Guadalcanal (39%) [16]. Transmission intensity in Mae Sod, northwestern Thailand is extremely low, as indicated by an estimated "incidence" of 0.8 infections per person per year (in 1991) based on dispensary attendance in a refugee camp located in this area [28].

2.2. Determination of msp1 haplotypes

msp1 haplotypes are defined as unique combinations of 5′ recombinant types and 3′ sequence types (Fig. 2). 5′ recombinant types are defined as unique combinations of allelic types of blocks 2, 4a, 4b, and 6, which comprise a 1.1 kb region in the 5′ region of *msp1*. Our typing method involves nested PCR, and can distinguish 24 distinct 5′ recombinant types generated by intragenic recombination between *msp1* alleles [11]. Protocols for determination of 5′ recombinant types by PCR using allelic-type-specific primers have been described in previous reports [16]. Preliminary study revealed that there was no significant difference in frequency distribution of 5′ recombinant types between in-patients (n=32) and out-patients (n=65) in Palawan (P=0.927). (Similarly, there was no significant difference in frequency between individuals with clinical malaria and those with asymptomatic malaria in the Solomon Islands [16]). Although the rate of multiple infections of 5′ recombinant type and the multiplicity of infection were slightly higher in in-patients than in out-patients (rate of multiple infections, 34.4% *versus* 32.3%; multiplicity of infection, 1.5 *versus* 1.42), these differences were not significant. Therefore, the data for in-patients and out-patients were combined for further analysis.

The 3′ sequence types are defined as unique combinations of single-nucleotide polymorphism in block 17, which is a 0.4-kb region in the 3′ end of *msp1*. The sequence of block 17 was determined by direct sequencing of isolates that harbored a single 5′ recombinant type. A DNA fragment covering the entire coding region of *msp1* was amplified using Takara LA Taq (Takara Shuzo, Japan) in a 20-μl reaction mixture for 40 cycles essentially as described elsewhere [23,24], except that the primers used were UPF1 (5′-GGCTAATGTAAAATGCAAAAATAAATGT-3′) and DWR1 (5′-ACATGACTAAAATATCACTATTCCTGT-3′). The PCR product was diluted 10-fold, and a 2-μl aliquot was used as the template for nested PCR amplification for 20 cycles in a 50-μl reaction mixture using primers UPF3 (5′-AATAAATGTATACATATTTTTGCTAAGTCA-3′) and DWR3 (5′-TTAAGGTAACATATTTTAACTCCTACA-3′). The nested PCR product was purified using the QIAquick PCR purification kit (QIAGEN, Germany), and was sequenced from both directions using the BigDye Terminator Cycle Sequencing Kit (ver 3.1) and an ABI 3100 sequencer. The sequencing primers were C17aFs [11] and DWR3. Sequences were verified by re-sequencing a DNA template independently amplified from genomic DNA. Five amino acid substitution sites have previously been identified in block 17 in field isolates from various geographic areas [15,29] (E or Q at residue 1644; T or N at 1691; SR or NG at 1700–1701; and L or F at 1716 [8]). In addition, a new substitution from L to I at residue 1740 was identified in the present study in two isolates from Palawan Island (see Results). We hereafter designate unique combinations of these six residues as 3′ sequence type.

2.3. Genotyping of pfcrt

To assess the prevalence of polymorphism of the *P. falciparum* chloroquine resistance transporter gene (*pfcrt*) in Palawan Island, we sequenced the 2nd exon of *pfcrt*, which contains a polymorphic region at residues 72 to 76. Isolates harboring a single 5′ recombinant type (n=59) were subjected to direct sequencing. Simultaneously, 71 Thai isolates were sequenced for comparison. Amplification of the 2nd exon and sequencing were performed as previously described [16], except that the dRhodamine Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) was used. For those samples containing a mixture of both CQ-sensitive and CQ-resistant *pfcrt* alleles, DNA fragments carrying CQ-sensitive and CQ-resistant alleles were separately amplified by semi-nested PCR using two reverse primers (CQ76*Kr* [5′-TTTGTTTAAAGTTCTTTTAGCAAAA-A*TTT*-3′], specific to the CQ-sensitive allele; and CQ76*Tr* [5′-TTTGTTTAAAGTTCTTTTAGCAAAAA*TTG*-3′], specific to the CQ-resistant allele), in combination with the forward primer CQf4 [16]. In this case, the 1st PCR product, diluted 100-fold, was subjected to nested PCR using AmpliTaq Gold (Applied Biosystems, Foster City, CA) for 12 cycles.

2.4. Statistical analysis

Frequency distributions of *msp1* 5′ recombinant types and *msp1* haplotypes were compared between different parasite populations using the two-tailed χ^2 test, with Yates' correction for data sets of fewer than 10 items, and Fisher's exact probability test. Pairs of polymorphic blocks 2, 4a, 4b and 6 and polymorphic sites in block 17 were subjected to the R^2 test to assess linkage disequilibrium (LD), as described elsewhere [11]. To avoid false detection of LD deriving from inclusions of minor alleles, non-informative pairs (frequency of minor alleles <10% in a polymorphic block or nucleotide site) were excluded from the R^2 test. Significance of LD was assessed using the χ^2 test and the two-tailed Fisher's exact probability test. The haplotype diversity index (h) and its standard error (SE) were calculated as described elsewhere [16]. Probability values of P<0.05 were considered to indicate statistical significance.

3. Results

3.1. Diversity and multiplicity of 5′ recombinant types (blocks 2 to 6)

Table 1 summarizes diversity profiles of 5′ recombinant types in *P. falciparum* isolates from Palawan, in comparison with those from the Solomon Islands and Thailand. The number of distinct 5′ recombinant types in Palawan (n=9) was close to the number in the Solomon Islands (n=8), but was much lower than the number in Thailand (n=14). The rate of multiple infections of 5′ recombinant types (33%) and multiplicity of infection (*i.e.*, mean number of 5′ recombinant types per isolate) (1.44) in Palawan were also similar to those of the Solomon Islands, but were substantially lower than those of Thailand. These results indicate that the diversity and extent of mixed

Table 1
Diversity profiles of *P. falciparum* isolates in Palawan, the Philippines in comparison with the Solomon Islands and Thailand

	Palawan	Solomon Islands[a]	Thailand[a]
Number of isolates	114	90	111
Number of PCR positives	97 (85.1%)	82 (91.1%)	107 (96.4%)
Number of 5′ recombinant types/area	9	8	14
Rate of multiple infections	32/97 (33.0%)	29/82 (35.4%)	103/107 (96.3%)
Multiplicity of infection[b]	1.44	1.41	3.6

[a] Data from Sakihama et al. [16].
[b] Mean number of *msp1* 5′ recombinant types per isolate.

infections of 5′ recombinant types are relatively limited in Palawan and the Solomon Islands.

Frequency distribution of 5′ recombinant types greatly differed among the three geographic areas (Fig. 3) ($P<10^{-7}$ for all pairs compared). The observed differences are due to different frequencies of 5′ recombinant types #2, #16, #21, #22, #23 and #24.

3.2. The 3′ sequence types (block 17)

A total of 57 Palawan isolates were successfully sequenced, and six distinct 3′ sequence types were detected (Fig. 4): E/TSR/L ($n=17$), Q/KNG/L ($n=18$), E/KNG/L ($n=3$), Q/TSR/L ($n=3$), Q/KNG/F ($n=14$), and E/TSR/L* ($n=2$). E/TSR/L* is a new type containing a previously unknown mutation from L (codon: TTA) to I (ATA) at amino acid position 1740 (Fig. 4), which is downstream from a putative post-translational cleavage site between residues 1726 and 1727 [30].

3.3. Diversity of msp1 haplotypes

Table 2 summarizes the distribution of *msp1* haplotypes, unique combinations of 5′ recombinant type and 3′ sequence type in Palawan, in comparison with those of the Solomon Islands and Thailand. The number of *msp1* haplotypes detected in Palawan ($n=15$) was comparable to the number in Thailand ($n=16$), but was much greater than the number in the Solomon Islands ($n=8$). Consistent with this, haplotype diversity (h) in Palawan (0.88) was comparable to that of Thailand (0.89) ($P=0.66$), but was greater than that of the Solomon Islands (0.80) ($P=0.033$).

We observed marked difference in the distribution of *msp1* haplotypes among Palawan, the Solomon Islands and Thailand ($P<10^{-5}$) (Fig. 5). Of the 15 distinct *msp1* haplotypes found in Palawan, seven were unique to Palawan, of which MMMM-Q/KNG/F had the highest frequency. Three haplotypes (KKKK-Q/KNG/L, MMKM-E/TSR/L, and MMMM-E/TSR/L) in Palawan were shared with the Solomon Islands and Thailand. Five haplotypes were shared with the Solomon Islands only, and five haplotypes were shared with Thailand only. These prevalence patterns of *msp1* haplotypes suggest that *msp1* haplotypes are geographically stable in the three regions studied.

3.4. Linkage disequilibrium in msp1

Based on the frequency distributions of *msp1* haplotypes summarized in Table 2, we analyzed linkage disequilibrium

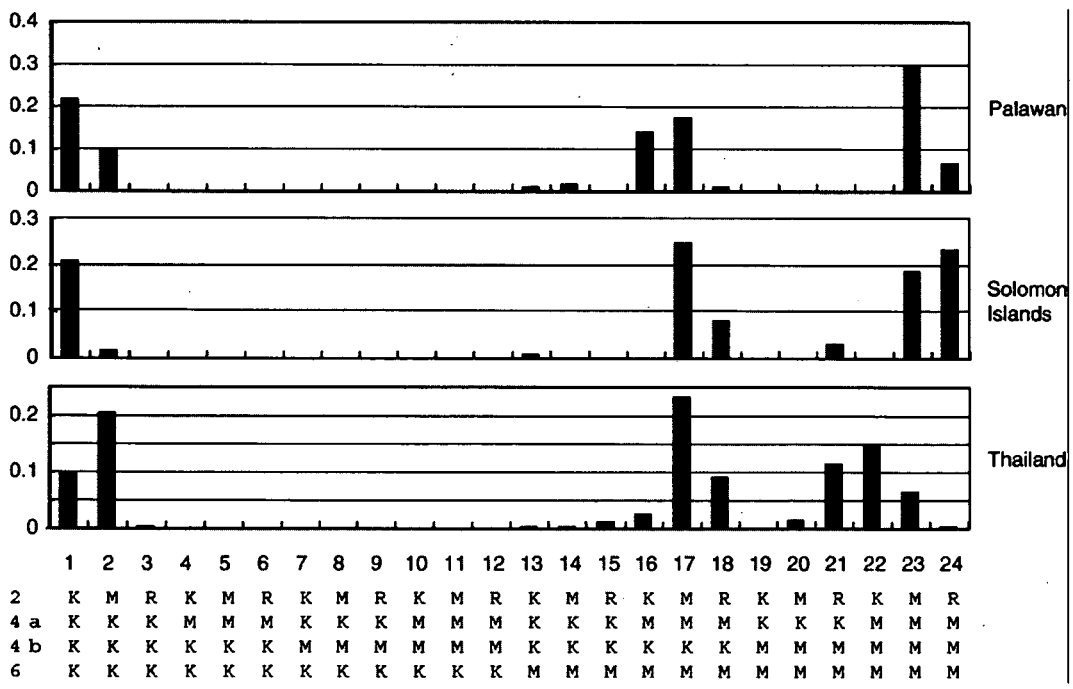

Fig. 3. Frequency distribution of *P. falciparum msp1* 5′ recombinant types in isolates from Palawan, Solomon Islands, and Thailand. Twenty-four potential 5′ recombinant types, defined as unique combinations of allelic types in blocks 2, 4a, 4b, and 6, are shown at the bottom, in which K, M and R are abbreviations of three allelic types: K1, MAD20, and RO33, respectively.

