

ALLELIC VARIATION IN THE CIRCUMSPOROZOITE PROTEIN OF *PLASMODIUM FALCIPARUM* FROM THAI FIELD ISOLATES

SOMCHAI JONGWUTIWES, KAZUYUKI TANABE, MARIANNE K.
HUGHES, HIROJI KANBARA, AND AUSTIN L. HUGHES

Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; Laboratory of Biology, Osaka Institute of Technology, Osaka, Japan; Department of Biology, Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania; Department of Protozoology, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan

Abstract. Allelic variation in the *Plasmodium falciparum* circumsporozoite (CS) protein gene has been examined by sequencing the entire gene in 15 isolates from an endemic area of Thailand. The isolates contain a total of six new allelic forms of the tetrapeptide repeats and eight variants of the T cell epitope (TCE) region of the CS gene. All nucleotide substitutions in the TCE are nonsynonymous. There is no apparent association between the sequence patterns in the repeats and in the TCE. Comparison of the TCE with published sequences has shown that most variants of our isolates are not identical to those found in different geographic areas, suggesting geographic variation in genetic diversity of the CS protein. In a phylogenetic tree, the new Thai alleles did not cluster together, suggesting a considerable heterogeneity within some geographic areas. Furthermore, analyses of tetrapeptide repeats from a number of isolates and strains showed evidence of three genetic mechanisms for the generation of variation in the repeats of the CS gene: point mutation, duplication of one or more repeat units, and intragenic recombination.

The circumsporozoite (CS) protein of *Plasmodium falciparum* has been considered to be a target of host immune responses and contains immunodominant B cell epitopes in a tetrapeptide repeat region and a limited number of T cell epitopes in the C-terminal region. Antibodies against the tetrapeptide repeats inhibit hepatocyte invasion by sporozoites in vitro. The recognition of the CS T cell epitope is also important for mounting immunity against sporozoites and infected hepatocytes.¹ Analyses of the CS sequences from culture-adapted parasite strains have revealed substantial variations in these regions² and the CS gene of each parasite isolate so far characterized has a different sequence.²⁻⁹

Since the C-terminus of the CS protein contains several helper (CD4+) T cell epitopes and at least one cytotoxic (CD8+) T cell epitope,¹⁰⁻¹⁴ sequence variation in this portion of the gene has been studied in a number of field isolates obtained from different geographic areas.¹⁵⁻¹⁸ However, the extent of variation in the whole CS gene, including the tetrapeptide repeat region, is poorly known. We have sequenced the complete CS gene of 15 Thai isolates. Among these isolates, there are seven different CS sequences with extensive polymorphism in the number and pattern of the tetrapeptide repeats and in T cell

epitopes of the 3' portion of the gene. Analysis of the repetitive sequences of variants has provided evidence of possible mechanisms for the generation of different repetitive sequences.

MATERIALS AND METHODS

Source of parasite DNA. We isolated DNA from 19 isolates of *Plasmodium falciparum* from patients in the Mae Sod district in Thailand as described previously.¹⁹ Three culture-adapted strains, K1 (Kanchanaburi, Thailand),²⁰ T9/94 (Tak, Thailand),²¹ and MAD20 (Papua New Guinea),²² were also included in this study.

Polymerase chain reaction. Amplifications of DNA encompassing the whole CS gene were performed by polymerase chain reaction (PCR) with two primers (Figure 1): P1, 5'-ATTA-GAATTCATGATGAGAAAATTAGC-3' (nucleotides 68-94, positions after the 7G8 sequence³); and P2, 5'-ACTCAAGCTTAGA-TGTGTTCT-3' (nucleotides 1321-1341). Additional PCR amplifications for the 3' portion encompassing the 352 basepairs of the gene (nucleotides 948-1341) were also conducted with two primers: P2 and P3, 5'-AATCAAGCTTA-TGGACAAGGT-3' (nucleotides 948-968). Mismatches that are artificially introduced to create

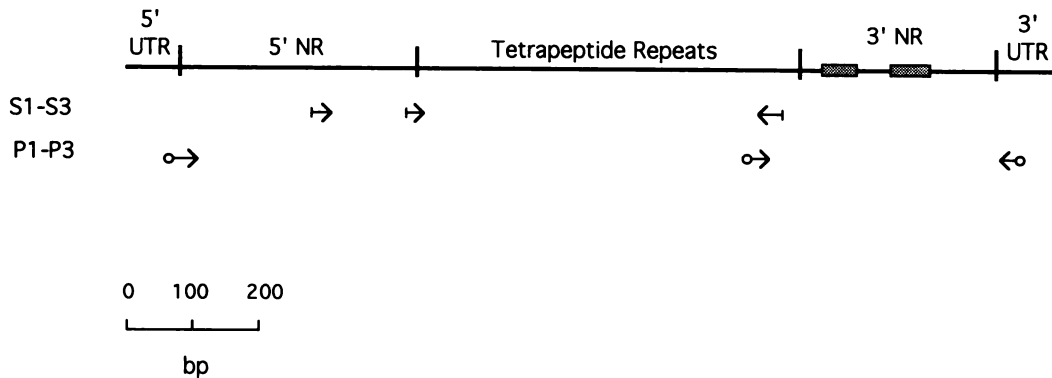


FIGURE 1. Map of the circumsporozoite gene of *Plasmodium falciparum*. Primers for the polymerase chain reaction (P1-P3) and sequencing (S1-S3) are indicated beneath the map. The regions used in analyses are as follows (nucleotides numbered after 7G8 sequence³): 5'NR = 5' nonrepeat region, nucleotides 78-443; tetrapeptide repeat region, nucleotides 444-938; 3'NR = 3' nonrepeat region, nucleotides 1019-1313; shaded boxes = Th2R (nucleotides 1055-1106) and Th3R (nucleotides 1157-1217). UTR = untranslated region; bp = base-pairs.

