which the origin of the polymorphism was proposed to predate the *P. falciparum*–*P. reichenowi* divergence (thus termed 'ancient origin'), or TMRCA of the extant *P. falciparum* population, respectively [53,54]. Early origins of the polymorphism older than TMRCA of extant *P. falciparum* populations would suggest that *rhoph1/clag* polymorphisms confer an advantage to the parasite and were positively selected for during the recent evolution of *P. falciparum*.

In summary, four factors appear to affect current *rhoph1/clag* polymorphism; (i) older origin than TMRCA of the extant *P. falciparum* population; (ii) gene conversion and (iii) copy number polymorphism for *rhoph1/clag* on chr 3; and (iv) positive diversifying selection. Multigene families play important roles in many aspects of malaria biology, e.g., responsibility for redundancy of erythrocyte invasion or antigenic variation of parasite-infected erythrocytes. Given the abundance of multigene families in the *P. falciparum* genome [55], combination of the mechanisms described in this study can be a powerful driving force to generate high biologic redundancy for parasite survival.

Acknowledgements

We thank NIAID intramural editor Brenda Rae Marshall for assistance. Sequence data of *P. reichenowi* were obtained from The Sanger Centre website (http://www.sanger.ac.uk/Projects/P_reichenowi/). This work was supported in part by Grants-in-Aid for Scientific Research 15406015 (to MT), 18390131 (to KT), and 17590372 and 17406009 (to OK) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA (to XS).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2007.11.004.

References

[1] Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. Cell 2006;124:755–66.

[2] Kaneko O. Erythrocyte invasion: vocabulary and grammar of the *Plasmodium* rhoptry. Parasitol Int 2007;56:255–62.

[3] Bannister LH, Mitchell GH, Butcher GA, Dennis ED. Lamellar membranes associated with rhoptries in erythrocytic merozoites of *Plasmodium knowlesi*: a clue to the mechanism of invasion. Parasitology 1986;92:291–303.

[4] Sam-Yellowe TY, Perkins ME. Interaction of the 140/130/110 kDa rhoptry protein complex of *Plasmodium falciparum* with the erythrocyte membrane and liposomes. Exp Parasitol 1991;73:161–71.
[5] Rungruang T, Kaneko O, Murakami Y, et al. Erythrocyte surface glycosylphosphatidyl inositol anchored receptor for the malaria parasite. Mol Biochem Parasitol 2005;140:13–21.
[6] Lustigman S, Anders RF, Brown GV, Coppel RL. A component of an antigenic rhoptry complex of *Plasmodium falciparum* is modified after merozoite invasion. Mol Biochem Parasitol 1988;30:217–24.
[7] Cowman AF, Baldi DL, Duraisingh M, et al. Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. FEBS Lett 2000;476:84–8.
[8] Holder AA, Freeman RR, Uni S, Aikawa M. Isolation of a *Plasmodium falciparum* rhoptry protein. Mol Biochem Parasitol 1985;14:293–303.
[9] Brown HJ, Coppel RL. Primary structure of a *Plasmodium falciparum* rhoptry antigen. Mol Biochem Parasitol 1991;49:99–110.
[10] Shirano M, Tsuboi T, Kaneko O, Tachibana M, Adams JH, Torii M. Conserved regions of the *Plasmodium yoelii* rhoptry protein RhopH3 revealed by comparison with the *P. falciparum* homologue. Mol Biochem Parasitol 2001;112:297–9.
[11] Ling IT, Kaneko O, Narum DL, et al. Characterisation of the *rhoph2* gene of *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2003;127:47–57.
[12] Kaneko O, Tsuboi T, Ling IT, et al. The high molecular mass rhoptry protein, RhopH1, is encoded by members of the *clag* multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2001;118:237–45.
[13] Holt DC, Gardiner DL, Thomas EA, et al. The cytoadherence linked asexual gene family of *Plasmodium falciparum*: are there roles other than cytoadherence? Int J Parasitol 1999;29:939–44.
[14] Ling IT, Florens L, Dluzewski AR, et al. The *Plasmodium falciparum clag9* gene encodes a rhoptry protein that is transferred to the host erythrocyte upon invasion. Mol Microbiol 2004;52:107–18.
[15] Kaneko O, Yim Lim BY, Iriko H, et al. Apical expression of three RhopH1/Clag proteins as components of the *Plasmodium falciparum* RhopH complex. Mol Biochem Parasitol 2005;143:20–8.
[16] Ghoneim A, Kaneko O, Tsuboi T, Torii M. The *Plasmodium falciparum* RhopH2 promoter and first 24 amino acids are sufficient to target proteins to the rhoptries. Parasitol Int 2007;56:31–43.
[17] Polley SD, Conway DJ. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics 2001;158:1505–12.
[18] Baum J, Thomas AW, Conway DJ. Evidence for diversifying selection on erythrocyte-binding antigens of *Plasmodium falciparum* and *P. vivax*. Genetics 2003;163:1327–36.
[19] Siddiqui WA, Tam LQ, Kramer KJ, et al. Merozoite surface coat precursor protein completely protects *Aotus* monkeys against *Plasmodium falciparum* malaria. Proc Natl Acad Sci USA 1987;84:3014–8.
[20] Cooper JA, Ingram LT, Bushell GR, et al. The 140/130/105 kilodalton protein complex in the rhoptries of *Plasmodium falciparum* consists of discrete polypeptides. Mol Biochem Parasitol 1988;29:251–60.
[21] Doury JC, Bonnefoy S, Roger N, Dubremetz JF, Mercereau-Puijalon O. Analysis of the \ weight rhoptry complex of *Plasmodium falciparum* using monoclonal antibodies. Parasitology 1994;108:269–80.
[22] Trager W, Jensen JB. Human malaria parasites in continuous culture. Science 1976;193:673–5.
[23] Jongwutiwes S, Tanabe K, Nakazawa S, Uemura H, Kanbara H. Coexistence of GP195 alleles of *Plasmodium falciparum* in a small endemic area. Am J Trop Med Hyg 1991;44:299–305.
[24] Kaneko O, Soubes SC, Miller LH. *Plasmodium falciparum*: invasion of *Aotus* monkey red blood cells and adaptation to *Aotus* monkeys. Exp Parasitol 1999;93:116–9.
[25] Su X, Ferdig MT, Huang Y, et al. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. Science 1999;286:1351–3.
[26] Mu J, Ferdig MT, Feng X, et al. Multiple transporters associated with malaria parasite responses to chloroquine and quinine. Mol Microbiol 2003;49:977–89.
[27] Thompson JD, Higgins DG, Gibson TJ, Clustal W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–80.
[28] Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004;5:150–63.
[29] Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 1986;3:418–26.
[30] Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003;19:2496–7.
[31] McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 1991;351:652–4.
[32] Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 1993;135:599–607.
[33] Haubold B, Wiehe T. Statistics of divergence times. Mol Biol Evol 2001;18:1157–60.
[34] Betrán E, Rozas J, Navarro A, Barbadilla A. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics 1997;146:89–99.
[35] Mu J, Duan J, Makova KD, et al. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. Nature 2002;418:323–6.
[36] Chung WY, Gardiner DL, Anderson KA, Hyland CA, Kemp DJ, Trenholme KR. The CLAG/RhopH1 locus on chromosome 3 of *Plasmodium falciparum*: two genes or two alleles of the same gene? Mol Biochem Parasitol 2007;151:229–32.
[37] Escalante AA, Ayala FJ. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. Proc Natl Acad Sci USA 1994;91:11373–7.
[38] Goodman M. The genomic record of Humankind's evolutionary roots. Am J Hum Genet 1999;64:31–9.
[39] Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 1988;335:167–70.
[40] Kidgell C, Volkman SK, Daily J, et al. A systematic map of genetic variation in *Plasmodium falciparum*. PLoS Pathog 2006;2:e57.
[41] Sam-Yellowe TY, Shio H, Perkins ME. Secretion of *Plasmodium falciparum* rhoptry protein into the plasma membrane of host erythrocytes. J Cell Biol 1988;106:1507–13.
[42] Douki JB, Sterkers Y, Lepolard C, et al. Adhesion of normal and *Plasmodium falciparum* ring-infected erythrocytes to endothelial cells and the placenta involves the rhoptry-derived ring surface protein-2. Blood 2003;101:5025–32.
[43] Sam-Yellowe TY, Fujioka H, Aikawa M, Hall T, Drazba JA. A *Plasmodium falciparum* protein located in Maurer's clefts underneath knobs and protein localization in association with Rhop-3 and SERA in the intracellular network of infected erythrocytes. Parasitol Res 2001;87:173–85.
[44] Vincensini L, Richert S, Blisnick T, et al. Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering *Plasmodium falciparum* proteins to the surface of its host cell. Mol Cell Proteomics 2005;4:582–93.
[45] Ferreira MU, Kaneko O, Kimura M, Liu Q, Kawamoto F, Tanabe K. Allelic diversity at the merozoite surface protein-1 (MSP-1) locus in natural *Plasmodium falciparum* populations: a brief overview. Mem Inst Oswaldo Cruz 1998;93:631–8.
[46] Sakihama N, Ohmae H, Bakote B, Kawabata M, Hirayama K, Tanabe K. Limited allelic diversity of *Plasmodium falciparum* merozoite surface protein 1 gene from populations in the Solomon Islands. Am J Trop Med Hyg 2006;74:31–40.
[47] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 1999;50:213–9.
[48] Nielsen KM, Kasper J, Choi M, et al. Gene conversion as a source of nucleotide diversity in *Plasmodium falciparum*. Mol Biol Evol 2003;20:726–34.

[49] Freitas-Junior LH, Bottius E, Pirrit LA, et al. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature 2000;407:1018–22.

