

## Sequence Variation in the T-Cell Epitopes of the *Plasmodium falciparum* Circumsporozoite Protein among Field Isolates Is Temporally Stable: a 5-Year Longitudinal Study in Southern Vietnam

Amadu Jalloh,<sup>1\*</sup> Huynh van Thien,<sup>2</sup> Marcelo U. Ferreira,<sup>3</sup> Jun Ohashi,<sup>4</sup> Hiroyuki Matsuoka,<sup>5</sup> Toshio Kanbe,<sup>1</sup> Akihiko Kikuchi,<sup>1</sup> and Fumihiko Kawamoto<sup>6</sup>

Division of Molecular Mycology and Medicine, Department of Advanced Medical Science, Center for Neurological Disease and Cancer, Nagoya University Graduate School of Medicine, 65 Tsurumai, Showa-ku, Nagoya, Aichi 466-8550, Japan<sup>1</sup>; Bao Loc General Hospital, Lam Dong Province, Bao Loc, Vietnam<sup>2</sup>; Departamento de Parasitologia, Instituto de Ciencias Biomedicas da USP, Av. Prof. Lineu Prestes 1374, Cidade Universitaria, 05508-900 Sao Paulo (SP), Brazil<sup>3</sup>; Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan<sup>4</sup>; Division of Medical Zoology, Department of Infection and Immunity, Jichi Medical School, 3311-1 Yakushiji, Minami-kawachi Tochigi-ken, 329-0498, Japan<sup>5</sup>; and Department of Social and Environmental Medicine, Institute of Scientific Research, Oita University, Oita 879-5593, Japan<sup>6</sup>

Received 2 September 2005/Returned for modification 28 October 2005/Accepted 11 January 2006

**In an effort to decipher the nature and extent of antigen polymorphisms of malaria parasites in a setting where malaria is hypomesoendemic, we conducted a 5-year longitudinal study (1998 to 2003) by sequencing the Th2R and Th3R epitopes of the circumsporozoite protein (CSP) of 142 *Plasmodium falciparum* field isolates from Bao Loc, Vietnam. Samples were collected during the high-transmission season, September through December 1998 ( $n = 43$ ), as well as from July 2000 to August 2001 ( $n = 34$ ), September 2001 to July 2002 ( $n = 33$ ), and August 2002 to July 2003 ( $n = 32$ ). Marked sequence diversity was noted during the high-transmission season in 1998, but no significant variation in allele frequencies was observed over the years ( $\chi^2 = 70.003$ , degrees of freedom = 57,  $P = 0.116$ ). The apparent temporal stability in allele frequency observed in this Bao Loc malaria setting may suggest that polymorphism in the Th2R and Th3R epitopes is not maintained by frequency-dependent immune selection. By including 36 isolates from Flores Island, Indonesia, and 19 isolates from Thaton, Myanmar, we investigated geographical patterns of sequence polymorphism for these epitopes in Southeast Asia; among the characterized isolates, a globally distributed variant appears to be predominant in Vietnam (75 of 142 isolates, or 52.8%) as well as in Myanmar (15 of 19 isolates, or 78.9%) and Indonesia (31 of 36 isolates, or 86.1%). Further analyses involving worldwide CSP sequences revealed distinct regional patterns, a finding which, together with the unique mutations observed here, may suggest a possible role for host or local factors in the generation of sequence diversity in the T-cell epitopes of CSP.**

As one of the most threatening tropical diseases, malaria continues to maintain its toll in the developing world and poses a heavy economic burden on many countries. Estimates show that 300 million to 500 million clinical cases occur per year, with at least a million deaths, most of which are seen among children under 5 years of age (8). Resistance to antimalarial drugs and insecticides, coupled with the lack of availability of an effective vaccine, is the leading factor behind the success of the parasite's continuing burden. Apart from its complex life cycle, which alternates between the human and the mosquito host, the malaria parasite also exhibits stages characterized by extensive genetic and antigenic diversity which may present adverse obstacles to antimalarial control measures.

The *Plasmodium falciparum* circumsporozoite protein (CSP), a leading malaria vaccine candidate antigen, is predominantly distributed on the surface of the sporozoites (40) and has

approximately 120 residues, with a molecular mass of about 58 kDa. The structure of CSP can be divided into a polymorphic central repeat region known to contain immunodominant B-cell epitopes (41) and flanked by a less variant 5' and a highly polymorphic 3' end where T-cell epitopes have been identified (19, 27). Clinical trials of *Plasmodium falciparum* CSP-based vaccines have shown that they induce both humoral and cellular immune responses (29, 39) and that the cellular immune responses are central in immunity against the exoerythrocytic stages of malaria.