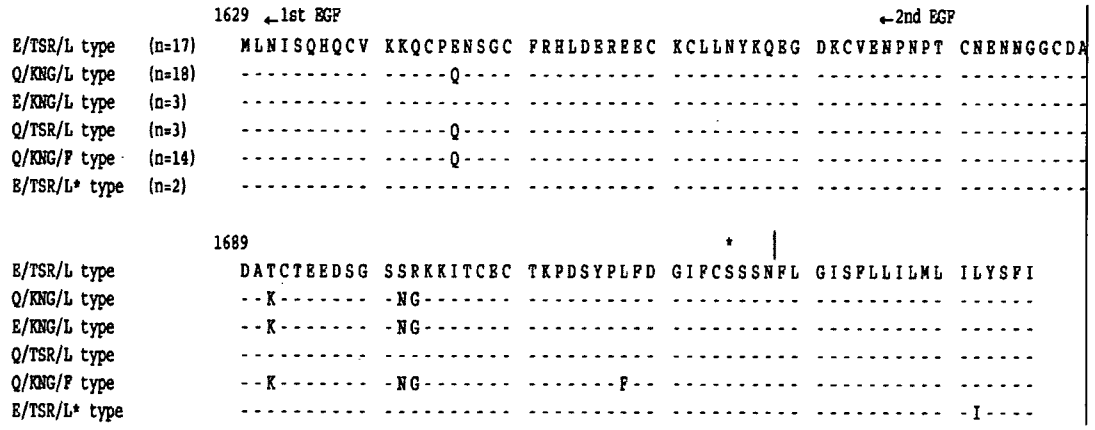

Fig. 4. Amino acid sequence alignment of *P. falciparum msp1* block 17 in isolates from Palawan, Philippines. Arrows indicate the starting points of two epidermal growth factor (EGF)-like domains, and an asterisk indicates a putative attachment site for a glycosylphophatidylinositol (GPI) anchor. A vertical line indicates a putative post-translational cleavage site between residues 1726 and 1727 [30].

(LD) between polymorphic blocks and polymorphic sites in block 17 (Fig. 6). Of the 21 informative pairs within and between polymorphic blocks/sites in Palawan, four pairs (between block 2 and two polymorphic sites in block 17) were not significant. Overall, LD in Palawan was very strong, and was comparable to that of the Solomon Islands (where one out of 15 pairs was not significant), but was stronger than that of Thailand (where seven out of 15 pairs were not significant). These findings suggest that the frequency of recombination events in *msp1* is relatively low in Palawan and the Solomon Islands, compared with Thailand.

3.5. Polymorphism in pfcrt

Of the 59 Palawan isolates sequenced, 33 were mono-infections with either the CQ-sensitive *pfcrt* allele ($n=7$) or the CQ-resistant *pfcrt* allele ($n=26$), and 26 had a mixture of both CQ-sensitive and CQ-resistant alleles. The CQ-resistant *pfcrt*

Table 2
Diversity of *P. falciparum msp1* haplotypes in isolates from Palawan, the Philippines incomparison with the Solomon Islands and Thailand

Location	5′ recombinant type	3′ sequence type						No. of *msp1* haplotype[a]	*h*±SE
		Q/KNG/L	Q/KNG/F	E/KNG/L	Q/TSR/L	E/TSR/L	(Total)		
Palawan	KKKK	9	0	0	0	0	(9)		
	MKKK	5	0	0	0	0	(5)		
	MKKM	1	0	0	0	0	(1)		
	KMKM	1	0	3	1	5	(10)		
	MMKM	1	0	0	0	9[b]	(10)		
	MMMM	1	14	0	1	5	(21)		
	RMMM	0	0	0	1	0	(1)		
	(Total)	(18)	(14)	(3)	(3)	(19)	(57)	15	0.88±0.02
Solomon Islands	KKKK	9	0	0	0	0	(9)		
	MMKM	0	1	0	0	13	(14)		
	RMKM	0	0	0	0	1	(1)		
	MMMM	1	0	5	0	3	(9)		
	RMMM	0	0	0	0	14	(14)		
	(Total)	(10)	(1)	(5)	(0)	(31)	(47)	8	0.80±0.03
Thailand	KKKK	2	0	0	0	0	(2)		
	MKKK	9	0	4	0	0	(13)		
	KMKM	0	0	0	0	1	(1)		
	MMKM	0	2	12	0	2	(16)		
	RMKM	0	0	3	0	0	(3)		
	RKMM	0	0	3	0	1[c]	(4)		
	KMMM	1	1	2	0	3	(7)		
	MMMM	0	0	1	0	1	(2)		
	(Total)	(12)	(3)	(25)	(0)	(8)	(48)	16	0.89±0.03

[a] *msp1* haplotypes are defined as unique associations of 5′recombinant type and 3′ sequence type.
[b] Includes two isolates with a new amino acid substitution at residue 1740.
[c] One isolate with E/TSG/L type.

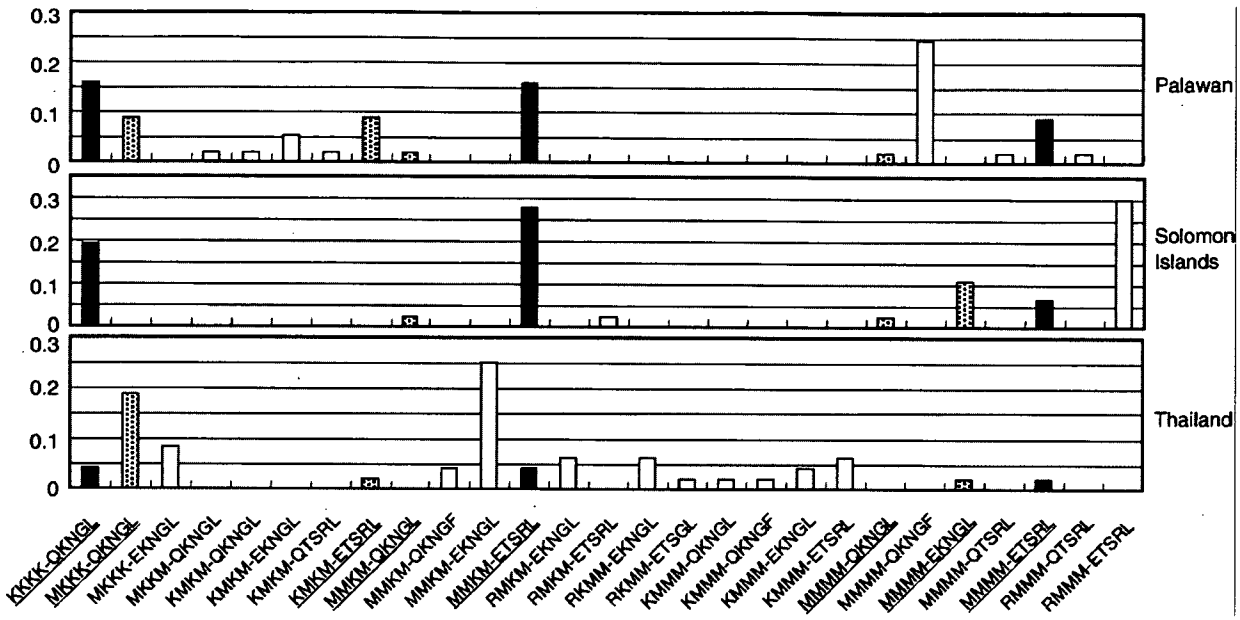

Fig. 5. Frequency distribution of *P. falciparum msp1* haplotypes in Palawan, the Solomon Islands, and Thailand. *msp1* haplotypes, defined as unique combinations of 5′ recombinant types and 3′ sequence types, are indicated along the horizontal axis. MMKM-E/TSR/L type includes two isolates with a new amino acid substitution at residue 1740 (see Fig. 4). Closed column, haplotype detected in all three areas; half-tone column, haplotype detected in two of the three areas; open column, haplotype detected in only one area.

allele was of Papua New Guinean type, with the amino acid sequence SVMNT at residues 72 to 76 [31]. We did not detect the Southeast Asian type with the amino acid sequence CVIET at residues 72 to 76 [32]. In contrast, all of the Thai isolates (n=71) had the CVIET-type *pfcrt* allele. Strikingly, the wild-type CQ-sensitive *pfcrt* allele, with the amino acid sequence CVMNK at residues 72 to 76, was detected in 33 Palawan isolates, with a prevalence rate of 38.8%. This is in sharp contrast to the monomorphic prevalence of the Papua New Guinean type CQ-resistant *pfcrt* allele in the Solomon Islands (n=57) and Vanuatu (n=142) [16,33]. There was no significant difference in the distribution of *msp1* haplotypes between isolates harboring the CQ-sensitive *pfcrt* allele and those with the CQ-resistant *pfcrt* allele (data not shown), suggesting that there is no genetic linkage between the gene *msp1* in chromosome 9 and the gene *pfcrt* in chromosome 7.

4. Discussion

It is to be expected that the extent of allelic diversity of *P. falciparum msp1* would be low in a low-transmission area such as Palawan Island, because *msp1* allelic diversity is mainly generated by meiotic recombination events that takes place solely in the mosquito stage of the *P. falciparum* life cycle. In the present study, the frequency of recombination events in *msp1*, as inferred from LD, was relatively low on Palawan Island. However, the extent of *msp1* allelic diversity on Palawan Island was higher than that of the Solomon Islands, where malaria transmission is very high. These findings support our hypothesis that the extent of *msp1* allelic diversity is not always dependent on the intensity of malaria transmission. This hypothesis was originally raised in our previous study of parasite populations from the Solomon Islands, in which *msp1* allelic diversity was very limited despite a high level of malaria transmission. In contrast, the present study of the Palawan

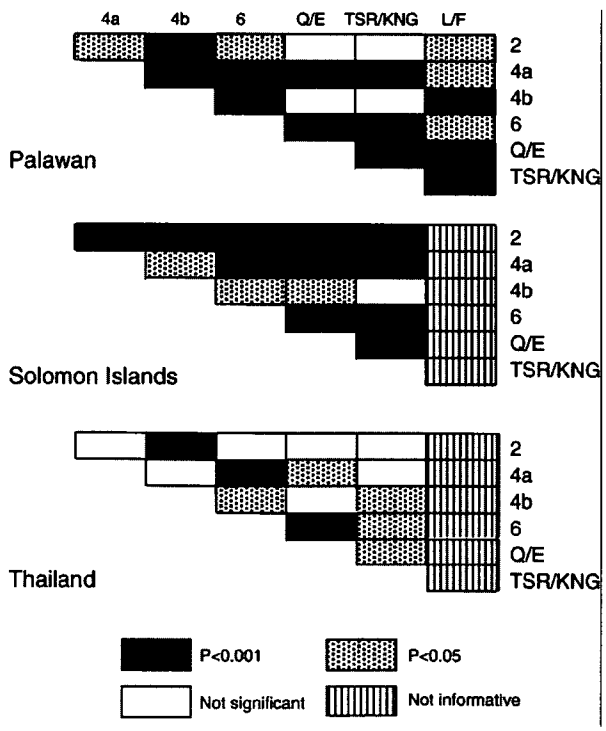

Fig. 6. Linkage disequilibrium in *P. falciparum* merozoite *msp1* in populations from Palawan, the Solomon Islands, and Thailand. Pairs of polymorphic blocks (2, 4a, 4b, and 6) and three polymorphic residue sites (Q/E, TSR/KNG, and L/F) in block 17 were subjected to the R^2 test.

population shows the opposite situation: *i.e.,* relatively high *msp1* allelic diversity with a low level of transmission. Therefore, it is likely that different mechanisms are involved in the lack of association between the extent of *msp1* allelic diversity and the intensity of malaria transmission.