either *Eco* RI or *Hind* III sites are indicated in bold. The PCR conditions were described previously.²³

Subcloning and sequencing. The amplified DNA of the entire CS gene from each isolate was digested with *Eco* RI and *Hind* III and the 3' portion fragment was digested with *Hind* III. The fragments were subcloned into phagemid pUC118 and transformed into *Escherichia coli* strain JM109. The sequence was determined by the dideoxy chain termination technique using denatured template and the Sequenase version 2.0 kit (United States Biochemicals, Cleveland, OH). Three synthetic sequencing primers were used to obtain overlapping sequences of the whole gene (Figure 1): S1, 5'-AATAGTAGATCACTTGGAG-3' (nucleotides 279-297); S2, 5'-CAACCAGGGGATGGTAATCC-3' (nucleotides 420-439); and S3, 5'-GTCATTTGGCATATTGTGAC-3' (nucleotides 957-976). Sequences of the entire CS gene were analyzed for 15 field isolates and three culture-adapted parasites, K1, T9/94, and MAD20. These isolates and four other field isolates were sequenced for the 3' portion using different PCR-amplified subclones. Nucleotide sequence data reported in this paper have been submitted to the GenBank[®] database with the accession numbers M83149-M83174.

Statistical analysis of DNA sequences. A reliable alignment of the repeat region cannot be obtained because of the variation among alleles in numbers of repeat units (40-51 in the com-

plete sequences analyzed here). However, it is possible to align both the 5' nonrepeat region (5'NR) and the 3' nonrepeat region (3'NR).²⁴ Preliminary analysis of complete alleles in these regions identified two groups of alleles that were identical in both the 5'NR and the 3'NR: clones 806, 814, and 838; and clones 827, 842, and strain K1. Clones 814, 838, 842, and strain K1 were eliminated from subsequent analyses that involved both nonrepeat regions. In addition, six published sequences of nonrepeat CS protein regions were available for this analysis: 7G8,³ Wellcome,⁴ T4,⁵ T9/98,⁸ T9/101,⁸ and NF54-CVD.⁹ Preliminary analysis of the 3'NR only, which included the partial sequences from this study, showed that in this region, there were two groups of identical sequences: clones 808, 815, 947, 806, T9/101, and strain T4; and clones 822 and T9/94 and strains Wellcome and MAD20. Clones 815, 947, 806, T9/101, and T9/94 and strains T4, Wellcome, and MAD20 were eliminated from subsequent analysis of the 3'NR.

To obtain evidence of positive diversifying selection outside the tetrapeptide repeat region, we computed numbers of synonymous substitutions per synonymous site (d_s) and nonsynonymous substitutions per nonsynonymous site (d_n) by the method of Nei and Gojobori.²⁵ The 5'NR, the T cell epitope (TCE) consisting of Th2R and Th3R,^{11, 12} and the 3'NR excluding the TCE were analyzed separately as was done by Hughes²⁴ (Figure 1). In most genes, the rate of synonymous (silent) nucleotide substitution exceeds