In an attempt to understand the nature and the extent of the polymorphism patterns of malaria parasites in a setting of malaria hypomesoendemicity, we have conducted a 5-year longitudinal study of the most variable terminal 3' region of the CSP gene among *Plasmodium falciparum* field isolates collected from Bao Loc, southern Vietnam, over 5 years (1998 to 2003). We compare our data with the reported CSP sequences and also discuss the geographical patterns of polymorphism at this locus in Southeast Asia and globally. Our findings show that the genetic diversity of the Th2R and Th3R epitopes in this Bao Loc malaria setting is less than that in areas of hyperendemicity and that the variation patterns tend to corre-

\* Corresponding author. Mailing address: Division of Molecular Mycology and Medicine, Department of Advanced Medical Science, Center for Neurological Disease and Cancer, Nagoya University Graduate School of Medicine, 65 Tsurumai, Showa-ku, Nagoya, Aichi 466-8550, Japan. Phone: (81) 52-744-2460. Fax: (81) 52-744-2459. E-mail: alphamadu@yahoo.com.

late with malaria transmission intensity. Furthermore, the polymorphisms appear to be stable over time in this Bao Loc setting of relatively low malaria endemicity. Findings from our analyses involving worldwide CSP sequences seem to suggest that the sequence polymorphisms observed in the Th2R and Th3R epitopes of CSP not only appear to be geographically restricted but also may be a result of a balanced state(s) of interaction between the parasite and its hosts.

#### MATERIALS AND METHODS

**Study population.** Bao Loc lies about 180 km northeast of Ho Chi Minh City and is a district with a growing population of 151,000 inhabitants. The city belongs to Lam Dong Province, which is the southernmost of the four provinces belonging to the Tay Nguyen Highlands. Lam Dong is in an area in the southern highlands of Vietnam where malaria is mesoendemic and where many minority ethnic groups reside. Although malaria has disappeared from many areas in Vietnam, it is still endemic in the central and southern provinces of the Tay Nguyen Highlands (38), where all four *Plasmodium* species that cause human malaria are present year round, with a peak in transmission shortly after the rainy season (October to December). The prevalence of parasitemia in the general population typically fluctuates between 10% and 30%, of which approximately 75% of the cases are due to *P. falciparum* (unpublished observations). Local malaria vectors include *Anopheles dirus* sensu lato and *Anopheles minimus* sensu lato. After 10 years of intensive control efforts, the rates of malaria morbidity and mortality in Vietnam have decreased by 60% and 97%, respectively.

**Sample collection.** This study was approved by the ethical committees of the Nagoya University Graduate School of Medicine and the Bao Loc District General Hospital. Following the receipt of informed consent, venous blood samples were collected from malaria patients presenting at the Bao Loc General Hospital over the period from September 1998 to December 1998 and from July 2000 to August 2003. Bao Loc is a region for which no previous epidemiologic data exist for this malaria antigen gene, although epidemiologic data are available for other candidate genes (16, 33). The blood samples were categorized into four groups: (i) group I (BL1998), September 1998 to December 1998; (ii) group II (BL2001), July 2000 to August 2001; (iii) group III (BL2002), September 2001 to July 2002; and (iv) group IV (BL2003), August 2002 to July 2003. We further included field isolates collected from Thaton, Myanmar ( $n = 19$ ), during a field survey in 1999 and another 36 collected from Flores Island, Indonesia, in 2003 ( $n = 14$ ) and 2004 ( $n = 22$ ).

**DNA extraction, amplification, and sequencing.** Slide-positive *P. falciparum* infections, as demonstrated by Giemsa staining, were analyzed. Genomic DNA was extracted from frozen venous blood samples by using a DNeasy kit (QIAGEN, Hilden, Germany), following the manufacturer's instructions. The carboxy-terminal portion of the CSP gene was amplified by using primers PfCSF (TGT AGA TGA AAA TGC TAA TGC) and PfCSR (CGA CAT TAA ACA CAC TGG A) in a 25- $\mu$ l PCR mixture containing 0.5  $\mu$ M of each primer, 1 to 3  $\mu$ l of template DNA, 2.5  $\mu$ l of 10 $\times$  buffer, 200  $\mu$ M of each deoxynucleoside triphosphate, and 0.5 units of AmpliTaq Gold DNA polymerase (PE Applied Biosystems). Thermal conditions were incubation at 95°C for 5 min, followed by 35 cycles of 94°C for 30 s, 58°C for 50 s, and 72°C for 1 min. A final extension of 72°C for 5 min was also included. The 268-bp amplicons were resolved on 2.0% agarose gels following electrophoresis in the presence of ethidium bromide (0.5 mg/ml). Distilled water and genomic DNA from *P. vivax* were used as negative controls. Positive samples (i.e., PCR fragments) were purified with a High Pure PCR product purification kit (Boehringer Mannheim, Mannheim, Germany), as directed by the manufacturer, and a precipitation step was achieved with Etachimate (Tokyo, Japan). DNA quantity was estimated by both electrophoresis and measurement of absorption values by determination of the optical density at 260 nm, after which a 1:10 dilution was prepared. Sequencing reactions were performed with the forward and reverse primers in separate reaction mixtures with the BigDye Terminator sequencing kit (version 1.1; PE Applied Biosystems). The sequences in both directions were then read with an ABI 310 DNA sequencer. Whenever a mutant or singleton was encountered, a third independent reaction mixture was made, and the product was resequenced to confirm the mutation.