[50] Dzikowski R, Deitsch K. Antigenic variation by protozoan parasites: insights from *Babesia bovis*. Mol Microbiol 2006;59:364–6.

[51] Kraemer SM, Kyes SA, Aggarwal G, et al. Patterns of gene recombination shape *var* gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. BMC Genomics 2007;8:45.

[52] Tanabe K, Sakihama N, Hattori T, et al. Genetic distance in housekeeping genes between *Plasmodium falciparum* and *Plasmodium reichenowi* and within *P. falciparum*. J Mol Evol 2004;59:687–94.

[53] Polley SD, Weedall GD, Thomas AW, Golightly LM, Conway DJ. Orthologous gene sequences of merozoite surface protein 1 (MSP-1) from *Plasmodium reichenowi* and *P. gallinaceum* confirm an ancient divergence of *P. falciparum* alleles. Mol Biochem Parasitol 2005;142:25–31.

[54] Roy SW, Ferreira MU, Hartl DL. Evolution of allelic dimorphism in malarial surface antigens. Heredity 2008;100:103–10.

[55] Michon P, Stevens JR, Kaneko O, Adams JH. Evolutionary relationships of conserved cysteine-rich motifs in adhesive molecules of malaria parasites. Mol Biol Evol 2002;19:1128–42.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Infection, Genetics and Evolution 7 (2007) 180–187
Infection, Genetics and Evolution
www.elsevier.com/locate/meegid

A comparative study of the genetic diversity of the 42 kDa fragment of the merozoite surface protein 1 in *Plasmodium falciparum* and *P. vivax*

M. Andreina Pacheco[a], Amanda C. Poe[b], William E. Collins[b], Altaf A. Lal[b,1], Kazuyuki Tanabe[c], Simon K. Kariuki[d], Venkatachalam Udhayakumar[b], Ananias A. Escalante[a,*]

[a] *School of Life Sciences, Arizona State University, P.O. Box 874501, Tempe, AZ 85287-4501, United States*
[b] *Malaria Branch, Division of Parasitic Diseases, Centers for Disease Control and Prevention, Chamblee, GA, United States*
[c] *International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan*
[d] *Center for Vector Biology and Control Research, Kenya Medical Research Institute, Kisumu, Kenya*

Received 12 May 2006; received in revised form 22 August 2006; accepted 23 August 2006
Available online 28 September 2006

Abstract

We investigated the genetic diversity of the 42 kDa fragment of the merozoite surface protein 1 (MSP-1) antigen in *Plasmodium falciparum* and *P. vivax*, as well as in non-human primate malarial parasites. This fragment undergoes a proteolytic cleavage generating two fragments of 19 kDa ($MSP\text{-}1_{19}$) and 33 kDa ($MSP\text{-}1_{33}$) that are critical in erythrocyte invasion. We found that overall the $MSP\text{-}1_{33}$ fragment exhibits greater genetic diversity than the $MSP\text{-}1_{19}$ regardless of the species. We have found evidence for positive natural selection only in the human malaria parasites by comparing the rate of non-synonymous versus synonymous substitutions. In addition, we found clear differences between the two major human malaria parasites. In the case of *P. falciparum*, positive natural selection is acting on the $MSP\text{-}1_{19}$ region while the $MSP\text{-}1_{33}$ is neutral or under purifying selection. The opposite pattern was observed in *P. vivax*. Our results suggest different roles of this antigen in the host–parasite immune interaction in each of the major human malarial parasites.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Malaria; Merozoite; MSP-1; MSP-1 19 kDa; Genetic diversity; Vaccine; *Plasmodium*

1. Introduction

The malaria burden is particularly high in sub-Saharan Africa where *Plasmodium falciparum* is predominant. However, malaria "out of Africa" is characterized by the presence of *P. vivax*, the second most important malaria parasite in terms of its morbidity. Although there are clear biological and genetic differences between these two parasites (Coatney et al., 1971), they overlap in their geographic distribution and there is increasing evidence for their interaction (Snounou and White, 2004).

Among the antigens currently under consideration in malaria vaccine formulations, one of the most promising candidates is the major merozoite surface protein 1 (MSP-1) (Good et al., 2004). The MSP-1 antigen is expressed as a large protein of 190–200 kDa on the parasite surface (Holder and Freeman, 1982). This precursor undergoes two steps of proteolytic cleavage during the merozoite maturation. First, it is cleaved into four major fragments of 83, 30, 38 and 42 kDa (further referred to as $MSP\text{-}1_{83}$, $MSP\text{-}1_{30}$, $MSP\text{-}1_{38}$, and $MSP\text{-}1_{42}$) then, before erythrocyte invasion the $MSP\text{-}1_{42}$ fragment undergoes a second cleavage resulting in the generation of 33 and 19 kDa ($MSP\text{-}1_{33}$ and $MSP\text{-}1_{19}$) fragments where the latter remain on the merozoite surface during invasion.

Plasmodium spp. MSP-1 exhibits extensive genetic polymorphism (Tanabe et al., 1987; Putaporntip et al., 2002) that appears to be maintained by positive natural selection in *P. falciparum* (Hughes, 1992; Escalante et al., 1998; Conway et al.,

* Corresponding author. Tel.: +1 480 965 3739; fax: +1 480 965 6899.
E-mail address: ananias.escalante@asu.edu (A.A. Escalante).
[1] Present address: Health Attaché and HHS Regional Representative for South Asia, U.S. Embassy, New Delhi, India.

1567-1348/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2006.08.002

2000) and *P. vivax* (Putaporntip et al., 2006). Similar observations have been made about other malarial vaccine antigens (see Escalante et al., 2004) on which the host immune system is considered the driving selective force that allows for the accumulation and frequent switch of suitable mutations in the parasite population. Under this scenario, mutations are maintained longer in the parasite population than expected if genetic drift were the sole process acting on the genetic polymorphism.

The conclusion that positive selection maintains the genetic diversity of genes encoding malarial antigens is supported, among others lines of evidence, by the observation in *P. falciparum* that non-synonymous nucleotide substitutions (those that change the amino acid) are more common than synonymous substitutions (mutations that do not change the amino acid) (Hughes and Hughes, 1995; Escalante et al., 1998, 2004). Since natural selection acts on phenotypic differences, an excess of non-synonymous substitutions over synonymous is considered evidence that natural selection is favoring the maintenance of genetic polymorphism.

In the case of *Plasmodium* spp. MSP-1, most of the genetic diversity analyses have subdivided the gene into blocks (segments) based on their level of genetic diversity but not using any other biological criteria (Tanabe et al., 1987; Putaporntip et al., 2002, 2006); however, few studies have been done considering the proteolytic fragments as functional units (Escalante et al., 1998).

The MSP-1_{42} and MSP-1_{19} fragments have received special attention in *P. falciparum* as part of vaccine formulations given that they are relatively conserved and antibodies against these fragments inhibit the parasite invasion into the red blood cells (Yang et al., 1999; Stanisic et al., 2004). In addition, the critical role of the MSP-1_{19} fragment in the erythrocyte invasion is conserved even among distantly related species (O'Donnell et al., 2000).

An important characteristic of *P. vivax* is that it invades reticulocytes, a process that is mediated by specific proteins such as the reticulocyte binding proteins and Duffy receptor (Galinski et al., 1992; Chitnis and Miller, 1994). However, MSP-1 in *P. vivax* also appears to play an important role in this process (Rodríguez et al., 2002; Espinosa et al., 2003; Han et al., 2004; Sachdeva et al., 2004). Indeed, peptides with high specific binding activity (HSBA) to reticulocytes have been found in the MSP133 (Espinosa et al., 2003; Rodríguez et al., 2002).

This investigation aims to compare the genetic diversity of the MSP-1_{42} in *Plasmodium* spp. focusing on *P. falciparum* and *P. vivax*. We have analyzed 120 sequences of the MSP-1_{42} of *P. falciparum* and 75 sequences of the homologous region in *P. vivax*, and we have explored the genetic diversity of the MSP-1_{33} and MSP-1_{19} fragments. In the case of *P. vivax*, we further explored its genetic diversity by comparing it with the homologous regions in primate malarial parasites that are closely related to *P. vivax* (Escalante et al., 2005). Although we find evidence that positive natural selection is acting on the observed polymorphism in MSP-1_{42}, it operates differently in each of the two major human malarial parasites. We conclude that inferences made about *P. falciparum* MSP-1 cannot simply be "translated" into *P. vivax*.

2. Materials and methods

The gene encoding the 42 kDa fragment of MSP-1 or MSP-1_{42} was amplified by polymerase chain reaction (PCR). The primers forward-GAA TGA TAT TCC TAA GAA GTT AGA GG and reverse-GAT AGA TTA TTT AAT AAG AAA AAA GAA CTT GGC CAA GAC AAA ATG C were used to amplify the partial *P. falciparum* 3′ sequences. The PCR conditions for amplifying the MSP-1_{42} from *P. falciparum* were: a partial denaturation at 94 °C for 1 min and 30 cycles with 1 min at 94 °C, 1 min at 50 °C and 3 min extension at 72 °C. A final extension of 3 min was added in the last cycle. The primers forward-GAC CAA GTA ACA ACG GGA G and reverse-CAA AGA GTG GCT CAG AAC C were used for *P. vivax*, *P. cynomolgi*, *P. inui*, and *P. knowlesi*. In the case of *P. fragile*, we used the forward primer GAC CAA GTA ACA ACG GG. The PCR conditions for amplifying the MSP-1_{42} from *P. vivax* and non-human primate malarias were: a partial denaturation at 94 °C for 3 min and 35 cycles with 1 min at 94 °C, 45′ at 50–58 °C and 2 min extension at 72 °C, a final extension of 10 min was added in the last cycle.