The frequency of recombination events in *msp1* is determined not only by the frequency of malaria transmission but also by other variables: the frequency of mixed-allele infections, the mean number of alleles per person, and the repertoire of alleles in a local area [12]. Mixed-allele infections are required for recombination events to occur between heterologous *msp1* alleles. Therefore, those variables would affect the rate of inbreeding, and a high inbreeding rate could lead to limited recombination events *de facto* in spite of extensive genetic shuffling during meiosis in the mosquito stage in a high transmission area. In the parasite populations from the Solomon Islands studied previously, values of those variables are rather low, suggesting a high inbreeding rate, which in turn can limit recombination events (and subsequently lower the allelic diversity of *msp1*). In the present study, the Palawan parasite population had low rates of mixed 5′ recombinant type infections and multiplicity (the mean number of 5′ recombinant types per person), with levels similar to those of the Solomon Islands. However, diversity profiles, the number of *msp1* haplotypes, and haplotype diversity (*h*) were higher in Palawan than in the Solomon Islands. Therefore, the observed relatively high allelic diversity of *msp1* in Palawan may be due to a mechanism not directly related to the frequency of malaria transmission.

Then, a question arises: why *msp1* allelic diversity is relatively high in Palawan, where recombination events are limited? We consider that the high prevalence of CQ-resistant *pfcrt* allele observed in Palawan may be partly involved in the *msp1* diversity. It is generally believed that the parasites having CQ-resistant *pfcrt* allele does not necessarily mean that they are resistant to CQ, and that first line treatment of CQ might have minor influence on parasite transmission by asymptomatic individuals, where parasites are known to persist for long periods. These arguments may be valid in the situation in highly endemic areas of malaria such as African tropics, where repeated malaria infections and asymptomatic cases are common and CQ is still effective in asymptomatic individuals harboring CQ-resistant *pfcrt* allele, who are immune or semi-immune to malaria due to synergistic effect of immunity and the drug. However, these arguments do not always apply to geographic areas other than African tropics, particularly to islands, where *P. falciparum* populations are isolated and the influx of new genotypes from outside is very limited. In addition to this limited gene flow, malaria-immune asymptomatic cases are rare in areas with low or moderate malaria transmission such as Palawan Island and northwestern Thailand. (Asymptomatic cases are exceptionally not rare in the Solomon Islands, where malaria transmission is intense). Taking these malaria epidemiological settings into considerations, it is likely that frequent and extensive mass administrations of CQ on islands as done in the Solomon Islands would ere bottleneck of local parasite populations and duction in the parasite genotypes. Therefore, the of CQ-resistant *pfcrt* allele would reflect selective pressure of CQ in the past and reduction in parasite genotypes including *msp1* alleles. Conversely, a high prevalence of CQ-sensitive *pfcrt* allele would reflect less intense selective pressure, which is not so strong to cause population bottleneck and reduction in the diversity of *msp1* alleles. We observed here that the wild-type CQ-sensitive *pfcrt* allele remained at a substantially high level (39%) in Palawan. No wild-type *pfcrt* allele was found in the Solomon Islands, where nation-wide mass administration of CQ was implemented in the 1980s (in Guadalcanal, 1984–1987) [34]. The population bottleneck caused by mass CQ administration in the Solomon Islands would presumably have been intense, and thereby caused a marked reduction in the diversity of *P. falciparum* genotypes in the area. Consistently, we observed a low level of *msp1* allelic diversity. Additionally, continued usage of CQ as the first-line drug, while it is still effective against *P. vivax*, might have contributed to persistence of a low level of parasite genetic diversity. On the other hand, no such mass drug administration has ever been performed in Palawan. Thus, persistence of a high prevalence of CQ-sensitive parasites may have contributed to the relatively high allelic diversity of *msp1* observed in Palawan.

In contrast to malaria epidemiological settings on islands, monomorphic prevalence of a CQ-resistant *pfcrt* allele did not apparently limit the genetic diversity of *msp1* in the Thai parasite population, a situation opposite to on islands. We consider a likely explanation as follows: the Thai parasite population is not isolated within a limited land area and subject to recurrent gene flows of novel genotypes from outside, most probably from Myanmar where malaria transmission is higher than in Thailand. Multiplicity of infection was very high in northwestern Thailand (3.6), whereas the intensity of malaria transmission was extremely low with expected EIR of 0.8 [28]. Elevated levels of mixed genotype infections may well increase the frequency of recombination events of *msp1* alleles in the mosquito and subsequently a high diversity of *msp1* alleles would be maintained in the Thai population. Rohr et al. have recently demonstrated that, in the Thai–Myanmar border area, the majority of multiple or polyclonal infections with *P. falciparum* are generated by the bite of a single mosquito infected with multiple parasite genotypes [Rohr C, Singlam S, Pumpieng K, Nosten F, Anderson TJC. Kinship relationships in polyclonal malaria infections. In: Abstracts of the 11th International Congress of Parasitology (Glasgow, Scotland 6–11th Aug 2006). Abstract No. a1960], consistent with the high multiplicity of infection frequently reported in this area despite of an extremely low transmission intensity [35].

The present study provides the first molecular evidence that the prevalence of CQ resistance in *P. falciparum* in Palawan did not reach 100% in 1997. CQ resistance in Palawan was first reported by Baird et al. [19], who conducted a field survey in April to May 1995 and found that 23 to 39% of *P. falciparum* isolates were CQ-resistant in an *in vitro* test. In addition, Bustos et al. [36,37] observed a cure rate of only 30.4% (7/23) in the control CQ monotherapy group during a clinical trial of antimalarial drug therapy conducted at Palawan Provincial Hospital between October 1994 and February 1995. These reports suggest that

CQ-sensitive and CQ-resistant parasite populations co-existed in Palawan in the mid-1990s. Here, we have confirmed this co-existence in 1997, with a prevalence rate of 39% for the CQ-sensitive *pfcrt* allele. Identification of the Papua New Guinea-type *pfcrt* allele in Palawan as the sole CQ-resistant *pfcrt* allele is consistent with a recent report on the distribution of *pfcrt* alleles in the Philippines: all twelve *P. falciparum* isolates collected at Malaria Control Service Palawan Office between 1997 and 1999 (four isolates per year) had the CQ-resistant *pfcrt* allele of Papua New Guinea type, and the Southeast Asia type CQ-resistant *pfcrt* allele was only detected in the vicinity of a refugee processing center that accommodated refugees from Southeast Asia at the time of sample collection (between 1989 and 1993) [38]. Further study is required to monitor subsequent change in the prevalence of CQ resistance in *P. falciparum* in Palawan (*e.g.*, whether the prevalence of CQ resistance has since reached 100%) and elucidate how this change has influenced recombination-based diversity of the *msp1* gene. Follow-up monitoring of the prevalence of CQ resistance will also contribute to optimization of drug-based malaria control in Palawan.

In summary, the present study shows relatively high allelic diversity of *msp1* in Palawan, compared with the Solomon Islands, whereas the frequency of recombination events in *msp1* (as determined by LD analysis) was limited, presenting additional evidence for our hypothesis that allelic diversity of *msp1* is not always dependent on the intensity of malaria transmission. Persistence of a substantially high level of the wild-type CQ-sensitive *pfcrt* allele may have contributed to the higher allelic diversity of *msp1* in Palawan, where the parasite population is isolated and gene flow is limited.

Acknowledgements

We wish to thank staff members of the Public Health Laboratory of Palawan Provincial Hospital and Clinical Laboratory of Malaria Control Service Palawan Office, Puerto Princesa, Palawan, the Philippines, for performing sample collection. We also wish to thank all blood donors who participated in this study. This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (18073013, 18GS03140013), Japan Society for Promotion of Science (18390131), and the Japanese Ministry of Health, Labor and Welfare (H17-Sinkou-ippan-019).

References

[1] Conway DJ, Cavanagh DR, Tanabe K, Roper C, Mikes ZS, Sakihama N, et al. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nat Med 2000;6:689–92.

[2] Woehlbier U, Epp C, Kauth CW, Lutz R, Long CA, Coulibaly B, et al. Analysis of antibodies directed against merozoite surface protein 1 of the human malaria parasite *Plasmodium falciparum*. Infect Immun 2006;74:1313–22.

[3] Holder AA, Guevara Patino JA, Uthaipibull C, Syed SE, Ling IT, Scott-Finnigan T, et al. Merozoite surface protein 1, immune evasion, and vaccine against asexual blood stage malaria. Parasitologia 1999;41:409–14.

[4] Renia L, Ling IT, Marussig M, Miltgen F, Holder AA, Mazier D. Immunization with a recombinant C-terminal fragment of *Plasmodium yoelii* merozoite surface protein 1 protects mice against homologous but not heterologous *P. yoelii* sporozoite challenge. Infect Immun 1997;65:4419–23.

[5] Martinelli A, Cheesman S, Hunt P, Culleton R, Raza A, Mackinnon M, et al. A genetic approach to the *de novo* identification of targets of strain specific immunity in malaria parasites. Proc Natl Acad Sci U S A 2005;102:814–9.

[6] Cheesman S, Raza A, Carter R. Mixed strain infections and strain-specific protective immunity in the rodent malaria parasite *Plasmodium chabaudi chabaudi* in mice. Infect Immun 2006;74:2996–3001.

[7] Tanabe K, Mackay M, Goman M, Scaife JG. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J Mol Biol 1987;195:273–87.

[8] Miller LH, Roberts T, Shahabuddin M, McCutchan TF. Analysis of genetic diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol Biochem Parasitol 1993;59:1–14.

[9] Conway DJ, Roper C, Odoula AM, Arnot DE, Kremsner PG, Grobusch MP, et al. High recombination rate in natural populations of *Plasmodium farciparum*. Proc Natl Acad Sci U S A 1999;96:4506–11.

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

[11] Sakihama N, Kimura M, Hirayama K, Kanda T, Na-Bangchang K, Jongwutiwes S, et al. Allelic recombination and linkage disequilibrium within Msp1 of *Plasmodium falciparum*, the malignant human malaria parasite. Gene 1999;230:47–54.

[12] Sakihama N, Kaneko A, Hattori T, Tanabe K. Limited recombination events in merozoite surface protein-1 alleles of *Plasmodium falciparum* on islands. Gene 2001;279:41–8.

[13] Nei M. Molecular Evolutionary Genetics. New York: Colombia University Press; 1987.

[14] Tanabe K, Sakihama N, Rooth I, Björkman A, Färnert A. High frequency of recombination-driven allelic diversity and temporal variation of *Plasmodium falciparum msp1* in Tanzania. Am J Trop Med Hyg in press.

[15] Da Silveira LA, Ribeiro WL, Kirchgatter K, Wunderlich G, Matsuoka H, Tanabe K, et al. Sequence diversity and linkage disequilibrium within the merozoite surface protein-1 (MSP1) locus of *Plasmodium falciparum*: A longitudinal study in Brazil. J Eukaryot Microbiol 2001;48:433–9.

[16] Sakihama N, Ohmae H, Bakote'e 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.

[17] Fidock DA, Nomura T, Talley AK, Cooper RA, Dzekunov SM, Ferdig MT, et al. Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol Cell 2000;6:861–71.