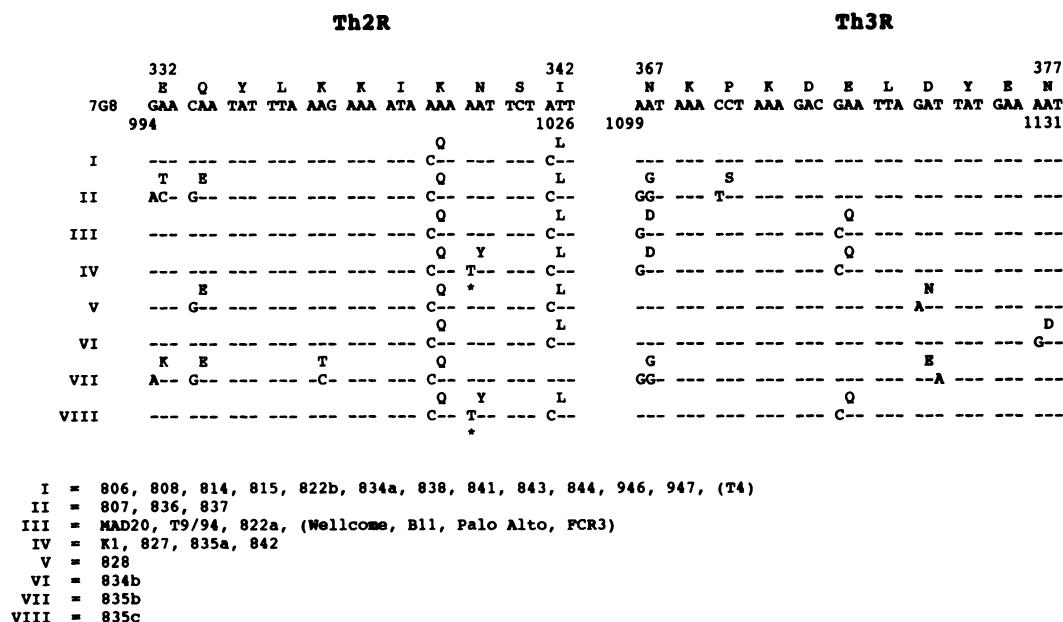


FIGURE 2. Variation in Th2R and Th3R in field isolates and laboratory strains of *Plasmodium falciparum*. Positions for nucleotides and amino acids are according to the 7G8 sequence³ as shown at the top. Subclones from one isolate are designated alphabetically after the isolate number. An identical nucleotide is indicated by a dash and amino acid substitutions are listed above the codon. Sequences shown are regions where substitutions occur. Roman numbers are combination groups for Th2R and Th3R in gene clones obtained from isolates that are listed at the bottom. One nucleotide substitution that has not been previously reported is indicated by an asterisk underneath the alignments.

that of nonsynonymous (amino acid-altering) substitution because amino acid substitutions are likely to be deleterious and thus eliminated by negative (purifying) selection. A region having $d_N > d_S$ is therefore evidence of positive selection.²⁶ McCutchan and others²⁷ have argued that in *P. falciparum* CS protein sequences, the extensive homoplasmy (similarity due to parallel evolution rather than to common ancestry) is evidence of positive selection favoring certain amino acid replacements. However, other factors, including shared functional constraint, can result in extensive homoplasmy. Thus, a comparison of synonymous and nonsynonymous rates of nucleotide substitution provides stronger evidence of positive selection.

Phylogenetic tree construction was done by the neighbor-joining method,²⁸ estimating the number of amino acid replacements per site (d_{aa}) by the Poisson-correction formula.²⁹

RESULTS

Sequence variation in nonrepeat regions. In the 5'NR, all alleles sequenced in this study are

identical to previously published CS protein alleles. Fifteen field isolates and two strains, K1 and MAD20, have sequences identical to 7G8,³ while T9/94 is the same as B11.² In the 3'NR, considerably more variation was found. Figure 2 shows changes in nucleotide and amino acid sequences of PCR-derived subclones from 19 wild isolates and three culture-adapted strains. Our field isolates have sequence variations in Th2R (amino acid residues 326-343 according to the 7G8 sequence)¹⁰ and Th3R (residues 361-380),¹¹ but not in CS.T3 (residues 378-398).³⁰ Variations in the cytotoxic T cell epitope (residues 351-395)^{12, 14} occur in the region overlapping Th3R. There are seven substitutions in Th2R, of which one, T at position 1,018, is newly identified in this study. There are also seven substitutions in Th3R; all substitutions in both of these regions create nonsynonymous codon changes.

Figure 2 also shows the eight observed combinations of substitutions in Th2R and Th3R that accounted for all alleles in this study. Group I, with 12 subclones, is identical to the previously

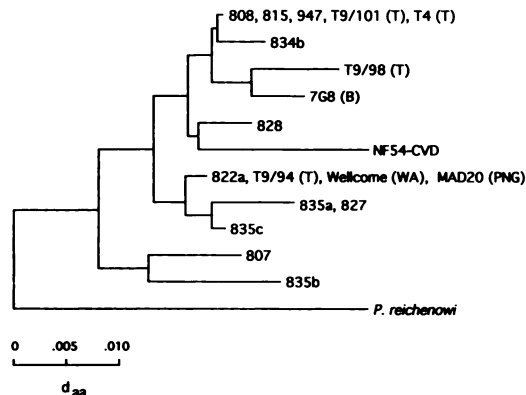


FIGURE 3. Neighbor-joining tree based on 3'NR of 11 *Plasmodium falciparum* alleles. The origins of the isolates are in parentheses: T = Thailand; B = Brazil; WA = West Africa; PNG = Papua New Guinea; d_{aa} = number of amino acids replacements per site.

published T4 isolate;⁵ while Group III sequences are identical to those of Wellcome, B11, Palo Alto, FCR3, and 22 isolates from Papua New Guinea.¹⁵⁻¹⁷ Isolates 822, 834, and 835 yielded gene clones with different combinations of substitutions, indicating mixed infections of variants in patients. Outside Th2R and Th3R, all sequences were identical with two exceptions: one group III allele, T9/94, differs from the sequence reported by Doolan and others¹⁷ in having T at position 907; and group IV alleles shared a substitution at residue 317 (Ala:GCT instead of Asp:GAT).

Figure 3 shows a phylogenetic tree based on 3'NR of all sequences, including partial sequences. *Plasmodium reichenowi*³¹ CS protein was used as an outgroup. The new Thai alleles are dispersed throughout the tree, suggesting that heterogeneity among these alleles is at least as extensive as is heterogeneity among alleles isolated in other parts of the world.