The amplified product was purified, cloned using the pGEM-TEasy Vector System I from Promega (USA), and sequenced. Both strands were sequenced from at least two clones. The alignment was performed using ClustalW Version 1.7 with manual editing using the alignment reported by Miller et al. (1993) in the case of *P. falciparum* and those reported by Putaporntip et al. (2002, 2006) in the case of *P. vivax* and related species.

In the case of *P. falciparum*, we sequenced the MSP-1 42 kDa in 34 isolates from Asembo Bay, western Kenya in this investigation. In addition, a total of 20 isolates (5 from India, 9 from Venezuela, and 6 from Thailand) were sequenced for the 3′ end. We used in our investigation prior published sequences (Chang et al., 1988; Qari et al., 1998; Jongwutiwes et al., 1992, 1993; Tanabe et al., 2004) and unpublished sequences under the accession numbers U20726–U20733 and U20653–U20656. A total of 120 MSP-1_{42} sequences were considered in our analyses. In addition, we included 55 sequences of the MSP-1_{19} reported in the literature (Kaneko et al., 1997; Kumar et al., 2005) and unpublished sequences under the accession numbers AF29507–AF29537 in order to obtain a complete picture of the MSP-1_{19} alleles that have been reported.

In the case of *P. vivax*, we report five sequences from laboratory isolates (Rio Meta, Sumatra I, Indonesia I, Mauritania I, and Vietnam II) and used the sequences reported in literature (Putaporntip et al., 2000, 2002) for a total of 75 sequences. In addition, we analyzed 10 sequences from different isolates of *P. cynomolgi* (the sequence AY869723 from the GenBank together with new sequences from the strains B strain, Berok, Cambodian, Ceylonensis, Gombok, Mulligan, PT1, PT2, and RO), 15 sequences from isolates of *P. inui* (Celebes I and II, Hackeri, Hawking, Leaf Monkey I and II, Leucosphyrus, Mulligan, N-34, OS, Perak, Perlis, Philippine, Taiwan I and II), a sequence of *P. knowlesi* (Hackery strain), *P. hylobati* (parasite from gibbons), and *P. fragile* (Nilgiri strain). Information about the biology of these species and the origin of

the isolates can be found elsewhere (Coatney et al., 1971). All the primate malaria strains were provided by the Centers for Disease Control and Prevention. The sequences reported in this study are deposited in the GenBank with the accession numbers DQ907617–DQ907702.

2.1. Statistical analysis

We estimate genetic polymorphism by using the parameter π, which estimates the average number of substitutions between any two sequences. The average number of synonymous (Ds) and non-synonymous substitutions (Dn) between a pair of sequences was investigated to explore the effect of natural selection. The average numbers of synonymous and non-synonymous substitutions are estimated using two methods: Nei and Gojobori's method (1986) with the Jukes and Cantor correction, and the Li's method (1993) as implemented in the MEGA program (Kumar et al., 2001). We estimated the difference between Ds and Dn, its standard deviation was calculated using bootstrap with 1000 pseudo-replications for Ds and Dn, as well as a two tail Z-test on the difference between Ds and Dn (Nei and Kumar, 2000). The null hypothesis is that Ds = Dn; thus, we assumed as null hypothesis that the observed polymorphism was neutral.

The Tajima's D statistic and F^* from Fu and Li were estimated for testing the hypothesis that the allele frequency spectrum is compatible with the neutral model (Tajima, 1989; Fu and Li, 1993). Under the neutral model, Tajima's D and F^* are approximately equal to zero, thus any deviation from zero would indicate a departure from neutrality in the allele frequency spectrum.

Evidence for recombination was assessed by using the Rm parameter that estimates the minimum number of recombination events in the history of the sample. Rm is obtained using the four-gamete test (Hudson and Kaplan, 1985) and, as the name of the parameter indicates, it is a conservative estimate of the number of recombination events.

In the case of *P. vivax* and related non-human primate malarial parasites, the gene genealogy of the MSP-1_{42} alleles was determined by using the Neighbor–Joining (Saitou and Nei, 1987) method with the Tamura-Nei model. The reliability of the nodes in the NJ tree was assessed by the bootstrap method with 1000 pseudo-replications. The genealogy was estimated using the MEGA program (Kumar et al., 2001). The assumption of neutrality was also tested in *P. vivax* MSP-1 by using the McDonald and Kreitman test (McDonald and Kreitman, 1991), which compares the intra-and interspecific number of synonymous and non-synonymous sites; significance was assessed by using a Fishers exact test for the 2×2 contingency table as implemented in the programs DNAsp Version 4.0 (Rozas et al., 2003). In this analysis, we compare *P. vivax* with *P. cynomolgi* and *P. inui* (see below).

3. Results

Table 1 shows the genetic diversity found in the MSP-1_{42} fragments in *P. falciparum* and *P. vivax*. Overall, the genetic diversity of *P. falciparum* is twice that observed in *P. vivax* (π of 0.05042 versus 0.02184). Analysis of the genetic diversity of the MSP-1_{33} and MSP-1_{19} fragments confirmed previous observations that the MSP-1_{19} fragment is more conserved than the MSP-1_{33} fragment (Table 1) in both human malarial parasites. *P. vivax* MSP-1_{19} has only one polymorphic site while in *P. falciparum* the substitutions are concentrated in five residues within the epidermal growth factor like domains (EGF). In an extended alignment that included all the MSP-1_{19} sequences reported in the literature at the time of this study ($n = 175$); we found 11 alleles reported based in these five residues, among them, there are four common alleles that have a worldwide distribution: E-KNG-L ($n = 54$), E-TSR-L ($n = 41$), Q-KNG-F ($n = 20$), Q-KNG-L ($n = 33$). It is worth noting that some alleles, although reported in low frequency, have been found in two continents; such are the cases of E-KNG-F ($n = 8$ reported in India and Kenya), E-KSR-L ($n = 4$ reported in Kenya, South Africa, and Vanuatu), and Q-TSR-L ($n = 3$ reported in India and Papua New Guinea). The allele E-TSG-L ($n = 9$) has been reported three times in India (including this study) and is the one observed in *P. reichenowi*, the most closely related species to *P. falciparum* found in chimpanzees (Coatney et al., 1971).

We found two recombination-convergent events using the Rm method (Hudson and Kaplan, 1985); these events are

Table 1
Polymorphism found in the MSP-1_{42} in *P. falciparum* and *P. vivax*

	π	Ds	Dn	Ds − Dn (S.D.)	Z	Tajima D	F^*
P. falciparum (n = 120)							
42 KDa	0.05042	0.0821	0.0541	0.0280 (0.011)	Ds > Dn (P<0.05)	−0.11184 n.s.	0.13353 n.s.
33 KDa	0.06551	0.1236	0.0741	0.0494 (0.020)	Ds > Dn (P<0.05)	0.10150 n.s.	0.86091 n.s.
19 KDa	0.00884	0.0013	0.0107	−0.009 (0.004)	Ds < Dn (P<0.05)	−1.72070 (0.10>P>0.05)	−4.78810 (P<0.05)
P. vivax (n = 75)							
42 KDa	0.02184	0.0125	0.0249	−0.0123 (0.005)	Ds < Dn (P<0.05)	2.19241 (P<0.05)	2.09599 (P<0.05)
33 KDa	0.03249	0.0162	0.0325	−0.0160 (0.006)	Ds < Dn (P<0.05)	2.31357 (P<0.05)	2.24458 (P<0.05)
19 KDa	0.0006	0.0005	0.0006	0.0001 (0.000)	Ds = Dn	−1.02018 n.s	−1.02018 n.s

π, nucleotide diversity; n, number of sequences; Dn is the nucleotide diversity of non-synonymous mutations per non-synonymous sites and Ds is the nucleotide diversity of synonymous mutations per synonymous site using the Nei and Gojobori method; Ds − Dn are the difference of Ds and Dn with their standard deviation, SD, estimated by bootstrap with 1000 pseudo replicates; Z is the Z-test (Nei and Kumar, 2000); Tajima D and F^* are tests for detecting departures from the neutral model.

illustrated using the relative positions of the residues in the allele E-KNG-L, specifically between the position held by the amino acids E and K (separated by 138 bp) and between the positions filled by amino acids K and G (separated by 30 bp). Recombination events have been previously reported in MSP-1_{19} (Qari et al., 1998).

In order to explore the role of natural selection we further analyzed the genetic polymorphism in the MSP-1_{42} as a unit by estimating the number of synonymous (Ds) and non-synonymous (Dn) substitutions per site estimated by the Nei and Gojobori method with the Jukes and Cantor correction. When this comparison is made, both parasites exhibit opposite patterns: MSP-1_{42} in *P. falciparum* shows more synonymous than non-synonymous substitutions while the homologous region in *P. vivax* shows more non-synonymous than synonymous substitutions. In both cases the differences are significant with a Z-test (Nei and Kumar, 2000) (Table 1). The Li's method gives identical results. We explore departure from neutrality by using the Tajima's D test (Tajima, 1989) and F^* test (Fu and Li, 1993). These tests should to be used with caution since they aim to detect departures from a neutral panmictic population, an assumption that is violated by these geographically and temporally spaced samples. Nevertheless, we used them to explore the distribution of haplotypes in our samples as was used previously to compare *P. vivax* and *P. knowlesi* (Putaporntip et al., 2006). These tests could not detect departure from neutrality in *P. falciparum*, although they did so in *P. vivax* when the complete MSP-1_{42} was considered as a unit.