[18] Web version of the final report available from http://www.census.gov.ph/census2000/. CD-ROM version available from http://www.census.gov.ph/data/publications/pufindex.html.

[19] Baird JK, Caneta-Miguel E, Masbar S, Bustos DG, Abrenica JA, Layawen AV, et al. Survey of resistance to chloroquine of falciparum and vivax malaria in Palawan, The Philippines. Trans R Soc Trop Med Hyg 1996;90:413–4.

[20] Malaria Control Service Annual Report. Manila, Philippines: Department of Health; 1997.

[21] Schultz GW. Biting activity of mosquitos (Diptera: Culicidae) at a malarious site in Palawan, Republic of the Philippines. Southeast Asian J Trop Med Public Health 1992;23:464–9.

[22] Custodio GGV. National Objective for Health. Philippines 1999–2004. Manila: Department of Health; 1999 [web version available at http://www.doh.gov.ph/noh/].

[23] Sakihama N, Matsuo T, Mitamura T, Horii T, Kimura M, Kawabata M, et al. Relative frequencies of polymorphisms in variation in Block 2 repeats and 5' recombinant types of *Plasmodium falciparum* msp1 alleles. Parasitol Int 2004;53:59–67.

[24] Sakihama N, Mitamura T, Kaneko A, Horii T, Tanabe K. Long PCR amplification of *Plasmodium falciparum* DNA extracted from filter paper blots. Exp Parasitol 2001;97:50–4.

[25] Oberst RB, Schultz GW, Laughlin LW, Sy NE, Santos MN, Casimiro C. Epidemiological study of malaria in Palawan. Phil J Microbiol Infect Dis 1988;17:41–5.

[26] Hii JKL, Kanai L, Foligera A, Kan SKP, Burkot TR, Wirtz A. Impact of permethrin-impregnated mosquito nets compared with DDT house-

spraying against malaria transmission by *Anopheles farauti* and *An. punctulatus* in the Solomon Islands. Med Vet Entomol 1993;7:333–8.

[27] Babiker HA, Ranford-Cartwright LC, Currie D, Charlwood JD, Billingsley P, Teuscher T, et al. Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. Parasitology 1994;109:413–21.

[28] Luxemburger C, Thwai KL, White NJ, Webster HK, Kyle DE, Maelankirri L, et al. The epidemiology of malaria in a Karen population on the western border of Thailand. Trans R Soc Trop Med Hyg 1996;90:105–11.

[29] Qari SH, Shi Y-P, Goldman IF, Nahlen BL, Tiberenc M, Lal AA. Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP1), a potential vaccine antigen. Mol Biochem Parasitol 1998;92:241–52.

[30] Morgan WD, Birdsall B, Frenkiel TA, Gradwell MG, Burghaus PG, Seyd SEH, et al. Solution structure of an EGF module pair from the *Plasmodium falciparum* merozoite surface protein 1. J Mol Biol 1999;289:113–22.

[31] Mehlotra RK, Fujioka H, Roepe PD, Janneh O, Ursos LMB, Jacobs-Lorena V, et al. Evolution of a unique *Plasmodium falciparum* chloroquine-resistance phenotype in association with *pfcrt* polymorphism in Papua New Guinea and South America. Proc Natl Acad Sci U S A 2001;98:12689–94.

[32] Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, et al. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. Nature 2002;418:320–3.

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

[34] Kere NK, Parkinson AD, Samrawickerema WA. The effect of permethrin impregnated bednets on the incidence of *Plasmodium falciparum*, in children of north Guadalcanal, Solomon Islands. Southeast Asian J Trop Med Public Health 1993;24:130–7.

[35] Paul REL, Hackford I, Brockman A, Muller-Graf C, Price R, Luxemburger C, et al. Transmission intensity and *Plasmodium falciparum* diversity on the northwestern border of Thailand. Am J Trop Med Hyg 1998;58:195–203.

[36] Bustos DG, Canfield CJ, Caneta-Miguel E, Hutchinson DBA. Atovaquone-proguanil compared with choloroquine and choloroquine–sulfadoxine–pyrimethamine for treatment of acute *Plasmodium falciparum* malaria in the Philippines. J Infect Dis 1999;179:1587–90.

[37] Bustos DG, Lazaro JE, Gay F, Pottier A, Laracas CJ, Traore B, et al. Pharmacokinetics of sequential and simultaneous treatment with the combination chloroquine and sulfadoxine–pyrimethamine in acute uncomplicated *Plasmodium falciparum* malaria in the Philippines. Trop Med Int Health 2002;7:584–91.

[38] Chen N, Wilson DW, Pasay C, Bell D, Martin LB, Kyle D, et al. Origin and dissemination of chloroquine-resistant *Plasmodium falciparum* with mutant *pfcrt* alleles in the Philippines. Antimicrob Agents Chemother 2005;9:2102–5.

Available online at www.sciencedirect.com
ScienceDirect
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*.

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 multiclonal 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

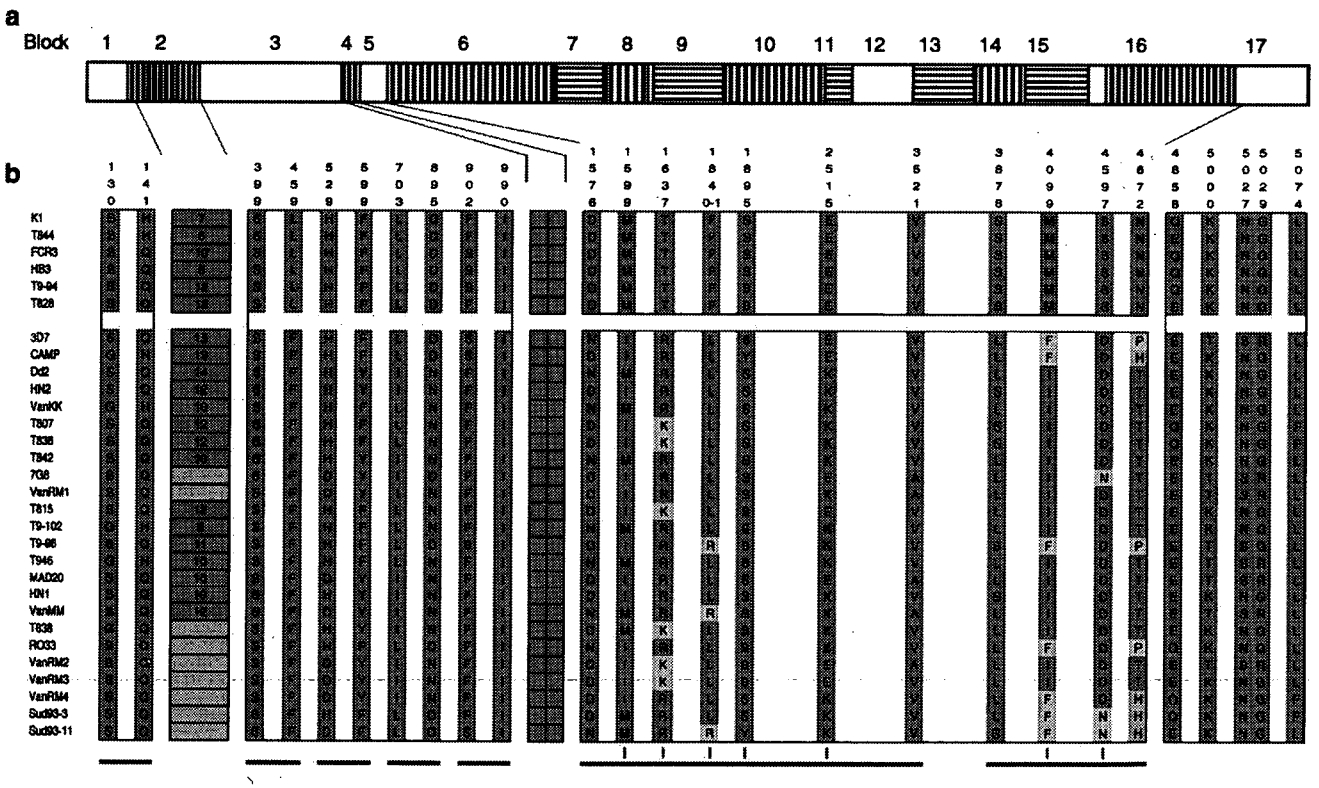

Fig. 1. Potential recombination sites in *Plasmodium falciparum msp1*. (a) The gene structure of *msp1*, showing five conserved blocks (open blocks), seven variable blocks (vertical-striped blocks) and five semi-conserve blocks (horizontal-striped blocks). (b) Potential recombination sites, inferred 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.

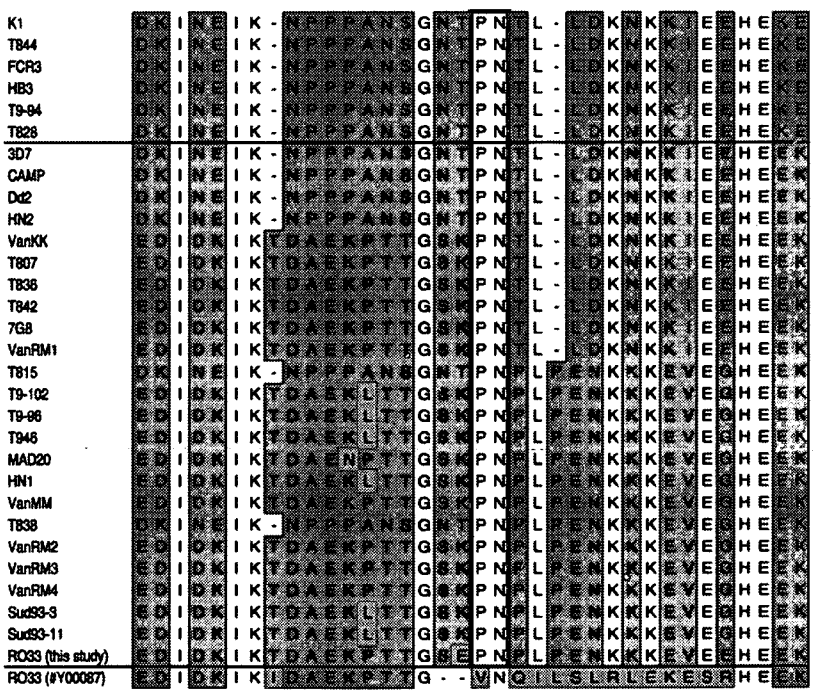

Fig. 2. A potential recombination site in *P. falciparum msp1* block 4. The site is boxed. The thirty *msp1* alleles shown in the left column were aligned manually, together with the reported sequence of the RO33 allele (Y00087).