The tetrapeptide repeat region. Our analysis of the repeat region revealed nine new sequences in addition to nine already known. The range of lengths of the repeat sequences of 18 alleles was 40-51 units. Figure 4 shows the 14 different combinations of codons that were found in all repeat units of all alleles ($N = 799$). The most frequently used, designated 1 in Figure 4 (AAT GCA AAC CCA), occurred in almost half of all repeat units; and the five most common combinations accounted for 78.5% of repeat units. Ten combinations of codons coding for NANP made

up 92.1% of repeat units; four combinations coding for NVDP made up the remaining 7.9%. There are 64 possible combinations of codons that could produce each of the two observed tetrapeptide repeats; thus, only a small proportion of possible codons actually occurs in *P. falciparum* repeat regions.

Examination of the repeat region showed that patterns of repeat units appeared, in various combinations, in all 18 alleles (Figure 5). In all alleles, the 5' end of the repeat region consisted of 4-8 units of types 1, 7, and 8 appearing in varying order; in all alleles except NF54, Wellcome, B11, and T9/94, type 9 also occurred. These were followed by an uninterrupted series of 8-24 type 1 units. There was a significant positive correlation between the number of uninterrupted type 1 repeats and the overall length of the repeat region ($R = 0.870$, $P < 0.001$), suggesting that the number of type 1 repeats has a strong influence on the length of the repeat region. Following the uninterrupted type 1 repeats were 4-11 units that tended to be of types 2, 3, 6, 10, and 11 but showed no consistent pattern in comparisons among sequences. Finally, in the 3' end of the repeat region, blocks of units appeared in the same order in a number of sequences. Figure 5 shows six different types of blocks of 4-6 units each that occurred in the 3' portion of the repeat region; 13 of 18 alleles could be placed in one of two groups with respect to the arrangement of blocks in the 3' end. The remaining five alleles all had at least two of these common blocks; however, the order in which the blocks appeared (5' to 3') varied among alleles.

We have found evidence of three possible mechanisms for generating new allelic forms of the repeat region: point mutation, duplication (presumably by unequal crossing-over²) involving one or more repeat units, and recombination. The repeat regions of Wellcome and B11, for example, differ by a single point mutation in the fourth codon of unit 24; similarly, the repeat regions of B11 and T9/94 differ by a single point mutation in the first codon of unit 20. The duplication of a single type 1 repeat in units 7-14 of 7G8 results in the MAD20 type, and K1 repeats are identical to those of MAD20 in units 1-15, while units 16-43 are the same as units 20-46 of T9/94. Thus, intragenic recombination between MAD20 and T9/94 could create K1 repeats. Interestingly, outside the tetrapeptide

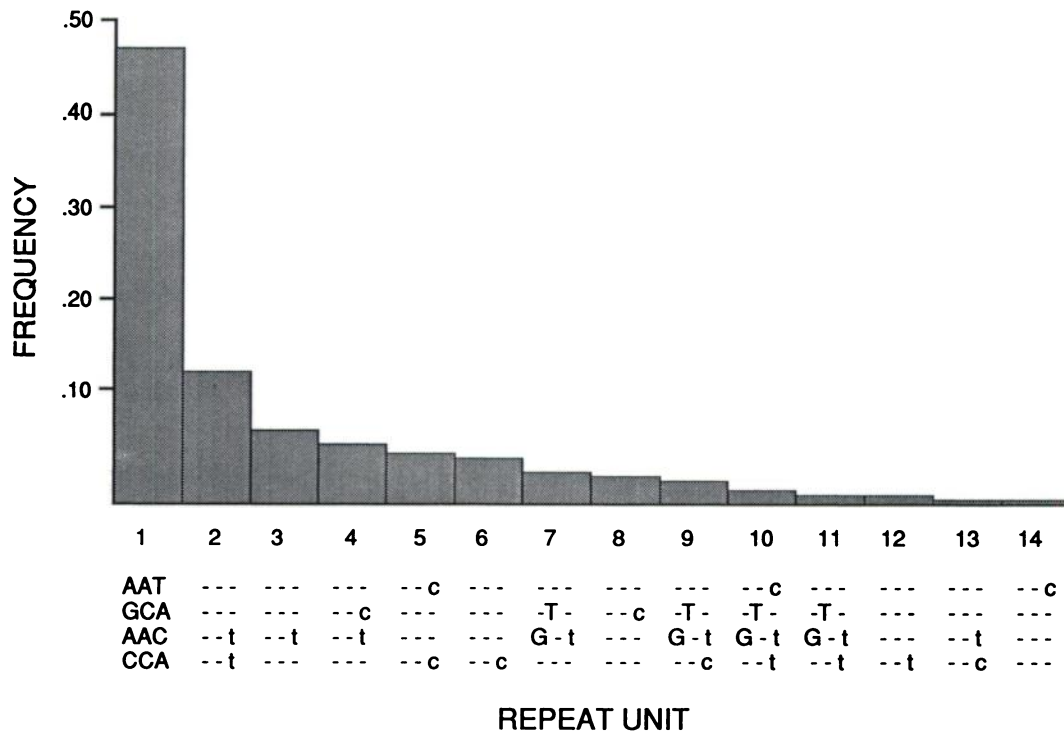


FIGURE 4. Frequency of occurrence of nucleotide sequences coding for 14 types of tetrapeptide repeat units observed in 18 allelic forms of *Plasmodium falciparum* repeat regions. An identical nucleotide is indicated by a dash. Nonsynonymous substitutions (relative to type 1) are indicated by capital letters; synonymous substitutions are indicated by lower case letters.

repeats, the 5' portion of K1 is identical to MAD20 while the 3' sequence of K1 is different from that of T9/94 in only two nucleotides.