We explored the diversity in the MSP-1_{33} and MSP-1_{19} fragments separately by comparing the number of synonymous and non-synonymous substitutions in each species. In the case of the MSP-1_{33} of *P. falciparum* there are more synonymous than non-synonymous substitutions ($P < 0.05$) (Table 1), while the contrary was observed in the MSP-1_{19} where there are more non-synonymous than synonymous substitutions ($P < 0.05$). These results suggest that while the MSP-1_{19} is under positive selection in *P. falciparum*, the MSP-1_{33} is under purifying selection; that is, natural selection favors the maintenance of amino acid polymorphism in the MSP-1_{19} while it holds back the rate of amino-acid polymorphism in the MSP-1_{33}. Differences between the MSP-1_{33} and MSP-1_{19} were also observed by using the Tajima's D and F^* tests (Table 1): there is not a departure from neutrality in the MSP-1_{33} while the MSP-1_{19} polymorphism rejects the expectation under the neutral model. Although the significance level by the Tajima's D test is weak for MSP-1_{19} ($0.05 < P < 0.1$), there is almost no synonymous variation, substantiating a departure from the neutrality in this region. It is important to notice that the Tajima's D and F^* tests have a negative value indicating that there is an excess of low frequency variants in the sample (Table 1).

In the case of *P. vivax* the pattern is the opposite. There are more non-synonymous than synonymous substitutions in the MSP-1_{33} while there is almost no variation in the MSP-1_{19} (Table 1). The polymorphism in the *P. vivax* MSP-1_{33} is not evenly distributed. Indeed, there is a region of 105 bp out of 848 bp in MSP-1_{33} (35 amino acids) where a clear excess of non-synonymous versus synonymous substitution is observed driving the overall MSP-1_{33} results. In addition, there is a departure from neutrality in the MSP-1_{33} when the Tajima's D and F^* tests are applied. However, contrasting with *P. falciparum*, the value of the test is positive as the result of an excess of variants in intermediate frequencies.

We further explore the hypothesis that positive selection is acting on the *P. vivax* MSP-1_{33} fragment by comparing it with its closely related non-human primate malarial parasites (Escalante et al., 2005). The genealogy of the MSP-1_{42} fragments from the species reported in this study is depicted in Fig. 1. *P. cynomolgi* appears as sister taxa of *P. vivax*; however, this clade does not have strong support. *P. cynomolgi* strains are subdivided into two clear clades; no evidence for allele families could be observed with this fragment. *P. inui* and *P. hylobati* are closely related as previously reported (Escalante et al., 2005). The close relationship of these two species was further supported by the presence of a repetitive sequence in the MSP-1_{33} fragment. Specifically, a motif with the residues NE*QEEI* is inserted in some of the *P. inui* isolates while *P. hylobati* has the residues NE*QEEIKIRQEEI*. We also found an insertion in *P. knowlesi* that emerged as a duplication of the motif *INNCQIEK* conserved in *P. inui* and *P. vivax* (Fig. 2). Given the lack of

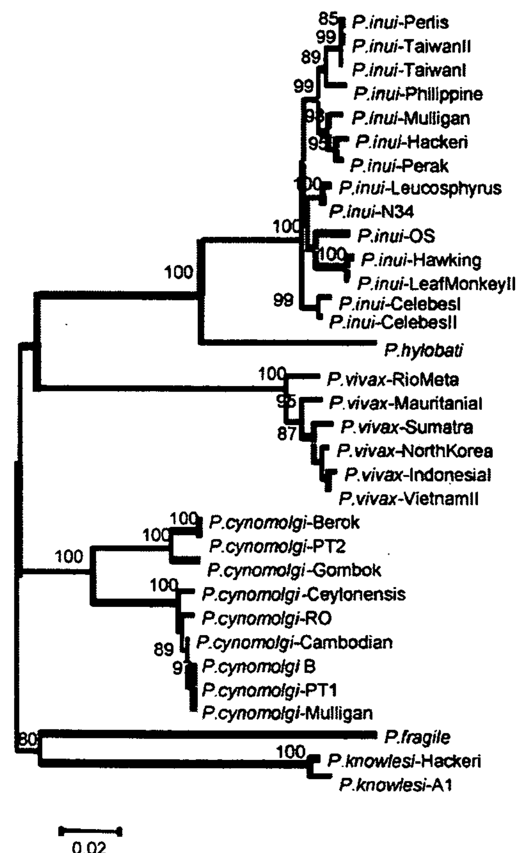

Fig. 1. Neighbor–Joining tree of the MSP-1_{42} alleles using Tamura-Nei's distance. The numbers on the nodes of the tree are percent of bootstrap values based on 1000 pseudo-replications. The sequences reported in this study are identified with their species and strain names.

```
Pvi-RioMeta        KTHLTAVNAQIKKVEDDIKK QDEEL-------KKIENEANKTAE       LVSKVNTYTDNLKKV---- -----INNCQLEK
Pvi-NorthKorea     KTHLDGVKTEIKKVEDDIKK QDEEL-------KKLGNVNSQDSK       LVSKVNTYTDNLKKV---- -----INNCQLEK
Pcy-Berok          KKHLDEVNAQIKEVEANINK QDEEL-------KQIESDTSKTAQ       LANKVQSYTENLKKF---- -----LNNYQIEK
Pcy-Gombok         KKHLDEVNAHIKEVEANINK QDEEI-------KKIGTDTTKTNE       LANKVHSYTENLKKF---- -----LNNYQIEK
Pin-Perak          KKQLDAVNKKIKEMEDEI-- ------------KKIPDEEPNSAT  >...< LVSMVTTYTNNLKKF---- -----INNCQIEK
Pin-LeafMonkeyII   KKQLDAVNKKIKEVEDEIND QEEEI-------EKISDEEQDAAI       LVSMVTTYTNNLKKF---- -----INNCQIEK
Phy                KKQFDAVNEKIKDLEDQIKE QEEEIKIRQEEIQRTSNDTNETDE       LVSMATTYTDNLKKF---- -----INNCQIEK
Pkn-Hackeri        KKHLEAVNAQIKEI------ -----------------EASVPGE       LVNMAHTYKENLKKFINNC QIEKSINNCQIEK
Pkn-A1             KQHLEAVNAQIKEI------ -----------------EASVPGE       LVNMAHTYKENLKKF---- -----INNCQIEK
Pfr                KNHMDAVHAHIQSI------ ----------------EKGDSETD       LMNKVHIYTDNLKKF---- -----MNKYPIEK
```

Fig. 2. Repetitive sequences observed in the MSP-1_{42}. The observed motifs are in italics. The dots (>. . .<) are indicating a non-repetitive portion of the protein that is not shown. The first three letters in the sequence codes indicate the species: Pvi, *P. vivax*; Pcy, *P. cynomolgi*; Pin, *P. inui*; *Phy*, *P. hylobati*; Pkn, *P. knowlesi*; Pfr, *P. fragile*.

Table 2
Polymorphism found in the MSP-1_{42} in other non-human *Plasmodium* spp.

	π	Ds	Dn	Ds − Dn (S.D.)	Z
P. cynomolgi (n = 10)					
42 KDa	0.03805	0.0871	0.0287	0.0585 (0.015)	Ds > Dn ($P<0.05$)
33 KDa	0.06551	0.1001	0.0312	0.0687 (0.018)	Ds > Dn ($P<0.05$)
19 KDa	0.02502	0.0469	0.0211	0.0257 (0.022)	Ds = Dn
P. inui (n = 15)					
42 KDa	0.02416	0.0284	0.0237	0.0049 (0.006)	Ds = Dn
33 KDa	0.02951	0.0358	0.0289	0.0071 (0.008)	Ds = Dn
19 KDa	0.0067	0.0051	0.0073	−0.0022 (0.005)	Ds = Dn

π, nucleotide diversity; n, number of sequences; Dn is the nucleotide diversity of non-synonymous mutations per non-synonymous sites and Ds is the nucleotide diversity of synonymous mutations per synonymous site using the Nei and Gojobori method. Ds − Dn are the difference of Ds and Dn with their standard deviation, S.D., estimated by bootstrap with 1000 pseudo replicates. Z is the Z-test (Nei and Kumar, 2000).

resolution of the phylogeny using this region, we used both *P. cynomolgi* and *P. inui* for comparison with *P. vivax*.

Table 2 shows the basic statistics for the MSP-1_{42} in these two non-human primate malarial parasites. As in the cases of the human parasites, the MSP-1_{33} fragment is more diverse than the MSP-1_{19}. However, in the case of the non-human primate malarias, there is no excess of non-synonymous substitutions over synonymous substitutions in the MSP-1_{42} as a unit or considering the MSP-1_{33} and MSP-1_{19} fragments separated. Thus, by comparing the rate of non-synonymous versus synonymous substitutions we could not detect evidence for positive selection acting on *P. cynomolgi* or *P. inui* MSP-1_{42}. An identical pattern can be observed in *P. knowlesi* when the two complete MSP-1_{42}, the one reported in this investigation and the one available in literature (Putaporntip et al., 2006) are compared, specifically Ds = 0.04275 and Dn = 0.00240 for MSP-1_{42}.

We then analyzed the genetic diversity of *P. vivax* MSP-1_{42} by using the McDonald and Kreitman test (McDonald and Kreitman, 1991) and compared it with both *P. cynomolgi* and *P. inui* samples. In the case of the complete 42 Kda, there was an overall excess of non-synonymous over synonymous in the *P. vivax* polymorphism when compared with *P. cynomolgi* ($p < 0.05$ using a Fisher's exact test). Similar results were found with *P. vivax* and *P. inui* ($p < 0.001$ using a Fisher's exact test). In both cases, the significance of the MK test was explained by an excess of amino acid replacements in the polymorphism of the *P. vivax* MSP-1_{33}. It is worth noting that no departure from neutrality was found when only MSP1_{19} was considered. It is also important to emphasize that no departure from neutrality was observed when *P. cynomolgi* and *P. inui* were compared considering the MSP-1_{42} as a unit, or separating it into the MSP-1_{33} and MSP-1_{19} fragments.