Table 1
Lack of recombination events in the central region (blocks 6 to 16) of *P. falciparum msp1*

Allelic type		Haplotype	Tanzania (1993)		Tanzania (1998)		Solomon Islands	
			Observed	Expected	Observed	Expected	Observed	Expected
block 2	block 6							
K1	K1	K2K6	17	10	8	6	27	6
MAD20	K1	M2K6	2	5	4	4	2	13
RO33	K1	R2K6	2	6	1	3	0	10
K1	MAD20	K2M6	81	88	100	102	1	22
MAD20	MAD20	M2M6	51	48	69	69	56	45
RO33	MAD20	R2M6	54	50	62	60	44	34
(total)			207	207	244	244	130	130
			$P=0.01$		$P=0.43$		$P<0.0001$	
block 6	block 16							
K1	K1	K6K16	19	3	12	1	23	6
MAD20	K1	M6K16	0	16	0	11	0	17
K1	MAD20	K6M16	0	16	0	11	0	17
MAD20	MAD20	M6M16	91	75	99	88	66	49
(total)			110	110	111	111	89	89
			$P<0.0001$		$P<0.0001$		$P<0.0001$	

Isolates from Tanzania ($n=34$) (Tanabe et al., in press), Thailand ($n=117$) and the Solomon Islands ($n=58$) having K1-type in blocks 6 to 16 always had K1 allelic sequences in blocks 4a and 4b (Sakihama et al., 2006). This concordance suggests that a region of no recombination events can be extended from blocks 6 to 16 to blocks 4 to 16. Previously, recombination events have been reported in blocks 4 to 6 in isolates having K1-type in block 6, using PCR-based haplotyping (Ferreira et al., 1998; Sakihama et al., 1999). However, we found that the lack of allele specificity of some primers used for the haplotyping could produce PCR-artifact recombinants (Sakihama et al., 2004). By improving the haplotyping method with new primers, which are highly specific to allelic types, we did not observe any recombination events in blocks 4 to 6 between K1-type and Mad20-type in isolates having K1 allelic type in block 6 (Sakihama et al., 2006).

3.2. Mapping of recombination sites

Of the 34 *msp1* sequences including 24 sequences obtained here, 30 were distinct *msp1* alleles. Six alleles (present in 9 parasite lines) were of K1-type in blocks 6 to 16 (designated as K1-type$_{6-16}$), and 24 alleles (25 isolates) were of Mad20-type in blocks 6 to 16 (designated as Mad20-type$_{6-16}$). Four of the six *msp1* alleles of K1-type$_{6-16}$ were completely identical (from parasite lines T9/94, Wellcome, T806, and T837). Two of the 24 Mad20-type$_{6-16}$ alleles had identical sequence (Sudan 60-93-11 and Sudan 121-93-12).

Sequence alignment revealed no apparent recombination events in blocks 6 to 16 between K1-type and Mad20-type sequences (Fig. 1b). In K1-type alleles, block 4a and 4b sequences were always concordant with K1-type$_{6-16}$, suggesting a region of no apparent recombination between K1-type and Mad20-type alleles spanning from block 4 to block 16.

Potential recombination sites were mapped using the four-gamete test. The minimum number of recombination sites predicted by the four-gamete test was 1 and 4 in blocks 1 and 3, respectively (Fig. 1b). Nine recombination sites were also predicted in blocks 6 to 16 in Mad20-type$_{6-16}$. The six sequences of K1-type$_{6-16}$ obtained were nearly identical ($\pi=0.0005\pm0.0002$, $n=6$), and thus the test to detect recombination events could not be applied to sequences of K1-type$_{6-16}$. In block 17, the test did not predict any recombination sites. However, at least two recombination sites have been identified from field isolates: sites between nucleotides 4858 (amino acid Q or E) and 5000 (K or T), and between nucleotides 5029 (G or R) and 5074 (L or F) (Qari et al., 1998; Tanabe et al., 2000; Tanabe et al., in press).

Due to the impossibility of obtaining reliable alignments in highly variable blocks 2 and 4, the four-gamete test was not applicable to sequence of these blocks. Recombination events have recently been suggested to occur in block 2, a highly polymorphic sequence region characterized by the presence of absence of 9-bp repeats and the type of the repeats (Takala et al., 2002). Here, we did not detect recombination events in block 2 in any of the 34 sequences analyzed (see Supplementary Fig. 1). However, we obtained one recombinant-type sequence in 74 isolates from Tanzania: an allele having Mad20-type sequence at the 5′ end (90 bp) and RO33-type sequence at the 3′ end (75 bp) in block 2 (Tanabe, unpublished data). The unique allele was unlikely to be an artifact: we can exclude the possibility of artefactual recombination during PCR because the *msp1* sequence of the isolate was obtained from a blood sample with a single parasite genotype infection. Thus, we consider that recombination events occur in block 2 in nature, although at a low frequency. A recombination site has been noted previously in block 4, separating into blocks 4a and 4b, (Kaneko et al., 1996). This recombination site was confirmed by manual alignment of our sequences (Fig. 2): two parasite lines (T815 and T838) had the K1-type sequence and Mad20-type sequence in block 4a and 4b, respectively, and six lines (VanKK, T807, T836, T842, 7G8 and VanRM1) had the opposite recombinant-type sequence. Taken together, evidence from sequences so far obtained suggests that

there are a number of recombination sites throughout the entire *msp1* sequence.

3.3. Allelic dimorphism-associated restriction of recombination events

The conclusion of the study is that the recombination does not occur between the two major *msp1* allelic forms, K1-type and Mad20-type, of *P. falciparum msp1* within a central 3.5-kb region, comprising blocks 6–16. In contrast, there are numerous recombination sites in blocks 6 to 16 within Mad20-type sequences. *msp1* alleles having sequences of Mad20-type$_{6-16}$ also show recombination events in block 4. Therefore, *msp1* sequences can be divided into three major regions in terms of recombination events: (i) blocks 1 to 3, (ii) blocks 4 to 16, and (iii) block 17. In blocks 1, 3 and 17 inter-allelic recombination events potentially occur in all *msp1* alleles, whereas recombination events are restricted to Mad20-type sequences in blocks 4 to 16. Due to very limited sequence diversity in K1-type sequences in blocks 4 to 16, we cannot ascertain whether recombination events occur within K1-type sequences. More sequences will be needed to estimate recombination events within the K1 allelic group. Nevertheless, the present study indicates that the allelic dimorphism *most notable* in blocks 6 to 16 of *msp1*, as represented by the K1 allele and the Mad20 allele, greatly affects inter-allelic recombination events. *msp1* alleles of *P. vivax* do not show clear allelic dimorphism as seen in *P. falciparum msp1* and recombination sites are distributed throughout the entire *P. vivax msp1* gene (Putaporntip et al., 2002), highlighting a unique feature of recombination events in *P. falciparum msp1*.

A major question remains as to why recombination events do not occur in blocks 6 to 16 between the dimorphic groups (K1 and Mad20). Suppression of recombination events is commonly seen in genes under strong balancing selection, such as the self-incompatibility loci of plants (Clark and Kao, 1991). This is simply because mosaic alleles would be recognized as "self" by a large number of individuals in the population, resulting in severe restriction of their reproductive success. By analogy, it can be hypothesized that recombination in genes under strong immune-mediated selection might have been suppressed because recombinant Mad20/K1 sequences would be recognized by hosts exposed to either Mad20 or K1, making immune evasion more difficult. The finding that MSP-1 is a major target of protective allele-specific immunity (Martinelli et al., 2005) reinforces this hypothesis.

A recent evolutionary study shows that the K1 sequence in blocks 6 to 16 has a higher similarity to *P. reichenowi msp1* sequence than to the Mad20 sequence (Polley et al., 2005). We confirmed this finding by increasing the number of sequences: in an aligned region (3327 bp) in blocks 6 to 16, the genetic distance between K1-type (n=6) and Mad20-type (n=24) was 0.386±0.014, being much higher than that between *P. reichenowi* (n=1) and K1-type (0.086±0.005). (The genetic distance between *P. reichenowi* and Mad20-type was 0.387±0.014). Thus, the genetic distance within *P. falciparum* (K1 and Mad20-type alleles) is more divergent than between *P. falciparum* (K1 allele) and *P. reichenowi* in blocks 6 to 16. Such a high divergence may inhibit the formation of chromosomal chiasmata, resulting in the failure of homologous recombination between K1-type and Mad20-type sequences. However, the sequences of blocks 6 to 16 contain fairly conserved stretches in blocks 7, 9, 11, 12, 13, and 15 (see Supplementary Fig. 1), probably enabling the formation of chromosomal chiasmata in these sequence regions. The origin of *P. falciparum msp1* alleles has been suggested to be very ancient (Polley et al., 2005), far exceeding the divergence of *P. falciparum* and *P. reichenowi* at six million years ago. Interestingly, at the self-incompatibility locus of plants some interspecific sequence similarities are higher than intraspecific similarities, and some of the polymorphism predates speciation and has been maintained to the present (Ioerger et al., 1990), a situation similar to polymorphism of *P. falciparum msp1*. Suppression of recombination events over several million years of evolution could be a prerequisite for allelic dimorphism at the *P. falciparum msp1* locus. Therefore, we consider that evolutionary mechanisms leading to mechanical constraints on meiotic events, though unknown, are more likely to be involved in the suppression of recombination events in blocks 6 to 16 between two major allelic forms. Alternatively, it could be argued that recombination events do occur between the dimorphic sequences during meiosis, but parasites having a recombinant MSP-1 cannot survive due to functional constraints on MSP-1. To test such an epistatic selection theory, we examined *msp1* haplotypes in oocysts that were produced by crossing between gametes of 3D7 and HB3, whose *msp1* haplotypes were K2M6-16 and M2K6-16, respectively. As *msp1* is not expressed at the mosquito stage, recombinant haplotypes should not be affected by negative selection due to functional constraints on the protein, and the presence of such recombinants would provide support for the epistatic theory. Unfortunately, we did not detect recombinants in 404 oocysts produced by 18 independent crosses, either in blocks 6 to 16 or in blocks 2 to 6 (Supplementary Table 2). Thus, we cannot draw any conclusion regarding the likelihood of an epistatic selection mechanism.

3.4. Conclusion

Meiotic recombination is a genetic mechanism to generate allelic diversity in malaria parasites. This mechanism is used by *P. falciparum* to produce extensive allelic diversity in *msp1*. In this work, we obtained 24 distinct *msp1* alleles from 25 isolates having Mad20-type sequences in blocks 6 to 16. In those Mad20-type sequences, recombination events were seen to have occurred throughout the entire gene. In contrast, the number of alleles of K1-type sequence was very limited, and recombination events were restricted to blocks 1, 3 and 17. The present study provides no evidence of recombination events in blocks 6 to 16 occurring between the two major *msp1* allelic types, K1-type and Mad20-type, indicating that allelic dimorphism of *msp1*, which is remarkable in blocks 6 to 16, greatly affects inter-allelic recombination events. Further studies would be required to understand fully why *P. falciparum* possesses such an outstanding feature of recombination-associated allelic dimorphism in nature.

Acknowledgements

We thank an anonymous reviewer for constructive comments and suggestions, Richard Culleton for discussion, and S. Jongwutiwes, M. Kimura and T. Mitamura for parasite genomic DNA. This work was supported by Grants-in-Aid for Scientific Research 18390131, 18073013, 18GS03140013 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. This work is dedicated to the late John G. Scaife, who introduced me to the study of the genetic diversity of *msp1*.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2007.04.033.

References

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

Babiker, H.A., Abdel-Muhsin, A.A., Ranford-Cartwright, L., Satti, G., Walliker, D., 1998. Characteristics of *Plasmodium falciparum* parasites that survive the lengthy dry season in eastern Sudan where malaria transmission is markedly seasonal. Am. J. Trop. Med. Hyg. 59, 582–590.

Clark, A.G., Kao, T.-H., 1991. Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. Proc. Natl. Acad. Sci. U. S. A. 88, 9823–9827.

Conway, D.J., et al., 1999. High recombination rate in natural populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U. S. A. 96, 4506–4511.

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

Ferreira, M.U., Ribeiro, W.L., Tonon, A.P., Kawamoto, F.M., Rich, S., 2003. Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (*MSP1*) of *Plasmodium falciparum*. Gene 304, 65–75.

Hay, S.I., Rogers, D.J., Toomer, J.F., Snow, R.W., 2000. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, internet access and review. Trans. R. Soc. Trop. Med. Hyg. 94, 113–127.