We found no evidence of an association between the sequence patterns in the repeat region and that in the TCE in those alleles for which both these regions have been sequenced. The group I alleles of Figure 2, for example, are identical in their TCE but represent three different repeat region alleles. Groups III and IV are also represented by two and three types of repeat regions, respectively; on the other hand, groups II and VII contain alleles that are identical in both regions.

Diversifying selection on the T cell epitope. Table 1 shows values of d_s and d_N for the 5'NR, the TCE, and the remainder of the 3'NR (excluding the TCE) in pairwise comparisons among sequences. In the complete sequences, the numbers of synonymous substitutions in the 5'NR are small but not zero, indicating that contrary to the suggestion of Arnot,³² nothing pre-

vents synonymous substitution in the CS protein. In both the 5'NR and the 3'NR excluding the TCE, the numbers of nonsynonymous substitutions are also very small. In the TCE, on the other hand, d_N is significantly higher than d_s . This contrast is most marked in the sequences of Thai origin, which have the highest value of d_N of any of the comparisons. Among the partial sequences, the same pattern of $d_N > d_s$ in the TCE is apparent.

DISCUSSION

Our study is the first to examine sequence variation in the entire CS gene from field isolates and has revealed seven CS alleles in an endemic area of Thailand. We have also observed a nonsynonymous substitution in Th2R that has not been reported in any previous study of this region. Although they did not obtain complete sequences, Lockyer and others¹⁶ observed 12 variants with different amino acid se-