4. Discussion

The available data, mostly derived from *P. falciparum*, indicate the importance of the antibody response against block 2 (located in the 83 kDa or MSP-1_{83}) and the MSP-1_{42} fragments in developing protective immunity. In this study, we have described the selective forces operating on the polymorphism observed in the MSP-1_{42} fragment in the two major human malaria parasites. We have shown how the MSP-1_{33} and MSP-1_{19} fragments are under different selective pressures in each of the major human malarial parasites by using the rate of non-synonymous versus synonymous substitutions.

In the case of *P. falciparum*, the polymorphism in MSP-1_{33} appears to be neutral or under purifying selection while the polymorphism in MSP-1_{19} is under positive selection. In this case, our results are consistent with immunologic evidence suggesting that the MSP-1_{19} but not MSP-1_{33} elicits a protective immune response, though the latter being highly immunogenic (Ahlborg et al., 2002). Positive selection has been previously proposed as an important mechanism in maintaining the *P. falciparum* MSP-1 polymorphism in the form of balancing selection (Hughes, 1992; Conway et al., 2000); that is, natural selection maintains genetic polymorphism for a longer time than expected under a scenario where only

genetic drift is acting. A polymorphism under balancing selection is expected to have an excess of alleles in intermediate frequencies, a pattern that translates into positives Tajima's D and F^* tests. In the case of MSP-1_{19}, however, there is an excess of alleles in low frequency as evidenced by significant and negative values of the Tajima's D and F^* tests, not consistent with balancing selection. This could be the result of several factors. First, we found four alleles that are particularly common while several others are found in low frequency in our sample; low frequency alleles that are found even in different continents suggest an artifact due to a poor sampling effort. Indeed, lack of appropriate sampling could generate negative Tajima's D tests as a result of several sub-populations being analyzed together (Hammer et al., 2003). A second alternative is that a limited number of alleles are increasing in frequency, a scenario expected under a population expansion which coincides with the results reported for mitochondrial data (Joy et al., 2003).

Nevertheless, if the population demographic history and inappropriate sampling were the only factors leading to this result (significant and negative Tajimas's D and F^* tests), then the MSP-1_{33} should have shown a similar trend. The Tajimas's D and F^* tests for MSP-1_{33} are not only non-significant but also have an opposite sign. Interestingly, the MSP-1_{33} also shows more synonymous than non-synonymous substitutions. Therefore, we propose that the negative Tajimas's D and F^* tests, together with the excess of non-synonymous over synonymous substitutions in MSP-1_{19}, are the result of directional selection, that is, there are few MSP-1_{19} alleles increasing in frequency because they are positively selected.

Although the immune response against *P. falciparum* MSP-1_{19} is still under intense investigation, there is evidence suggesting that fine specificity rather than prevalence could be an important factor in the observed immune reactivity (Okech et al., 2004). Indeed, only partial cross-reactivity has been found in holoendemic areas among the most common MSP-1_{19} alleles (Udhayakumar et al., 1995; Shi et al., 1996; John et al., 2004). It has been also shown that immunity against MSP-1_{19} in *P. falciparum* has a short lifespan to the extent that its elicited antibody responses allow detecting differences in local transmission (Drakeley et al., 2005). Therefore, the pattern in the genetic polymorphism of MSP-1_{19} could be the result of differences of the most common alleles in their specificity and/or life spans of their elicited immune responses when compared with the less frequent MSP-1_{19} alleles, differences that give them a selective advantage favoring their transmission.

Our hypothesis that directional selection is operating on MSP-1_{19} does not contradict previous claims for balancing selection since they are well supported by the extensive divergence observed in MSP-1_{83}, MSP-1_{30}, and MSP-1_{38} fragments allowing the identification of two very distinctive allele families (Tanabe et al., 1987) that have been found to be an ancient polymorphism (Hughes, 1992; Polley et al., 2005) as well as evidence derived from population base studies of the MSP-1_{83} (Conway et al., 2000; Takala et al., 2006). Indeed such divergent allele families are not observed when only the MSP-1_{19} is considered.

In the case of *P. vivax*, however, the MSP-1_{33} and MSP-1_{19} fragments appear to be under different selective pressures than the ones just described in the homologous region in *P. falciparum*. We observed an excess of non-synonymous over synonymous substitutions in the MSP-1_{33} and not in the MSP-1_{19}; in addition, we found that the Tajimas's D and F^* tests are significant and positive for MSP-1_{33}, which is expected under the scenario of balancing selection although it could be the result of population structure, a clear possibility given the origin of the sample analyzed. Nevertheless, when we studied the genetic variation in the MSP-1_{33} and MSP-1_{19} by using the McDonald and Kreitmant test against *P. cynomolgi* and *P. inui* we found an excess of non-synonymous substitutions in the *P. vivax* MSP-1_{33} no matter which species we used to compare it with, suggesting that positive natural selection is operating in this fragment.

Our results support previous observations that *P. vivax* MSP-1_{33} could play an important role in reticulocyte invasion (Espinosa et al., 2003; Rodríguez et al., 2002). However, the polymorphism in the *P. vivax* MSP-1_{33} appears more complicated; indeed, there is a 105 bp fragment with high polymorphism located between regions where peptides with high specific binding activity (HSBA) to reticulocytes have been found (Espinosa et al., 2003; Rodríguez et al., 2002). These regions with HSBA are not only highly conserved among *P. vivax* isolates ($n = 75$) but also show more synonymous than non-synonymous substitutions when compared with *P. cynomolgi* (peptides 1735, 1738 and 1747 *sensu* Rodríguez et al., 2002 have Ks of 0.30, 0.31 and 0.22 versus Kn of 0.16, 0.025, and 0.11, respectively) and a similar pattern is observed when compared with *P. inui* (peptides 1735, 1738 and 1747 *sensu* Rodríguez et al., 2002 have Ks of 0.29, 0.37, and 0.041 versus Kn of 0.17, 0.10, and 0.16, respectively). This overall pattern indicates that these HSBA regions are under selective constraints to accumulate amino acid replacements; as a result, they could be a valuable target for a vaccine against *P. vivax* as has been suggested previously (Espinosa et al., 2003).

There is no information regarding the immunologic role played by the variation observed in *P. vivax* MSP-1_{33}. Elucidating whether it hampers effective natural immune responses against these conserved regions with HSBA to reticulocytes or whether it plays any other role requires further investigation. Nevertheless, it seems clear from this comparative analyses that we cannot simply extrapolate information derived from *P. falciparum* into *P. vivax* in the case of MSP-1_{42}.

In summary, we have investigated the genetic diversity of the sequence encoding the MSP-1_{42} in the two major human malarial parasites. We found evidence supporting positive natural selection as an important factor in the maintenance and generation of the observed polymorphism. However, we describe how natural selection is acting differently in the MSP-1_{33} and MSP-1_{19} fragments of the MSP-1_{42} in each of the two human malarial parasites. That is, our results suggest that these fragments, MSP-1_{33} and MSP-1_{19}, could play different roles in each of the two human malarial parasites.

Acknowledgements

This research is supported by grants from the National Institutes of Health (R01 GM60740) to A.A.E. and from the Japanese Ministry of Education, Culture, Sports, Science and Technology (14021125) to K.T. We thank Andrea McCollum and two anonymous reviewers for their commentaries that improved this manuscript. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References

Ahlborg, N., Ling, I.T., Howard, W., Holder, A.A., Riley, E.M., 2002. Protective immune responses to the 42-hilodalton (kDa) region of *Plasmodium yoelii* merozoite surface protein 1 are induced by the C-terminal 19-kDa region but not by large adjacent 33-kDa region. Infect. Immun. 70, 820–825.

Chang, S.P., Kramer, K.J., Yamaga, K.M., Kato, A., Case, S.E., Siddiqui, W.A., 1988. *Plasmodium falciparum*: Gene structure and hydropathy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate. Exp. Parasitol. 67, 1–11.

Chitnis, C.E., Miller, L.H., 1994. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. J. Exp. Med. 180, 497–506.

Coatney, G.R., Collins, W.E., Warren, M., Contacos, P.G., 1971. The Primate Malarias. U.S. Government Printing Office.

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

Drakeley, C.J., Corran, P.H., Coleman, P.G., Tongren, J.E., McDonald, S.L., Carneiro, I., Malima, R., Lusingu, J., Manjurano, A., Nkya, W.M., Lemnge, M.M., Cox, J., Reyburn, H., Riley, E.M., 2005. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc. Natl. Acad. Sci. U.S.A. 102, 5108–5113.

Escalante, A.A., Lal, A.A., Ayala, F.J., 1998. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. Genetics 149, 189–202.

Escalante, A.A., Cornejo, O.E., Rojas, A., Udhayakumar, V., Lal, A.A., 2004. Assessing natural selection in malarial parasites. Trends. Parasitol. 20, 388–395.

Escalante, A.A., Cornejo, O.E., Freeland, D.E., Poe, A.C., Durrego, E., Collins, W.E., Lal, A.A., 2005. A monkey's tale: The origin of *Plasmodium vivax* as a human malaria parasite. Proc. Nat. Acad. Sci. U.S.A. 102, 1980–1985.

Espinosa, A.M., Sierra, A.Y., Barrero, C.A., Cepeda, L.A., Cantor, E.M., Lombo, T.B., Guzman, F., Avila, S.J., Patarroyo, M.A., 2003. Expression, polymorphism analysis, reticulocyte binding and serological reactivity of two *Plasmodium vivax* MSP-1 protein recombinant fragments. Vaccine 21, 1033–1043.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Galinski, M.R., Medina, C.C., Ingravallo, P., Barnwell, J.W., 1992. A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. Cell 69, 1213–1226.

Good, M.F., Stanisic, D., Xu, H., Elliott, S., Wykes, M., 2004. The immunological challenge to developing a vaccine to the blood stages of malaria parasites. Immunol. Rev. 201, 254–267.