Hii, J.L.K., Kanai, L., Foligela, A., Kan, S.K.P., Burkot, T.R., Wirtz, R.A., 1993. Impact of permethrin-impregnated mosquito nets compared with DDT-house-spraying against malaria transmission by *Anopheles farauti* and *An. punctulatus* in the Solomon Islands. Med. Vet. Entomol. 7, 333–338.

Holder, A.A., et al., 1999. Merozoite surface protein 1, immune evasion, and vaccine against asexual blood stage malaria. Parasitologia 41, 409–414.

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.

Ioerger, T.R., Clark, A.G., Kao, T.-H., 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. U. S. A. 87, 9732–9735.

Jongwutiwes, S., Tanabe, K., Nakazawa, S., Uemura, H., Kanbara, H., 1991. Coexistence of GP195 alleles of *Plasmodium falciparum* in a small endemic area. Am. J. Trop. Med. Hyg. 44, 299–305.

Kaneko, O., Jongwutiwes, S., Kimura, M., Kanbara, H., Tanabe, K., 1996. *Plasmodium falciparum*: variation in block4 of the precuror to the major merozoite surface proteins (MSP1) in natural populations. Exp. Parasitol. 84, 92–95.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

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

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Bioinformatics 5, 150–163.

Miller, L., Roberts, H.T., Shahabuddin, M., McCutchan, T.F., 1993. Analysis of genetic diversity in the *Plasmodium falciparum* merozoite surface protein-1 (*MSP1*). Mol. Biochem. Parasitol. 59, 1–14.

Mahanty, S., Saul, A., Miller, L.H., 2003. Progress in the development of recombinant and synthetic blood-stage malaria vaccines. J. Exp. Biol. 206, 3781–3788.

Martinelli, A., Cheesman, S., Hunt, P., Culleton, R., Raza, A., Mackinnon, M., Carter, R., 2005. A genetic approach to the de novo identification of targets of strain specific immunity in malaria parasites. Proc. Natl. Acad. Sci. USA 102, 814–819.

Mu, J., et al., 2005. Recombination hot spots and population structure in *Plasmodium falciparum*. PLoS Biol. 3, 1734–1741.

Peterson, M.G., Coppel, R.L., Moloney, M.B.D., Kemp, D., 1988. The third form of the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. Mol. Cell. Biol. 8, 2664–2667.

Polley, S.D., Weedal, 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., et al., 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.

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.

Rooth, I., Björkman, A., 1992. Fever episodes in a holoendemic area of Tanzania: parasitological and clinical findings and diagnostic aspects related to malaria. Trans. R. Soc. Trop. Med. Hyg. 86, 479–482.

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.

Sakihama, N., et al., 1999. Allelic recombination and linkage disequilibrium within *Msp1* of *Plasmodium falciparum*, the malignant human malaria parasite. Gene 230, 47–54.

Sakihama, N., Mitamura, T., Kaneko, A., Horii, T., Tanabe, K., 2001. Long PCR amplification of *Plasmodium falciparum* DNA extracted from filter paper blots. Exp. Parasitol. 97, 50–54.

Sakihama, N., et al., 2004. Relative frequencies of polymorphisms of variation in block 2 repeats and 5′ recombinant types of *Plasmodium falciparum msp1* alleles. Parasitol. Int. 53, 59–67.

Sakihama, N., Ohmae, H., Bakote, B., Kawabata, M., Hirayama, K., Tanabe, K., 2006. Limited allelic diversity of *Plasmodium falciparum msp1* from populations in The Solomon Islands, a highly endemic area. Am. J. Trop. Med. Hyg. 74, 31–40.

Snounou, G., et al., 1999. Biased distribution of *msp1* and msp2 allelic variants in *Plasmodium falciparum* populations in Thailand. Trans. R. Soc. Trop. Med. Hyg. 93, 369–374.

Takala, S., Branch, O., Escalante, A.A., Kariuki, S., Wootton, J., Lall, A.A., 2002. Evidence for intragenic recombination in *Plasmodium falciparum*: identification of a novel allele family in block 2 of remrozite surface protein-1: Asembo Bay Cohort Project XIV. Mol. Biochem. Parasitol. 125, 163–171.

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., et al., 2000. Selection and genetic drift of polymorphisms within the merozoite surface proetin-1 gene of *Plasmodium falciparum*. Gene 241, 325–331.

Tanabe, K., et al., 2002. In vitro recombination during PCR of *Plasmodium falciparum* DNA: a potential pitfall in molecular population genetic analysis. Mol. Biochem. Parasitol. 122, 211–216.

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

Tanabe, K., et al., 2004b. Genetic distance in housekeeping genes between *Plasmodium falciparum* and *Plasmodium reichenowi* and within *P. falciparum*. J. Mol. Evol. 59, 687–694.

Tanabe, K., Sakihama, N., Rooth, I., Björkman, A., Färnert, A., in press. High frequency of recombination-drived allelic diversity and temporal variation of *Plasmodium falciparum* in Tanzania. Am. J. Trop. Med. Hyg.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

WHO, 2000. WHO Expert Committee on Malaria. WHO Tech. Rep. Ser. 892, 174–181.

Am. J. Trop. Med. Hyg., 76(6), 2007, pp. 1037–1045

HIGH FREQUENCY OF RECOMBINATION-DRIVEN ALLELIC DIVERSITY AND TEMPORAL VARIATION OF *PLASMODIUM FALCIPARUM MSP1* IN TANZANIA

KAZUYUKI TANABE,* NAOKO SAKIHAMA, INGEGERD ROOTH, ANDERS BJÖRKMAN, AND ANNA FÄRNERT

Laboratory of Malariology, International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan; Nyamisati Malaria Research Unit, Dar es Salaam, Tanzania; Unit of Infectious Diseases, Department of Medicine, Karolinska Institute, Karolinska University Hospital, Stockholm, Sweden

Abstract. A major mechanism for the generation allelic diversity in the *Plasmodium falciparum msp1* gene is meiotic recombination in the *Anopheles* mosquito. The frequency of recombination events is dependent on the intensity of transmission. Herein we investigate the frequency of recombination-driven allelic diversity and temporal variation of *msp1* in Rufiji, eastern coastal Tanzania, where malaria transmission is intense. We identified 5′ recombinant types, 3′ sequence types, and *msp1* haplotypes (unique associations of 5′ recombinant types and 3′ sequence types) to measure the extent and temporal variation of *msp1* allelic diversity. The results show that *msp1* haplotype diversity is higher in Tanzania as compared with areas with lower transmission rates. The frequencies of individual polymorphic regions/sites remained stable during the study period. However, the frequency distribution of *msp1* haplotypes varied between 1993 and 1998. These results suggest that frequent recombination events between *msp1* alleles intermittently generate novel alleles in high transmission areas.

INTRODUCTION

The 200-kDa merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum* is a leading vaccine candidate antigen.[1,2] MSP-1 contains at least two regions targeted by host immunity: block 2 near the N terminus and block 17 at the C terminus. Human antibodies against block 2 are associated with protection from clinical malaria in highly endemic areas in Africa.[3] Block 17 encodes a C-terminal 19-kDa polypeptide, a product processed from MSP-1,[4] which confers protection after immunization against challenge with live parasites in animals.[5,6] Sera from individuals living in highly endemic areas contain antibodies against the 19-kDa fragment that inhibit merozoite invasion into red blood cells.[7–9]

MSP-1 exhibits extensive polymorphism,[10,11] which is a potential obstacle to the development of effective vaccines. In animal models, MSP-1 has been shown to be the major antigen involved in inducing "strain-specific immunity," in which the host mounts an immune response that is more effective against the immunizing strain than it is against genetically divergent strains.[12,13] As is the case for other *P. falciparum* antigen genes, *msp1* polymorphism is generated via a number of different mechanisms; point mutations result in single-nucleotide polymorphisms (SNPs), insertion/deletion of repeats cause 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[14] and may be of ancient origin.[15] Repeat-length polymorphisms are common in *msp1*[10,11,16] to the extent that size polymorphism between alleles is widely used as a marker for parasite genotyping.[17] Aside from repeat-length polymorphisms, meiotic recombination is the major mechanism for the generation of *msp1* allelic diversity.[10] Potential recombination sites have previously been mapped to restricted regions within *msp1* (see Figure 1).[10,11] The frequency of recombination in *P. falciparum* is dependent, to a large extent, on the rate of transmission, because meiotic recombination occurs only in the mosquito host. Recombination-driven allelic diversity in *msp1* is expected to be high in areas of intense malaria transmission and lower in areas with less intense transmission dynamics. The validity of this assumption remains to be tested, however, as very few studies have directly measured recombination-driven *msp1* diversity in areas of high transmission.

To investigate the nature and frequency of *msp1* allelic diversity in a highly endemic area, we conducted a study of the prevalence of *msp1* haplotypes in isolates collected 1993, 1998, and 2003 in Rufiji, eastern coastal Tanzania, where malaria transmission is intense and perennial.[18] Our results show that the extent of recombination-driven allelic diversity in *msp1* is higher in Tanzania as compared with areas with lower transmission rates. The frequency distribution of *msp1* haplotypes varied through time, but the frequencies of individual polymorphic regions and sites remained stable throughout the 10-year period of study. These results suggest that frequent recombination events in *msp1* intermittently generate novel *msp1* alleles in a high transmission area.

MATERIALS AND METHODS

Study area and sample collection. *P. falciparum* isolates were collected during malaria surveys from individuals living in Nyamisati village in the Rufiji River Delta, 150 km south of Dar es Salaam, in eastern coastal Tanzania in February and March 1993 (N = 120), 1998 (N = 132), and January 2003 (N = 104). Almost all samples were taken from asymptomatic donors of all ages with a mean age of 14.2 years (range, 1–78) and 16.8 years (range, 1–63) in 1993 and 1998, respectively, and from those aged 10–19 years (mean, 13.8 years) in 2003. Malaria in the study area was holoendemic with perennial transmission with some increase during the two rainy seasons, April to June and December.[18] An annual entomological inoculation rate is not available for the study area, but it is known to be in the range of 94 to 667 infective mosquito bites/person/year in eastern Tanzania.[19] Insecticide-impregnated bed nets were distributed to all houses in the village in 1999. Slide-positive parasite rates were recorded for the 1993 sample (46%) but were unavailable for the other sampling dates because of technical reasons. However, para-

* Address correspondence to K. Tanabe, International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan. E-mail: kztanabe@biken.Osaka-u.ac.jp

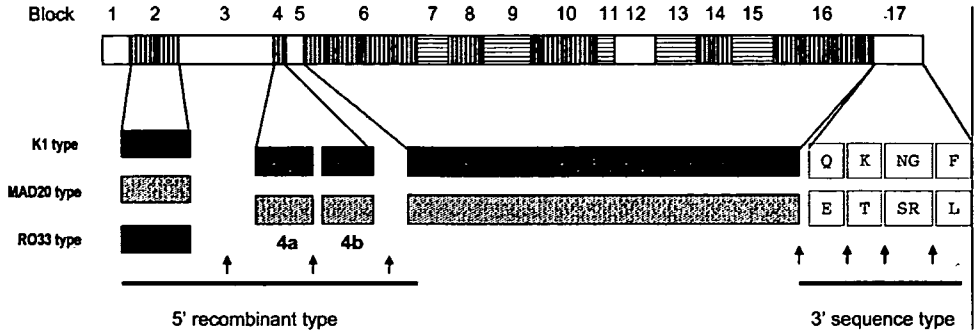

FIGURE 1. Determination of *P. falciparum msp1* haplotype, a unique association of 5′ recombinant types and 3′ sequence types. *msp1* is divided into 17 blocks, in which inter-allele conserved, semi-conserved, and variable blocks are indicated by open, horizontally hatched, and vertically hatched columns, respectively. For K1-type, MAD20-type, and RO33-type variable blocks, sequences are represented by densely toned, half-toned, and black bars, respectively. The 5′ recombinant types were determined by PCR amplification of blocks 2 to 6 using allelic type–specific primers of blocks 2 and 6, followed by nested PCR for blocks 4a and 4b using allelic type-specific primers of blocks 4a and 4b. Five amino acid substitutions in block 17 are indicated by the one-letter codes. The 3′ sequence type is the combination of those residues. Potential recombination sites are shown by arrows.

site positive rates in sampled people as checked by high-sensitivity PCR-based parasite detection (*msp1* typing method used in this study) were 78%, 77%, and 44% in 1993, 1998, and 2003, respectively.