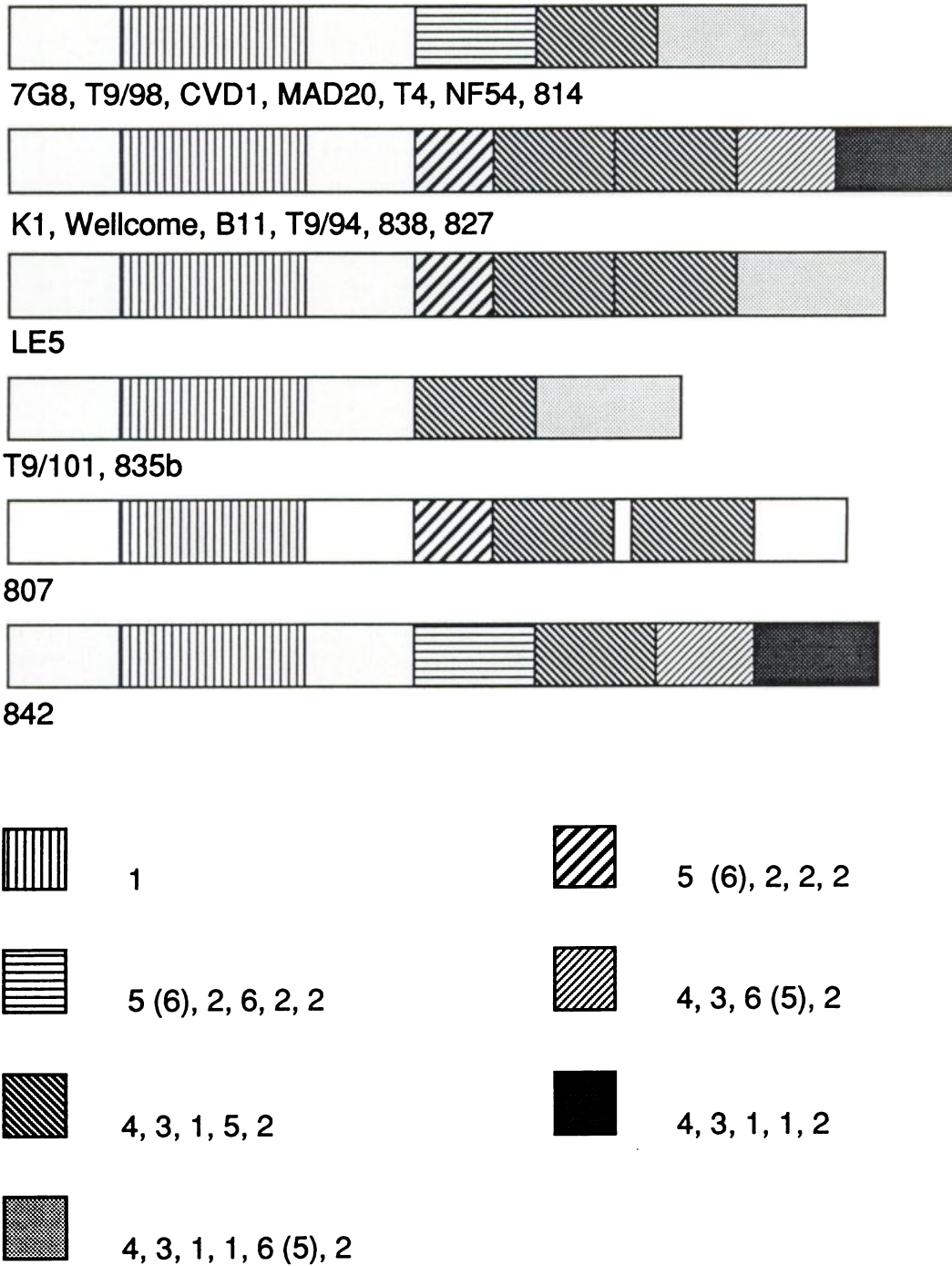


FIGURE 5. Schematic representation of *Plasmodium falciparum* tetrapeptide repeat regions. Numbers beside shaded squares at bottom identify repeat units in blocks of units occurring in more than one allele (see Figure 4). Numbers in parentheses indicate occasional variation in composition of blocks. Shaded areas (not drawn to scale) show the order of occurrence of blocks. Unshaded areas indicate portions of repeat region where no consistent pattern in repeat units could be found. Names of alleles having the patterns shown by the shaded areas are written under the bars.

TABLE 1

Numbers of synonymous (d_s) and nonsynonymous (d_n) nucleotide substitutions per 100 sites between alleles of the polymorphic *Plasmodium falciparum circumsporozoite gene**

Complete sequences (n)	5' nonrepeat region (65)†		T cell epitope (36)†		Remainder of the 3' region (42)†	
	d_s	d_n	d_s	d_n	d_s	d_n
All Thai origin sequences (8)	1.3 ± 1.0	0.5 ± 0.3	0.0 ± 0.0	5.4 ± 1.5‡	0.0 ± 0.0	0.2 ± 0.2
Thai origin (8) versus non-Thai (4)	1.5 ± 1.1	0.5 ± 0.3	0.1 ± 0.2	4.6 ± 1.3‡	0.0 ± 0.0	0.3 ± 0.2
Overall (12)	1.4 ± 1.1	0.5 ± 0.3	0.0 ± 0.2	5.0 ± 1.3‡	0.0 ± 0.0	0.2 ± 0.2
Partial sequences (n)	5' nonrepeat region		T cell epitope		Remainder of the 3' region	
	d_s	d_n	d_s	d_n	d_s	d_n
All Thai origin sequences (9)	–	–	0.0 ± 0.0	5.8 ± 1.5‡	0.0 ± 0.0	0.1 ± 0.2
Thai origin (9) versus non-Thai (2)	–	–	0.1 ± 0.3	5.8 ± 1.6‡	0.0 ± 0.0	0.4 ± 0.3
Overall (11)	–	–	0.0 ± 0.3	5.8 ± 1.4‡	0.0 ± 0.0	0.2 ± 0.2

* Values are the mean ± SEM. – = no data available for analysis.

† Values in parentheses are the no. of codons compared.

‡ Tests of the hypothesis that $d_s = d_n$; $P < 0.001$, by Student's *t*-test.

quences in Th2R and Th3R in five wild isolates from The Gambia. No variants in our isolates exhibit identical sequences to those found in Gambian isolates or Amazon isolates.^{15, 16} Thus, it seems likely that distribution of the CS alleles varies among different geographic areas, i.e., that certain CS alleles may be characteristic of particular geographic areas.^{16, 17} Similar geographically different allelic polymorphism has recently been reported in the CS gene of *P. vivax*.³³ However, the high rate of nonsynonymous nucleotide substitution among sequences of Thai origin (Table 1), and the uniform dispersal of our new Thai alleles throughout the phylogenetic tree of 3'NR (Figure 3) suggests that considerable heterogeneity exists within some geographic areas.

Much remains to be learned about the extent of variability in the CS gene. Sequence characterization of DNA of field isolates has so far concentrated on the C-terminal region of the gene, particularly the T cell epitopes. In this region, clinical isolates from the Amazon exhibited limited variation,¹⁵ whereas extensive variability was found in wild isolates from an endemic area of the Gambia.¹⁶ In Papua New Guinea, different studies have produced conflicting results regarding levels of diversity.^{17, 18} Shi and others¹⁸ hypothesized an association between high levels of *P. falciparum* polymorphism and high malaria endemicity. Our results are consistent with this hypothesis; however, to understand levels of gene diversity in any pop-

ulation, it is necessary to gather much larger samples than have so far been available for the CS gene.

It has been argued that the preponderance of nonsynonymous substitution in *P. falciparum* CS alleles is not due to positive selection favoring amino acid replacements but merely indicates that some factor is preventing synonymous substitution in this gene.³² This argument is based on the observation that there is a relatively low frequency of third-position differences in codons of *P. falciparum* surface proteins.³² Our data show nonzero values for d_s in both the 5'NR and TCE, indicating that at least low levels of synonymous substitution are possible in these regions. One explanation for low rates of synonymous substitution observed in CS alleles is that the A + T richness of the *P. falciparum* genome, particularly in third codon positions, reduces the observed rate of synonymous substitution at this position.²⁴ In the case of CS alleles, however, in addition to a relatively low rate of synonymous substitution, there is an elevated rate of nonsynonymous substitution in specific regions. In all CS alleles for which the entire 3'NR sequence is available, d_n in the TCE is significantly higher than is d_n in either the 5'NR or the remainder of the 3'NR (Table 1). Our data thus support the argument²⁴ that positive selection for avoidance of MHC recognition has caused the observed variation in 3'NR.