Hammer, M.F., Blackmer, F., Garrigan, D., Nachman, M.W., Wilder, J.A., 2003. Human population structure and its effects on sampling Y chromosome sequence variation. Genetics 164, 1495–1509.

Han, H.J., Park, S.G., Kim, S.H., Hwang, S.Y., Han, J., Traicoff, J., Kho, W.G., Chung, J.Y., 2004. Epidermal growth factor-like motifs 1 and 2 of *Plasmodium vivax* merozoite surface protein 1 are critical domains in erythrocyte invasion. Biochem. Biophys. Res. Commun. 320, 563–570.

Holder, A.A., Freeman, R.R., 1982. Biosynthesis and processing of a *Plasmodium falciparum* schizont antigen recognized by immune serum and a monoclonal antibody. J. Exp. Med. 156, 1528–1538.

Hudson, R.R., Kaplan, N.L., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147–164.

Hughes, A.L., 1992. Positive selection and intrallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. 9, 381–393.

Hughes, M.K., Hughes, A.L., 1995. Natural selection on *Plasmodium* surface proteins. Mol. Biochem. Parasitol. 71, 99–113.

John, C.C., O'Donnell, R.A., Sumba, P.O., Moormann, A.M., de Koning-Ward, T.F., King, C.L., Kazura, J.W., Crabb, B.S., 2004. Evidence that invasion-inhibitory antibodies specific for the 19 kDa fragment of merozoite surface protein-1 (MSP-1_{19}) can play a protective role against blood-stage *Plasmodium falciparum* infection in individuals in a malaria endemic area of Africa. J. Immunol. 173, 666–672.

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, 81–89.

Jongwutiwes, S., Tanabe, K., Kanbara, H., 1993. Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP-1) of *Plasmodium falciparum* from field isolates. Mol. Biochem. Parasitol. 59, 95–100.

Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., Beerli, P., Su, X.Z., 2003. Early origin and recent expansion of *Plasmodium falciparum*. Science 300, 318–321.

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.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Kumar, S.V., Ranjan, S., Saxena, V., Rajesh, V., Roy, S.K., Kochar, D., Ranjan, A., Das, A., 2005. *Plasmodium falciparum*: Genetic diversity of C-terminal region of MSP-1 in isolates from Indian sub-continent. Exp. Parasitol. 110, 384–388.

Li, W.H., 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitution. J. Mol. Evol. 36, 96–99.

McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351, 652–654.

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

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press.

O'Donnell, R.A., Saul, A., Cowman, A.F., Crabb, B.S., 2000. Functional conservation of the malaria vaccine antigen MSP-1_{19} across distantly related Plasmodium species. Nat. Med. 6, 91–95.

Okech, B.A., Corran, P.H., Todd, J., Joynson-Hicks, A., Uthaipibull, C., Egwang, T.G., Holder, A.A., Riley, E.M., 2004. Fine specificity of serum antibodies to *Plasmodium falciparum* merozoite surface protein, PfMSP-1(19), predicts protection from malaria infection and high-density parasitemia. Infect. Immun. 72, 1557–1567.

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. Mol. Biochem. Parasitol. 142, 25–31.

Putaporntip, C., Jongwutiwes, S., Seethamchai, S., Kanbara, H., Tanabe, K., 2000. Intragenic recombination in the 3′ portion of the merozoite surface protein 1 gene of *Plasmodium vivax*. Mol. Biochem. Parasitol. 109, 111–119.

Putapomtip, C., Jongwutiwes, S., Sakihama, N., Ferreira, M.U., Kho, W.G., Kaneko, A., Kanbara, H., Hattori, T., Tanabe, K., 2002. Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. Proc. Natl. Acad. Sci. U.S.A. 99, 16348–16353.

Putapomtip, 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*. Mol. Biochem. Parasitol. 146, 105–108.

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.

Rodríguez, L.E., Urquiza, M., Ocampo, M., Curtidor, H., Suarez, J., Garcia, J., Vera, R., Puentes, A., Lopez, R., Pinto, M., Rivera, Z., Patarroyo, M.E., 2002. *Plasmodium vivax* MSP-1 peptides have high specific binding activity to human reticulocytes. Vaccine 20, 1331–1339.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Sachdeva, S., Ahmad, G., Malhotra, P., Mukherjee, P., Chauhan, V.S., 2004. Comparison of immunogenicities of recombinant *Plasmodium vivax* merozoite surface protein 1 19- and 42-kiloDalton fragments expressed in *Escherichia coli*. Infect. Immun. 72, 5775–5782.

Saitou, N., Nei, M., 1987. The Neighbor–Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Shi, Y.P., Sayed, U., Qari, S.H., Roberts, J.M., Udhayakumar, V., Oloo, A.J., Hawley, W.A., Kaslow, D.C., Nahlen, B.L., Lal, A.A., 1996. Natural immune response to the C-terminal 19-kilodalton domain of *Plasmodium falciparum* merozoite surface protein 1. Infect. Immun. 64, 2716–2723.

Snounou, G., White, N.J., 2004. The co-existence of *Plasmodium*: sidelights from *falciparum* and *vivax* malaria in Thailand. Trends Parasitol. 20, 333–339.

Stanisic, D.I., Martin, L.B., Gatton, M.L., Good, M.F., 2004. Inhibition of 19-kDa C-terminal region of merozoite surface protein-1-specific antibody responses in neonatal pups by maternally derived 19-kDa C-terminal region of merozoite surface protein-1-specific antibodies but not whole parasite-specific antibodies. J. Immunol. 172, 5570–5581.

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

Takala, S.L., Escalante, A.A., Branch, O.H., Kariuki, S., Biswas, S., Chaiyaroj, S.C., Lal, A.A., 2006. Genetic diversity in the Block 2 region of the merozoite surface protein 1 (MSP-1) of *Plasmodium falciparum*: additional complexity and selection and convergence in fragment size polymorphism. Infect. Genet. Evol. 6, 417–424.

Tanabe, K., Mackay, M., Goman, M., Scaife, J.G., 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 195, 273–287.

Tanabe, K., Sakihama, N., Kaneko, A., 2004. Stable SNPs in malaria antigen genes in isolated populations. Science 303, 493.

Udhayakumar, V., Anyona, D., Kariuki, S., Shi, Y.P., Bloland, P.B., Branch, O.H., Weiss, W., Nahlen, B.L., Kaslow, D.C., Lal, A.A., 1995. Identification of T and B cell epitopes recognized by humans in the C-terminal 42-kDa domain of the *Plasmodium falciparum* merozoite surface protein (MSP)-1. J. Immunol. 154, 6022–6030.

Yang, C., Collins, W.E., Sullivan, J.S., Kaslow, D.C., Xiao, L., Lal, A.A., 1999. Partial protection against *Plasmodium vivax* blood-stage infection in Saimiri monkeys by immunization with a recombinant C-terminal fragment of merozoite surface protein 1 in block copolymer adjuvant. Infect. Immun. 67, 342–349.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Gene 397 (2007) 153–160
GENE
www.elsevier.com/locate/gene

Allelic dimorphism-associated restriction of recombination in *Plasmodium falciparum msp1* ☆

Kazuyuki Tanabe [a,b,*], Naoko Sakihama [a], David Walliker [c], Hamza Babiker [c], Abdel-Muhsin A. Abdel-Muhsin [c], Bernard Bakote'e [d], Hiroshi Ohmae [e], Nobuko Arisue [f], Toshihiro Horii [f], Ingegerd Rooth [g], Anna Färnert [h], Anders Björkman [h], Lisa Ranford-Cartwright [i]

[a] *Laboratory of Malariology, International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan*
[b] *Neo-Morgan Institute, Tokyo, Japan*
[c] *Institute of Immunology and Infection Research, University of Edinburgh, Edinburgh, UK*
[d] *Solomon Islands Medical Training and Research Institute, Honiara, Solomon Islands*
[e] *Department of Parasitology, National Institute of Infectious Diseases, Tokyo, Japan*
[f] *Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan*
[g] *Nyamisati Malaria Research Unit, PO Box, 663, Dar es Salaam, Tanzania*
[h] *Unit of Infectious Diseases, Department of Medicine, Karolinska University Hospital, 17276 Solna, Sweden*
[i] *Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK*

Received 17 November 2006; received in revised form 9 March 2007; accepted 25 April 2007
Available online 10 May 2007
Received by F.G. Alvarez-Valin

Abstract

Allelic dimorphism is a characteristic feature of the *Plasmodium falciparum msp1* gene encoding the merozoite surface protein 1, a strong malaria vaccine candidate. Meiotic recombination is a major mechanism for the generation of *msp1* allelic diversity. Potential recombination sites have previously been mapped to specific regions within *msp1* (a 5′ 1-kb region and a 3′ 0.4-kb region) with no evidence for recombination events in a central 3.5-kb region. However, evidence for the lack of recombination events is circumstantial and inconclusive because the number of *msp1* sequences analysed is limited, and the frequency of recombination events has not been addressed previously in a high transmission area, where the frequency of meiotic recombination is expected to be high. In the present study, we have mapped potential allelic recombination sites in 34 full-length *msp1* sequences, including 24 new sequences, from various geographic origins. We also investigated recombination events in blocks 6 to 16 by population genetic analysis of *P. falciparum* populations in Tanzania, where malaria transmission is intense. The results clearly provide no evidence of recombination events occurring between the two major *msp1* allelic types, K1-type and Mad20-type, in the central region, but do show recombination events occurring throughout the entire gene within sequences of the Mad20-type. Thus, the present study indicates that allelic dimorphism of *msp1* greatly affects inter-allelic recombination events, highlighting a unique feature of allelic diversity of *P. falciparum msp1*.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Malaria; *Plasmodium falciparum*; *msp1*; Merozoite surface protein 1; Recombination; Tanzania; The Solomon Islands

1. Introduction

Malaria is a major health problem in tropical and subtropical countries, where it is responsible for over one million deaths every year (WHO, 2000). The development of malaria vaccines is needed for improved malaria control. The merozoite surface protein 1 (MSP-1) of *Plasmodium falciparum* is a strong malaria

Abbreviations: bp, base pairs; kb, kilobase; MSP-1, merozoite surface protein-1; *msp1*, merozoite surface protein-1 gene.