All samples were collected after informed consent had been obtained from the donors or their guardians. Venous blood was collected into EDTA-containing tubes and stored at –20°C. Individuals with signs of clinical disease, i.e., fever and parasites, were treated with Fansidar. Parasite genomic DNA was extracted using the QIAamp DNA Blood Kit (Qiagen, Hilden, Germany). The volume of extracted DNA template was adjusted to be equivalent to the original blood volume. Ethical approval was obtained from the Ethical Committee of the National Institute for Medical Research, Tanzania, and the Ethical Committee of the Karolinska Institute, Sweden. Data previously obtained from clinical samples in Mae Sot in northwestern Thailand in 1995 and from survey samples in Guadalcanal Island in the Solomon Islands in 1994–1996 were used for geographical comparison.[20] We used clinical isolates (N = 111) from patients who attended a malaria clinic in Mae Sot in northwestern Thailand in 1995.[21] The mean age of the donors in Thailand was 24.6 years. A total of 90 isolates were collected in north Guadalcanal, the Solomon Islands: 40 clinical isolates from outpatients with a mean age of 18.3 years of a hospital in Honiara City and 50 isolates from four villages (Kaotave, Tadhimboko, Nugalitav, and Ruavatu).[20] In these rural villages, samples were collected in most cases from parasite-positive asymptomatic individuals during malariometric surveys, and most of the donors were primary-school children aged 8 to 15 years.

Determination of *msp1* polymorphisms. *P. falciparum msp1* (a 5-kb single-copy gene) consists of 17 distinct sequence blocks, according to the degree of sequence similarity among alleles (Figure 1).[10] Sequence variation in *msp1* is principally dimorphic (either one or the other of two major allelic types: K1 type and MAD20 type) in all variable blocks except block 2, which is trimorphic (represented by K1, MAD20, and RO33 types). To monitor the recombination-driven allelic diversity of *msp1*, we divided the gene into three regions: a 5′ 1.1-kb region (blocks 2 to 6), a central 3.5-kb region (blocks 6 to 16), and a 3′ 0.4-kb region (block 17), in which potential recombination sites have been mapped to the 5′ and 3′ regions (Figure 1). No recombination events occur in blocks 6 to 16.[16,21,22] The *msp1* haplotypes are thus defined as unique associations of 5′ recombinant types and 3′ sequence types in this study.

The 5′ recombinant types are defined as unique associations of allelic types of variable blocks 2, 4a, 4b, and 6. In total, 24 distinctive 5′ recombinant types are distinguishable: i.e., 24 = 3 × 2 × 2 × 2 (three allelic types designated as K, M, and R in block 2 and two allelic types designated as K and M in blocks 4a, 4b, and 6). The 5′ recombinant types were determined by our methods described previously.[20] In brief, they were determined by the following two steps: (i) first-round PCR to determine allelic types of blocks 2 and 6 using allelic-type-specific primers, and (ii) nested PCR to determine allelic types of blocks 4a and 4b (≈ 100 bp) using the first-round PCR products and allelic-type–specific primers. The PCR method allows us to determine the rate of multiple 5′ recombinant-type infections, here referred to as "polyinfection rate," and the mean number of 5′ recombinant-type infections per isolate (MORT). One microliter of template DNA was used for first-round PCR. 5′ Recombinant types were fully determined in 94 of 120 Tanzanian isolates collected in 1993, in 102 of 132 isolates in 1998 samples, and in 46 of 104 isolates in 2003. Thus, 68% (242/356) were PCR-positive in all samples obtained through malariometric surveys, indicating that our data represent a *P. falciparum* population in the study area.

The nucleotide sequence of block 17, which encodes the C-terminal 19-kDa polypeptide, was determined by direct sequencing after amplification of the full-length *msp1*. To see associations of 3′ sequence types (block 17 sequences) and 5′ recombinant types (blocks 2 to 6), only those isolates having a single 5′ recombinant type (i.e., mono-infection) were selected for further analysis. (We did not use cloning of the full-length *msp1* gene because artificial recombination readily occurs during amplification and cloning when samples with mixed genotypes are used.[22]) Because the number of isolates with mono-infections was limited in our Tanzanian samples, we increased the number of mono-infection samples by diluting genomic DNA templates by 20-fold. 5′ Recombinant types were again determined for the diluted samples, and those with a single 5′ recombinant type were selected. The

numbers of isolates sequenced were 38, 23, and 13 in samples collected in 1993, 1998, and 2003, respectively. No significant difference in the frequency distribution of 5′ recombinant types was found between undiluted original samples and diluted samples, indicating no bias of sampling after dilution (not shown). Amplification of the full-length *msp1* was first done with primers UPF1 (5′-GGCTAATGTAAAATGCAAAAATAAATGT) and DWR1 (5′-ACATGACTAAAATATCACTATTCCTGT) in a 20-μL reaction mixture containing 1 μL of template genomic DNA for 37 cycles using LA-*Taq* (TaKaRa, Tokyo, Japan). Two microliters of 10-fold diluted PCR products were amplified by nested PCR using primers UPF3 (5′-AATAAATGTATACATATTTTTGCTAAGTCA) and DWR3 (5′-TTAAGGTAACATATTTTAACTCCTACA) for 20 cycles. The PCR product was purified using the QIA Quick PCR purification kit (Qiagen) and directly sequenced from both directions using primers C17aFs (5′-CAAG(G/A)TATGTTAAACATTTCACAACA) and DWR3 with the BigDye Terminator Cycle sequencing kit (version 3.1) on an automated multicapillary ABI 3100 sequencer (Applied Biosystems, Foster City, CA). Sequences were verified by re-sequencing the PCR products independently amplified from the same DNA. To date, five major amino acid changes have been identified in block 17 from various geographic areas (E or Q at amino acid residue 1644; T or K at 1691; SR or NG at 1700–1701; and L or F at 1716; the positions are numbered according to Ref. 15) (Figure 1).[21,23] Hereafter, we refer to combinations of these residues as 3′ sequence type.

Unique associations of 5′ recombinant types and 3′ sequence types are referred to as *msp1* haplotypes. Partial sequencing of blocks 2 to 6 of the PCR amplicons (full-length *msp1*) confirmed 5′ recombinant types determined by PCR-based typing (Tanabe, unpublished). This indicates that our analysis of linkage between polymorphisms in the 5′ region and 3′ region is not affected by artificial recombination.

Statistical analyses. Frequency distributions of *msp1* 5′ recombinant types, 3′ sequence types, and *msp1* haplotypes were compared using the χ^2 test with Yates correction and Fisher's exact test for data sets fewer than 5. Differences in mean number of 5′ recombinant types per isolate (MORT) were tested for significance using a two-tailed Mann-Whitney *U* test. The diversity level of *msp1* haplotypes was expressed in two ways: (i) relative frequency of the number of unique *msp1* haplotypes per total number of *msp1* haplotypes, and (ii) expected heterozygosity (*h*). *h* and its variance were calculated as previously described.[20] Differences in the relative frequency were tested by *t* test. The frequency of recombination events in *msp1* was inferred from analysis of linkage disequilibrium within and between polymorphic blocks 2 to 6 and polymorphic sites in block 17. To assess linkage disequilibrium within *msp1*, pairs of polymorphic blocks 2, 4a, 4b, and 6 and polymorphic sites in block 17 were subjected to an R^2 test as described elsewhere.[24] Non-informative pairs (frequency < 10% in a polymorphic block or nucleotide site) were excluded from the R^2 test. Significance of linkage disequilibrium was assessed using the χ^2 test with Yates correction or two-tailed Fisher's exact probability test. A $P < 0.05$ was considered statistically significant.

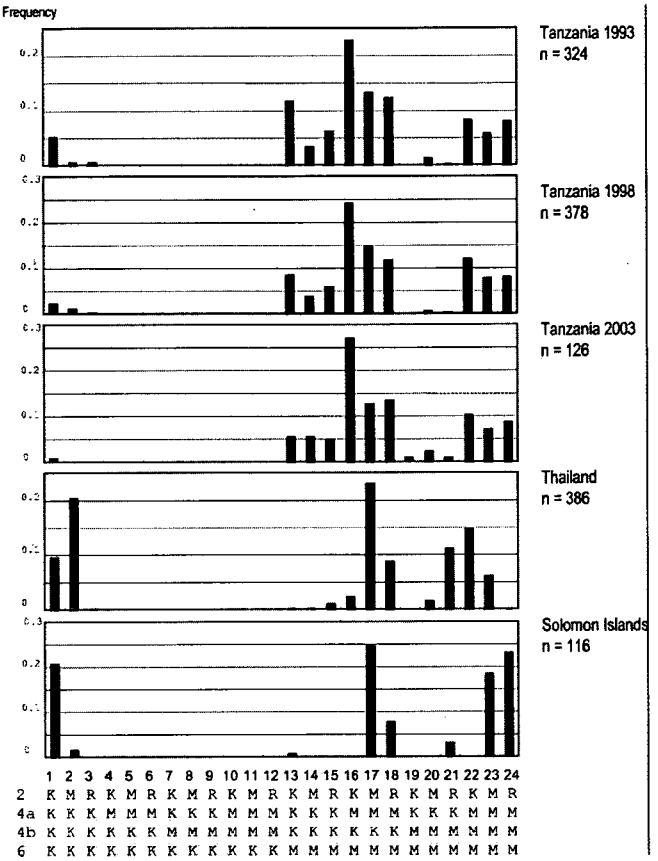

FIGURE 2. Frequency distribution of *P. falciparum msp1* 5′ recombinant types in Tanzania. Twenty-four distinct types—unique associations of allelic types in variable blocks 2, 4a, 4b, and 6—are shown at the bottom of the figure. Data from Thailand and Solomon Islands are from Sakihama et al.[20]

RESULTS

***msp1* 5′ recombinant types (blocks 2 to 6).** The frequency distributions of *msp1* 5′ recombinant types are shown in Figure 2. Types #1 to #12 are those with K1 allelic type in block 6, and types #13 to #24 are those with MAD20 allelic types. Most of the Tanzanian isolates were MAD20 allelic type in block 6 in 1993, 1998, and 2003. The overall pattern of fre-

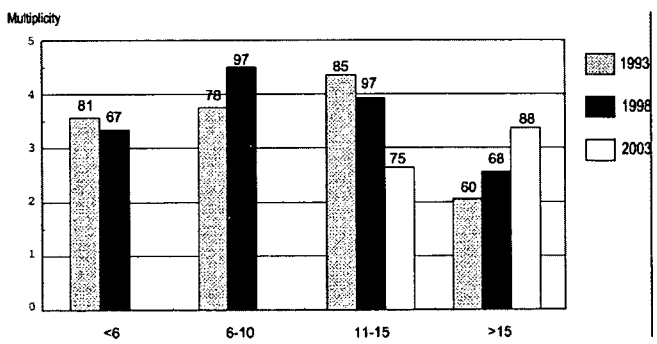

FIGURE 3. Age distribution of mean number of 5′ recombinant-type infections per isolates (MORT) in Tanzania. Ages are categorized into four classes: < 6, 6–10, 11–15, and > 15 yrs. Percentage of multiple 5′ recombinant infections is shown above each bar. Total numbers of isolates are 87, 95 and 44 in 1993, 1998 and 2003, respectively.

quency distribution of 5′ recombinant types was very similar from 1993 to 2003. The frequency distribution of 5′ recombinant types in Rufiji did not differ significantly from that reported previously in Tanga, northeastern Tanzania.[25] Tanzania was, however, significantly different from other geographic areas: Thailand and Solomon Islands ($P < 10^{-10}$), where frequencies of those types having K1 type in block 6 were substantially higher (19% in Solomon Islands and 30% in Thailand) as compared with Rufiji (< 7%).