The increased number of allelic forms of the repeat region from our sequences together with

published sequences has provided evidence of three possible mechanisms involved in the origin of variation in the tetrapeptide repeat region: point mutation, duplication of one or more repeat units, and intragenic recombination. The significant correlation between the number of uninterrupted type 1 repeat units and the length of the repeat region (Figure 5) suggests that repeated events of duplication of type 1 units have been a major factor responsible for the variation in length seen in this region.² The pattern of five repeat units occurring twice in six of the alleles (Figure 5) suggests that this entire block of repeats has been duplicated. With regard to intragenic recombination, this mechanism has been reported to generate a variant of merozoite surface antigen-2 repeats in *P. falciparum*.³⁴ Such mechanisms are formally very similar to those postulated in the minisatellite DNA of mammalian genomes.³⁵ However, variation in the CS repeat sequence is different in that it occurs in coding sequences and is therefore rather more constrained than the variation found in minisatellite DNA. The repeat region may be strongly constrained at the amino acid level because of its role in eliciting an ineffective immune response by the host.³⁶ There is also apparently some degree of constraint at the DNA level that causes only a small fraction of theoretically possible codons to be observed in this region. However, allelic forms generated by duplication of one or more repeat units or interallelic recombination are apparently not eliminated by selection as long as the tetrapeptide repeat structure remains intact after such events.

In the present study, we used the same Thai isolate as previously characterized for the merozoite surface protein-1 (MSP-1) alleles.^{19, 23} Investigation into possible association between the CS alleles (chromosome 3) and MSP-1 alleles (chromosome 9) among 10 isolates that contained a single gene clone revealed no association of the two genes although the number of isolates compared was limited. It follows that parasites with identical sequences of the entire CS gene in this study are not the same parasites since they are different at the MSP-1 locus. Thus, our sequence analysis of the two loci does not reveal any evidence of clonality in these natural populations of *P. falciparum*,³⁷ but seem to support the notion that malaria genes located on different chromosomes segregate independently during meiosis in the mosquito vector.³⁸

Acknowledgments: We thank Sugako Obama and Miki Kinoshita for excellent support throughout the study.

Financial support: This work was supported by grant-in-aids from the Ministry of Culture, Science and Education of Japan (04454193), by grants from the United States National Institutes of Health to Austin L. Hughes (R01-GM-43940 and K04-GM-00614) and Masatoshi Nei (GM-20293), and by the Hitachi Scholarship Foundation (Tokyo).

Authors' addresses: Somchai Jongwutiwes, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. Kazuyuki Tanabe, Laboratory of Biology, Osaka Institute of Technology, Osaka 535, Japan. Marianne K. Hughes and Austin L. Hughes, Department of Biology, Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, PA 16802. Hiroji Kanbara, Department of Protozoology, Institute of Tropical Medicine, Nagasaki University, Nagasaki 852, Japan.

Reprint requests: Somchai Jongwutiwes, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand.

REFERENCES

1. Nussenzweig V, Nussenzweig RS, 1990. Rationale for the development of an engineered sporozoite malaria vaccine. *Adv Immunol* 45: 283-335.
2. de la Cruz VF, Lal AA, McCutchan TF, 1987. Sequence variation in the putative functional domains of the circumsporozoite protein of *Plasmodium falciparum*. *J Biol Chem* 262: 11935-11939.
3. Dame JB, Williams JL, McCutchan TF, Weber JL, Wirtz RA, Hockmeyer WT, Maloy WL, Haynes JD, Schneider I, Roberts D, Sanders GS, Reddy EP, Diggs CL, Miller LH, 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* 225: 593-599.
4. Lockyer MJ, Schwarz RT, 1987. Strain variation in the circumsporozoite protein gene of *Plasmodium falciparum*. *Mol Biochem Parasitol* 22: 101-108.
5. Del Portillo HA, Nussenzweig RS, Enea V, 1987. Circumsporozoite gene of a *Plasmodium falciparum* strain from Thailand. *Mol Biochem Parasitol* 24: 289-294.
6. Caspers P, Gentz R, Matile H, Pink R, Sinigaglia F, 1989. The circumsporozoite gene from NF54, a *Plasmodium falciparum* isolate used in malaria vaccine trials. *Mol Biochem Parasitol* 35: 185-190.
7. Campbell JR, 1989. DNA sequence of the gene encoding a *Plasmodium falciparum* malaria candidate vaccine antigen. *Nucleic Acids Res* 17: 5854.
8. Lockyer MJ, 1991. Clonal variation in the *Plas-*