☆ Sequences obtained in this study have been deposited in DDBJ/EMBL/GenBank with accession numbers: AB276001–AB276018, AB300614 and AB300615.

* Corresponding author. Laboratory of Malariology, International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, 3-1, Yamada-Oka, Suita, Osaka 565-0871, Japan. Tel.: +81 6 6879 4260; fax: +81 6 6879 4262.
E-mail address: kztanabe@biken.osaka-u.ac.jp (K. Tanabe).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.04.033

vaccine candidate (Holder et al., 1999; Mahanty et al., 2003). However, the highly polymorphic nature of the gene encoding MSP-1 (*msp1*) presents a potential obstacle to the development of effective vaccines. As is the case for other *P. falciparum* antigen genes, *msp1* polymorphism is generated via a number of different mechanisms (Tanabe et al., 1987; Miller et al., 1993): point mutations result in single nucleotide polymorphisms (SNPs), insertion/deletion of repeats causes repeat length polymorphisms, and meiotic recombination involving the exchange of gene fragments between parental alleles produces novel alleles in the progeny. SNPs in *msp1* appear to be stable through time (Tanabe et al., 2004a) and may be of ancient origin (Polley et al., 2005). Repeat length polymorphisms are common in *msp1* (Tanabe et al., 1987; Miller et al., 1993; Ferreira et al., 2003) to the extent that size polymorphism between alleles is widely used as a marker for parasite genotyping (Snounou et al., 1999). Aside from repeat length polymorphisms, meiotic recombination is likely to be a major mechanism for the generation of *msp1* allelic diversity (Tanabe et al., 1987). Potential recombination sites have previously been mapped to specific regions within *msp1*: blocks 2 to 6 (a 5′ 1-kb region), and block 17 (a 3′ 0.4-kb region) (see Fig. 1) (Tanabe et al., 1987; Miller et al., 1993; Kerr et al., 1994; Qari et al., 1998). To date, no evidence has been shown for recombination events in blocks 6 to 16, a central 3.5-kb region (Miller et al., 1993; Peterson et al., 1988), which is surprising given the size of this region and the presence of blocks of high sequence conservation between alleles within this region.

However, evidence for the lack of recombination events in the central region is circumstantial because of the limited number of *msp1* sequences analysed ($n=4$) (Miller et al., 1993). Also, support for the lack of recombination events in blocks 6 to 16 by population genetic analysis of field isolates is weak because previous studies have been conducted using parasite populations from low malaria transmission areas such as Thailand and Brazil (Sakihama et al., 1999; Ferreira et al., 2003). The frequency of recombination events in *P. falciparum* is dependent, to a large extent, on the rate of transmission (Babiker and Walliker, 1997; Mu et al., 2005). Meiotic recombination only occurs in the mosquito host following blood feeding on a host infected with genetically distinct parasites of the same species. Such multi-clonal infections are more common in areas of higher endemicity than in areas of lower transmission (Babiker and Walliker, 1997; Tanabe et al., in press). The frequency of recombination events in *msp1* is therefore expected to be high in areas of intense malaria transmission and lower in areas with less intense transmission. Consistently, a high rate of recombination in *msp1* has been inferred in high transmission areas in Africa: for example, linkage disequilibrium between polymorphic sites in *msp1* declined

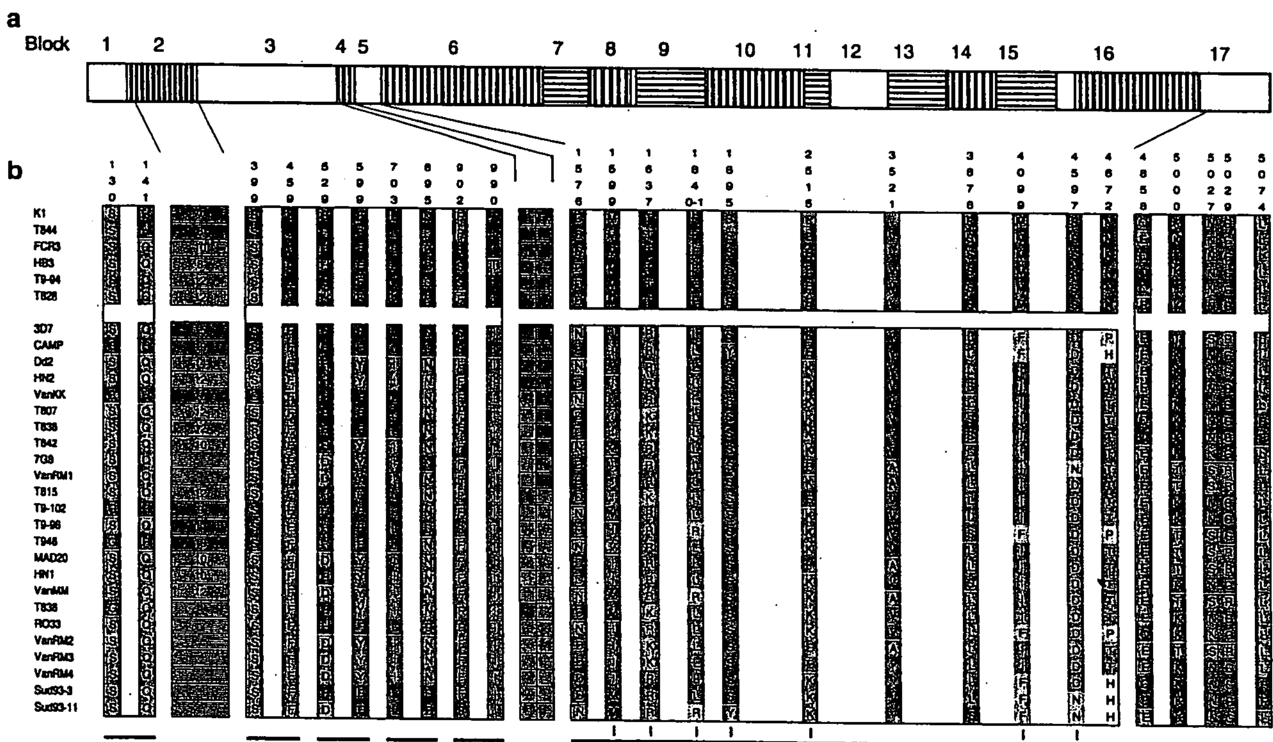

Fig. 1. Potential recombination sites in *Plasmodium falciparum msp1*. (a) The gene structure of *msp1*, showing five conserved blocks (open blocks), seven variable blocks (vertical-striped blocks) and five semi-conserve blocks (horizontal-striped blocks). (b) Potential recombination sites, inferred from the four-gamete test (red line under the sequence alignment). Sequences are divided into K1-type ($n=6$) and Mad20-type ($n=24$), depending on sequence type in blocks 6 to 16. Recombination sites are valid for the two types in blocks 1 and 3, but valid only for Mad20-type in blocks 6 to 16. Note that only amino acid residues that define recombination events (but not all the polymorphic residues) are listed in the alignments. See Supplementary Fig. 1, for other polymorphic sites. The same residues shaded different colours indicate synonymous substitutions at two nucleotide positions: 399 and 990. Positions are numbered after the 3D7 sequence (GenBank accession no. Z35327). Sequences of block 2 were grouped into three allelic types (K1-type in orange, Mad20-type in green, and RO33 type in violet), and the numbers of 9-bp repeats were inserted in each box. Sequences of blocks 4a and 4b were grouped into two allelic types, K1-type (orange) and Mad20-type (green).

rapidly with increasing nucleotide distance over 1 kb (Conway et al., 1999). However, the frequency of recombination events in the central region of *msp1* has not been addressed previously in a high transmission area. Therefore, the suggestion that recombination does not occur in blocks 6 to 16 is based on inconclusive evidence and needs to be studied further. In the present study, we have mapped potential allelic recombination sites in 34 full-length *msp1* sequences, including 24 new sequences, obtained from parasites of wide geographic origins. We also investigated recombination events in blocks 6 to 16 by population genetic analysis of *P. falciparum* populations in Tanzania, where malaria transmission is intense (Babiker and Walliker, 1997; Hay et al., 2000). The results clearly provide no evidence of recombination events in blocks 4 to 16 occurring *between* the two major *msp1* allelic types, K1-type and Mad20-type, but do show recombination events occurring throughout the entire gene *within* sequences of the Mad20-type.

2. Materials and methods

2.1. Parasite population and cultured-parasite lines

P. falciparum isolates were obtained from asymptomatic individuals with ages 1–78 years in cross-sectional surveys of malaria in Nyamisati village in the Rufiji River Delta, 150 km south of Dar es Salaam, in eastern coastal Tanzania in February–March 1993 ($n=94$) and 1998 ($n=102$) (Tanabe et al., in press), Malaria in the study area was holoendemic (Rooth and Björkman, 1992) with perennial transmission. *P. falciparum* isolates were also collected from symptomatic and asymptomatic individuals aged 4–60 years in 1995–1996 ($n=82$) in Guadalcanal Island, the Solomon Island (Sakihama et al., 2006), where the intensity of malaria transmission was very high with reported annual entomological inoculation rate of 584–1022 bites per person per year (Hii et al., 1993), a level comparable to Tanzania. Parasite genomic DNA was extracted as described elsewhere (Sakihama et al., 2001). DNA of isolates from Tanzania and the Solomon Islands was subjected to PCR-based haplotyping for the detection of recombinant types (see below).