Rates of multiple infections of the 5′ recombinant types (polyinfection rate) were 76.6%, 87.4%, and 78.3% in 1993, 1998, and 2003, respectively, and the mean number of 5′ recombinant types per isolate (MORT) was 3.48, 3.76, and 2.74, respectively. The reduction of MORT from 1998 to 2003 was significant ($P = 0.008$, Mann-Whitney U test). Both polyinfection rate and MORT are considerably higher in Tanzania than in the Solomon Islands (35.4–60.7% for polyinfection rate and 1.41–1.73 for MORT in Solomon Islands[20]). In Thailand, the polyinfection rate was 96.3% and MORT was 3.61, a level comparable to that observed in Tanzania. Thai isolates, however, were obtained from symptomatic patients, whereas Tanzanian isolates were from asymptomatic carriers, thus making direct comparison somewhat difficult. (There was no significant difference in polyinfection rate and MORT between individuals with clinical malaria and those with asymptomatic malaria in the Solomon Islands.[20])

There was a noticeable difference in age distribution of MORT from Tanzania (Figure 3). In 1993, MORT increased from age group < 6 years to age group 11–15 years and thereafter declined. MORT was significantly lower in age group > 15 years than other age groups ($P = 0.035$ against < 6 years, $P = 0.001$ against 6–10 years, and $P = 0.003$ against 11–15 years). In 1998, a peak MORT was observed in age group 6–10 years, followed by a significant reduction in age group > 15 years ($P = 0.003$). In contrast to the reduction in age group > 15 years in 1993 and 1998, MORT increased from age group

TABLE 1
Frequency distribution of *P. falciparum msp1* haplotypes in Tanzania

5′ Recombinant type	3′ Sequence type					Total	No. of *msp1* haplotypes
	QKNGL	QKNGF	EKNGL	EKNGF	ETSRL		
Tanzania 1993							
KKKK	1	0	0	0	0	1	
KKKM	2	0	1	0	0	3	
KMKM	2*	3	7	0	1	13	
KMMM	0	1 (QKSGF)	1	0	0	2	
MMKM	0	0	5	0	0	5	
(M/R)MMM	1	0	0	0	0	1	
RKKM	0	0	1	0	0	1	
RMKM	1	3	2*	2	0	8	
RMMM	3	0	0	0	1	4	
Total	10	7	17	2	2	38	20
Tanzania 1998							
MKKK	1	0	0	0	0	1	
KMKM	2	1	8	0	0	11	
KMMM	1 (QTSRL)	1	1	0	0	3	
MKKM	0	0	1	0	0	1	
MMKM	1	1	1	0	1+1 (EKSRL)	5	
RMKM	0	1	0	0	1	2	
Total	5	4	11	0	3	23	15
Tanzania 2003							
KKKK	1	0	0	0	0	1	
KKKM	1	0	0	0	0	1	
KMKM	3	2	2	0	0	7	
MMKM	0	0	0	0	1	1	
RMKM	2*	1	0	0	0	3	
Total	7	3	2	0	1	13	9
Thailand							
KKKK	2	0	0	0	0	2	
MKKK	9	0	4	0	0	13	
KMKM	0	0	0	0	1	1	
MMKM	0	2	12	0	2	16	
RMKM	0	0	3	0	0	3	
RKMM	0	0	3	0	1 (ETSGL)	4	
KMMM	1	1	2	0	3	7	
MMMM	0	0	1	0	1	2	
Total	12	3	25	0	8	48	16
Solomon Islands							
KKKK	9	0	0	0	0	9	
MMKM	1	0	0	0	13	14	
RMKM	0	0	0	0	1	1	
MMMM	0	1	5	0	3	9	
RMMM	0	0	0	0	14	14	
Total	10	1	5	0	31	47	8

Identical *msp1* haplotypes shared between 1993 and 1998 are boxed.
* A variant having a substitution from S to N at 1699 is included.

TABLE 2
Diversity of *P. falciparum msp1* haplotype in Tanzania

	No. of samples	No. of *msp1* haplotypes	Relative frequency	*P* value vs. Thailand vs. *Solomon*	$h \pm SE$*	*P* value vs. Thailand vs. *Solomon*
Tanzania						
1993	38	20	0.53	0.040 *0.002*	0.94 ± 0.02	0.212 *0.0002*
1998	23	15	0.65	0.007 *0.006*	0.89 ± 0.06	0.84 *0.21*
2003	13	9	0.69	0.015 *0.002*	0.94 ± 0.05	0.515 *0.022*
Thailand†	52	16	0.31		0.89 ± 0.03	
Solomon Islands	47	8	0.17		0.80 ± 0.03	

* h, expected heterozygosity as an index of haplotype diversity.[20]
† Data from Sakihama et al.[20]

11–15 years to age group > 15 years in 2003, but this trend was not statistically significant. (In 2003, sampling was limited to those of age > 10 years for technical reasons in the survey, and therefore MORT in age groups < 6 years and 6–10 years was not shown.) Polyinfection rates also showed similar patterns of age dependency. A sharp fall was noted from 11–15 years to > 15 years: 85% to 60% in 1993 and 97% to 68% in 1998.

3′ Sequence polymorphism (block 17). Five major nucleotide polymorphisms in block 17, all resulting in amino acid replacements, were observed in Tanzanian isolates ($N = 74$). We obtained 10 unique 3′ sequence types: Q-K-NG-L, Q-T-SR-L, Q-K-NG-F, Q-K-SG-F, E-K-NG-L, E-K-NG-F, E-T-SR-L, and E-K-SR-L (Table 1). In addition, minor variants showing Q-K-NNG-L ($N = 2$) and E-K-NNG-L ($N = 1$) were also observed, where the underlined "N" are substitutions for S at 1699, as detected earlier.[21,23] The number of 3′ sequence types was 5 and 4 in Thailand ($N = 48$) and Solomon Islands ($N = 47$), respectively.

Distribution and diversity of *msp1* haplotypes. The numbers of distinct *msp1* haplotypes were 20 in 38 isolates in 1993, 15 in 23 isolates in 1998, and 9 in 13 isolates in 2003 (Table 2). The *msp1* haplotype diversity, as expressed by relative frequency of the number of unique *msp1* haplotypes per total number of samples, was high in Tanzania in 1993 to 2003 (0.53–0.69) (Table 2). These levels were significantly higher than the level observed in Thailand (0.31; $P < 0.04$) and Solomon Islands (0.17; $P < 0.006$). Rare *msp1* haplotypes with a frequency of < 5% were abundant in Tanzania as compared with Thailand and Solomon Islands: 21/26 haplotypes (81%) in Tanzania in 1993 and 1998, 10/16 haplotypes (63%) in Thailand, and 3/8 haplotypes (38%) in Solomon Islands. Expected heterozygosity (h) was also high from 1993 to 2003 (Table 2). The difference in h reached statistical significance in 1993 and 2003 between Tanzania and Solomon Islands but not between Tanzania and Thailand.

Temporal variation in *msp1* polymorphisms. The frequencies of polymorphisms in polymorphic blocks 2, 4a, 4b, and 6 and five major polymorphic nucleotide sites in block 17 were compared from 1993 to 2003 (Table 3). A frequency variation was only observed in block 4a. Pairwise comparisons were also made between 1993 and 1998 ($P = 0.71$) and between 1998 and 2003 ($P = 0.06$). In contrast to the stable frequencies of individual polymorphisms, the frequency distribution of *msp1* haplotypes was clearly different between 1993 and 1998 (Figure 4) (χ^2 test, $P = 0.001$), indicating temporal variation of *msp1* haplotypes during this 5-year interval. (Rare *msp1* haplotypes were excluded from analysis: $N = 4$ in 1993 and $N = 2$ in 1998; see Table 1.) Among 26 distinct haplotypes found in 1993 and 1998 in a total of 61 isolates, only six haplotypes ($N = 35$) were shared between 1993 and 1998. Because of limited numbers of samples, a comparison with samples collected in 2003 was not made. The frequency distribution of *msp1* haplotypes in Tanzania was considerably different from that of Thailand and Solomon Islands (χ^2 test, $P < 10^{-10}$).

Linkage disequilibrium in *msp1*. To determine the frequency of recombination events in *msp1*, we performed linkage disequilibrium (LD) analysis, in which pairs of four polymorphic blocks (blocks 2, 4a, 4b, and 6) and four polymorphic sites were analyzed. Two sites at 1700 and 1701 in block 17 were always linked, and so they were combined for LD analysis. LD was undetectable in most pairs in Tanzania (Figure 5). Only one pair of 10 informative pairs in 1993 ($N = 37$) and one pair of 15 informative pairs in 1998 ($N = 23$) were significant. These pairs were within block 17. LD analysis was not carried out on samples from 2003, due to limited numbers ($N = 13$). These results indicate that the frequency of recombination events in *msp1* is high in the Tanzanian populations. In contrast, in Thailand and Solomon Islands 12 of 21 pairs

TABLE 3
Stable frequency of polymorphism in *Plasmodium falciparum msp1* in Tanzania

Block	Polymorphic type	*n* (frequency) 1993	1998	2003	*P* value
2	K1	156 (0.481)	176 (0.466)	56 (0.444)	0.869
	MAD20	79 (0.244)	104 (0.275)	35 (0.278)	
	RO33	89 (0.275)	98 (0.259)	35 (0.278)	
4a	K1	95 (0.293)	84 (0.222)	26 (0.206)	0.048
	MAD20	229 (0.707)	294 (0.778)	100 (0.794)	
4b	K1	247 (0.762)	271 (0.717)	88 (0.698)	0.261
	MAD20	77 (0.238)	107 (0.283)	38 (0.302)	
6	K1	21 (0.065)	13 (0.034)	1 (0.008)	0.06*
	MAD20	303 (0.935)	365 (0.966)	125 (0.992)	
17	1644:Q†	16 (0.432)	9 (0.391)	9 (0.692)	0.339
	1644:E	21 (0.568)	14 (0.609)	4 (0.308)	
	1691:T	2 (0.054)	3 (130)	1 (0.077)	0.775
	1691:K	35 (0.946)	20 (0.870)	12 (0.923)	
	1700-01:SR	3 (0.081)	4 (0.174)	1 (0.077)	0.772
	1700-01:NG	34 (0.919)	19 (0.826)	12 (0.923)	
	1716:L	28 (0.757)	19 (0.826)	10 (0.769)	0.926
	1716:F	9 (0.243)	4 (0.174)	3 (0.231)	

* Comparison between 1993 and 1998. Frequency in 2003 was not informative.
† Positions are after Miller et al.[11]