- modium falciparum* circumsporozoite protein gene. *Mol Biochem Parasitol* 45: 179–182.
9. Davis JR, Cortese JF, Herrington DA, Murphy JR, Clyde DF, Thomas AW, Bagar S, Cochran MA, Thanassi J, Levine MM, 1992. *Plasmodium falciparum*: in vitro characterization and human infectivity of a cloned line. *Exp Parasitol* 74: 159–168.
 10. Good MF, Maloy WL, Lunde MN, Margalit H, Cornette JL, Smith GL, Moss B, Miller LH, Berzofsky JA, 1987. Construction of synthetic immunogen: use of new T helper epitope on malaria circumsporozoite protein. *Science* 235: 1059–1062.
 11. de la Cruz VF, Maloy WL, Miller LH, Good MF, McCutchan TF, 1989. The immunological significance of variation within malaria circumsporozoite protein sequences. *J Immunol* 142: 3568–3575.
 12. Malik L, Egan JE, Houghten RA, Sadoff JC, Hoffman SL, 1991. Human cytotoxic T lymphocytes against the *Plasmodium falciparum* circumsporozoite protein. *Proc Natl Acad Sci USA* 88: 3300–3304.
 13. Zevering Y, Houghten RA, Frazer IH, Good MF, 1990. Major population differences in T cell response to a malaria vaccine candidate. *Int Immunol* 2: 945–955.
 14. Doolan DL, Houghten RA, Good MF, 1991. Location of human cytotoxic T cell epitopes within a polymorphic domain of the *Plasmodium falciparum* circumsporozoite protein. *Int Immunol* 3: 511–516.
 15. Yoshida N, Di Santi SM, Dutra AP, Nussenzweig RS, Nussenzweig V, Enea V, 1990. *Plasmodium falciparum*: restricted polymorphism of the T cell epitopes of the circumsporozoite protein in Brazil. *Exp Parasitol* 71: 386–392.
 16. Lockyer MJ, March K, Newbold CI, 1989. Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epitopes of the circumsporozoite protein. *Mol Biochem Parasitol* 37: 275–280.
 17. Doolan DL, Saul AL, Good MF, 1992. Geographically restricted heterogeneity of the *Plasmodium falciparum* circumsporozoite protein: relevance for vaccine development. *Infect Immun* 60: 675–682.
 18. Shi Y-P, Alpers MP, Pova MM, Lal AA, 1992. Diversity of the immunodominant determinants of the circumsporozoite protein of *Plasmodium falciparum* parasites from malaria-endemic regions of Papua New Guinea and Brazil. *Am J Trop Med Hyg* 47: 844–851.
 19. 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.
 20. Thaithong S, Beale GH, 1981. Resistance of ten Thai isolates of *Plasmodium falciparum* to chloroquine and pyrimethamine by in vitro tests. *Trans R Soc Trop Med Hyg* 75: 271–273.
 21. Thaithong S, Beale GH, Fenton B, McBride JS, Rosario V, Walker A, Walliker D, 1984. Clonal diversity in a single isolate of the parasite *Plasmodium falciparum*. *Trans R Soc Trop Med Hyg* 78: 242–245.
 22. Tanabe K, Mackay M, Goman M, Scaife JG, 1987. Allelic dimorphism in a surface antigen gene of malaria parasite *Plasmodium falciparum*. *J Mol Biol* 195: 273–287.
 23. Jongwutiwes S, Tanabe K, Nakazawa S, Yanagi T, Kanbara H, 1992. Sequence variation in the tripeptide repeats and T cell epitopes in p190 (MSA-1) of *Plasmodium falciparum* from field isolates. *Mol Biochem Parasitol* 51: 81–90.
 24. Hughes AL, 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* 127: 345–353.
 25. Nei M, Gojobori T, 1986. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.
 26. Hughes AL, Nei M, 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167–170.
 27. McCutchan TF, Lal AA, do Rosario V, Waters AP, 1992. Two types of sequence polymorphism in the circumsporozoite gene of *Plasmodium falciparum*. *Mol Biochem Parasitol* 50: 37–46.
 28. Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
 29. Nei M, 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
 30. Sinigaglia F, Guttinger M, Kilgus J, Doran DM, Matile H, Etlinger H, Trzeciak A, Gillessen D, Pink JRL, 1988. A human T-cell epitope recognized in association with most mouse and human MHC class II molecules. *Nature* 336: 778–780.
 31. Lal AA, Goldman IF, 1991. Circumsporozoite protein gene from *Plasmodium reichenowi*, a chimpanzee malaria parasite evolutionarily related to the human malaria parasite *Plasmodium falciparum*. *J Biol Chem* 266: 6686–6689.
 32. Arnot D, 1989. Malaria and the major histocompatibility complex. *Parasitol Today* 5: 138–142.
 33. Qari SH, Goldman IF, Pova MM, Santi S, Alpers MP, Lal AA, 1992. Polymorphism in the circumsporozoite protein of the human malaria parasite *Plasmodium vivax*. *Mol Biochem Parasitol* 55: 105–113.
 34. Marshall VM, Coppel RL, Martin RK, Oduola AMJ, Anders RF, Kemp DJ, 1991. A *Plasmodium falciparum* MSA-2 gene apparently generated by intragenic recombination between the two allelic families. *Mol Biochem Parasitol* 45: 349–351.
 35. Jeffreys AJ, Wilson V, Thein SL, 1985. Hyper-variable ‘mini-satellite’ regions in human DNA. *Nature* 314: 67–73.
 36. Enea V, Arnot D, 1988. The circumsporozoite gene in *Plasmodia*. Turner MJ, Arnot D. eds. *Molecular Genetics of Parasitic Protozoa*. Cold

- Spring Harbor, NY: Cold Spring Harbor Laboratory, 5–11.
37. Tibayrenc M, Kjellberg F, Ayala FJ, 1990. A clonal theory of parasitic protozoa. The population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc Natl Acad Sci USA* 87: 2414–2418.
38. Walliker D, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, London WT, Corcoran LM, Burkot TR, Carter R, 1987. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* 236: 1661–1666.