Full-length *msp1* sequences were obtained from twenty one culture-adapted *P. falciparum* lines originated from various geographic areas: 3D7 (The Netherlands), HB3 (Honduras), 7G8 (Brazil), RO33 (Ghana), FCR3 (Gambia), K1, T9/94, T9/96, T9/102, T806, T807, T815, T828, T836, T837, T838, T842, T844, T946 (Thailand), Dd2 (Indochina), Mad20 (Papua New Guinea) (Jongwutiwes et al., 1991; Tanabe et al., 2004b). Additional three field isolates, Sudan 60-93-3, Sudan 60-93-11 and Sudan 121-93-12 were collected in Asar village in the eastern Sudan in 1993–1994 (Babiker et al., 1998). Sequences obtained in this study have been deposited in DDBJ/EMBL/GenBank with accession numbers: AB276001–AB276018, AB300614 and AB300615. Previously reported sequences were also included for comparison: they were VanKK (AB116596), VanMM (AB116597), VanRM1 (AB116598), VanRM2 (AB116600), VanRM3 (AB116599), and VanRM4 (AB116601) from Vanuatu; HN1 (AF062348) and HN2 (AF062349) from Hainan Island, China; CAMP (X03831) (Malay); RO33 (Y00087, Z35326); 3D7 (Z35327); K1 (X03371); Wellcome (X02919) (unknown origin); and Mad20 (X05624).

2.2. Detection of recombinants

Potential recombination events in blocks 2 to 6 (a 5′ 1-kb region) and blocks 6 to 16 (a central 3.5-kb region) were investigated by PCR using allelic type-specific primers (see Fig. 1a for the location of blocks). PCR conditions have been described previously (Sakihama et al., 2006). One μl of template DNA was used for amplifications of blocks 2 to 6 and blocks 6 to 16 using allelic type specific primers. Amplification of blocks 2 to 6 was performed using forward primers, k2F2, m2FN and r2F, which are specific to K1-type, Mad20-type and RO33-type, respectively, in block 2 and reverse primers, k6R6 and m6R, specific to K1-type and Mad20-type, respectively, in block 6. Thus, six distinct haplotypes are detected in blocks 2 to 6: (i) K1-type in block 2 and K1-type in block 6, designated here as K2K6, (ii) M2MK6, (iii) R2K6, (iv) K2M6, (v) M2M6, and (vi) R2M6. Amplification of blocks 6 to 16 was done using forward primers, k6F and m6F, specific to K1-type and to Mad20-type in block 6, respectively, and k16R and m16R3, specific to K1-type and Mad20-type in block 16, respectively (Tanabe et al., 2002). Combinations of k6F and k16R and of m6F and m16R3 amplify K1-type and Mad20-type sequences in blocks 6 to 16 (i.e., K6K16 and M6M16), and those of k6F and m16R3 and of m6F and k16R yield K1/Mad20-type and Mad20/K1-type recombinant sequences (K6M16 and M6K16), respectively. Primer sequences are listed in Supplementary Table 1 online. The frequency of recombination events in blocks 2 to 6 and blocks 6 to 16 was estimated by the χ^2 test for observed and expected numbers of haplotypes, in which expected number of haplotypes was calculated based on observed frequency of polymorphic blocks, assuming no linkage disequilibrium between polymorphic blocks.

2.3. Sequencing of msp1

A DNA fragment covering the entire coding region of *msp1* was amplified in a 20-μl reaction mixture containing the following: 0.2 mM each of forward and reverse primers, UPF1 (5′-GGCTAATGTAAAATGCAAAAATAAATGT-3′) and DWR1 (5′-ACATGACTAAAATATCACTATTCCTGT-3′), 400 mM each of dNTP, 1 unit of Takara LA Taq (Takara Shuzo, Japan), 2 μl of 10× LA PCR Buffer II, 2.5 mM of $MgCl_2$, and 1 μl of genomic DNA template. Forty cycles of amplification (20 s at 93 °C and 5 min at 62 °C) were preceded by denaturation at 93 °C for 1 min and followed by final elongation at 72 °C for 10 min. PCR product obtained was diluted 10-fold and a 2-μl aliquot was used as template for 20-cycle nested PCR amplification in a 50-μl reaction mixture using primers UPF3 (5′-AATAAATGTATACA-TATTTTTGCTAAGTCA-3′) and DWR3 (5′-TTAAGGTAA-CATATTTTAACTCCTACA-3′). The final concentration of each component of the nested PCR mixture other than DNA template was identical with that in the first PCR mixture. The nested PCR product was purified using QIAquick PCR purification kit (QIAGEN, Germany) and sequenced in both directions using BigDye Terminator Cycle Sequencing Kit (ver

3.1) (Applied Biosystems, Foster City, CA) in an ABI 3100 Genetic Analyzer (Applied Biosystems, CA). Sequencing primers used in this study are in Supplementary Table 2 online. Sequences were verified by re-sequencing of a DNA template independently amplified from the same genomic DNA. The rate of sequence error of this sequencing protocol was 6×10^{-9} per site, calculated by a frequency of errors after sequencing of three independent PCR products from the original DNA templates of known *msp1* sequences, K1 allele and Mad20 allele. Such a low error rate indicates that the *msp1* sequences obtained in this study are unlikely to have errors because the total sequence length obtained here was approximately 1.2×10^5 kb.

2.4. Estimation of recombination sites

Nucleotide sequences were aligned using CLUSTAL X (Thompson et al., 1997) with manual corrections, and regions of highly variable sequences (blocks 2 and 4) and gaps were excluded from the analysis (Supplementary Fig. 1). The *msp1* sequences obtained were compared to published sequences and complete identity to that published was found for the 3D7, K1 and Mad20 sequences. However, numerous differences were found in the RO33 sequences. The RO33 sequence obtained here was not so distinctive as the Sud60-93-3 sequence. This was particularly evident in block 4, in which our sequence is very similar to that seen in the Mad20 allele (see Fig. 2). The RO33 sequence obtained in this study was used for further analysis. The minimum number of recombination events (Rm) and the likely positions of recombination events were estimated by the four-gamete test (Hudson and Kaplan, 1985), implemented in DnaSP ver 4.0 (Rozas et al., 2003) using all 34 full-length MSP1 sequences. Nucleotide diversity (π) was calculated using MEGA ver 3.1 (Kumar et al., 2004) with correction by the Kimura's 2-parameter method (Kimura, 1980).

3. Results and discussion

3.1. Detection of msp1 recombinant haplotypes

PCR-based haplotyping of *msp1* blocks 2 to 6 identified all six haplotypes in the 196 Tanzanian isolates collected in 1993 and 1998: K2K6, M2K6, R2K6, K2M6, M2M6, and R2M6 (Table 1), in which the numbers of haplotypes identified were greater than the number of samples because of multiple haplotype infections per sample. Relatively fewer haplotypes were observed with the K1 allelic type in block 6 compared with those having the Mad20 allelic type in block 6. Nevertheless, all possible recombinant types in blocks 2 to 6 were observed in the parasite populations collected in both 1993 and 1998. The frequency distribution of the observed numbers of haplotypes in 1998 did not differ significantly from expected numbers of haplotype, suggesting frequent recombination in blocks 2 to 6. In contrast, in blocks 6 to 16, only two haplotypes, K6K16 and M6M16, were observed, with no recombinant haplotypes (K6M16 and M6K16), indicating the absence of recombination events in blocks 6 to 16 within the same sample set. We also conducted PCR-based haplotyping on isolates from the Solomon Islands ($n=82$), where the prevalence of K1-type in block 6 was much higher than in Tanzania (Sakihama et al., 2006). We identified only two haplotypes in blocks 6 to 16 (K6K16 and M6M16), whereas five haplotypes were detected in blocks 2 to 6 (K2K6, M2K6, K2M6, M2M6, and R2M6), thus confirming the absence of recombination events in blocks 6 to 16.

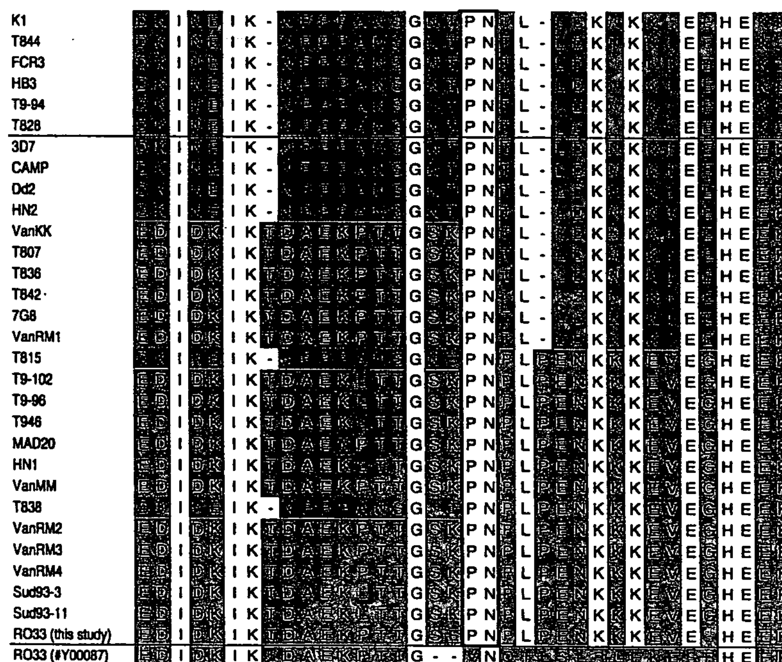

Fig. 2. A potential recombination site in *P. falciparum msp1* block 4. The site is boxed. The thirty *msp1* alleles shown in the left column were aligned manually, together with the reported sequence of the RO33 allele (Y00087).