**Data analyses.** To determine the concurrence of the findings of the present study with the findings of previous reports, the sequences were aligned and the amino residues were numbered with reference to the sequence of the 7G8 clone from Brazil (10). The alignment was done with the software Genetyx Mac (version 11.0), whereas the DNA sequence polymorphism software DnaSP (ver-

TABLE 1. Analyses of the C-terminal region (231 bp) of the CSP sequences of *P. falciparum* field isolates from different countries

Country	No. of sequences	No. of haplotypes	Fu and Li's (17) $F^*$	Tajima's (35) $D$	Avg (SE) $d^a$
Brazil	8	2	-1.80	-1.53	0.005 (0.002)
Venezuela	10	7	0.86	0.65	0.030 (0.009)
Kenya	18	13	-0.16	-0.17	0.027 (0.007)
Cameroon	9	7	-0.81	-0.84	0.021 (0.006)
Gambia	44	21	1.36	0.77	0.027 (0.007)
Senegal	10	8	1.14	0.52	0.021 (0.007)
India	11	3	-0.63	-1.03	0.004 (0.003)
Vanuatu	6	2	-1.20	-1.13	0.003 (0.002)
Thailand	23	8	-0.62	-0.52	0.014 (0.004)
Kanbauk <sup>b</sup>	6	3	-1.40	-1.30	0.006 (0.003)
Thaton <sup>b</sup>	19	3	-0.58	-0.82	0.005 (0.002)
Indonesia	36	5	-1.78	-1.93 <sup>c</sup>	0.003 (0.001)
Vietnam	142	20	0.43	-0.80	0.009 (0.003)
All	342	68	-1.23	-0.96	0.016 (0.004)

<sup>a</sup> Within-population average distances ( $d$ ) and standard errors (SE) were estimated by MEGA2.1 by the bootstrap method with 1,000 replications, and the values are shown.

<sup>b</sup> Note that both Kanbauk and Thaton are located in Myanmar.

<sup>c</sup> A value statistically different from zero ( $P < 0.05$ ).

sion 4.0) (32) was used to calculate genetic diversity parameters and MEGA2.1 (25) was used to reconstruct a phylogenetic tree. A chi-square test of independence was done by treating the CSP sequences as independent entities through the use of the statistical software package SPSS (version 10.0; SPSS, Inc., Chicago, IL). To examine spatial diversity patterns, the GenBank sequences of field isolates collected from different geographical areas were also included. Of these, 6 (GenBank accession numbers AB116602 to AB116607) were from Vanuatu (36); 44 (GenBank accession numbers AY878598 to AY878641) were from The Gambia (unpublished); 48 sequences (GenBank accession numbers AF540441 to AF540488) included 11 from India, 9 from Cameroon, 10 from Venezuela, and 18 from Kenya (14); 10 isolates (GenBank accession numbers AJ269961 to AJ269970) were from Senegal, 8 (GenBank accession numbers AJ269971 to AJ269978) were from Brazil; 6 (GenBank accession numbers AJ269955 to AJ269960) were from Kanbauk, Myanmar (12); and 23 (GenBank accession numbers M83171 to M83151) were from Thailand (24). All of these sequences, together with those sequenced here (from Vietnam, Indonesia, and Myanmar), were read by DnaSP (version 4.0) and grouped into their respective countries. We used a nonparametric statistical method that uses Monte Carlo simulations to estimate the levels of significance for geographic subdivision among the nucleotide sequences, i.e., by testing of the null hypothesis of no genetic differentiation between subpopulations at different localities (20). This test is incorporated in the gene flow and genetic differentiation (21) command in DnaSP (version 4.0), and 1,000 permutations were done. Among the genetic distances that resulted from this test, Nei's Dxy value (30), which is based on the average number of pairwise nucleotide substitutions per site between populations, was exported to MEGA2.1 for reconstruction of a neighbor-joining tree. Intrapopulation average pairwise genetic distances ( $d$ ) and standard errors were also estimated by use of the Tamura-Nei ( $\gamma$ ) parameter at 1,000 replications by using MEGA2.1, and the results are shown in Table 1. Values of neutrality tests are also shown. Tajima's  $D$  test (35) is based on the differences between the number of segregating sites and the average number of nucleotide differences, whereas Fu and Li's (17)  $F^*$  statistic is based on the differences between the number of singletons (mutations appearing only once among the sequences) and the average number of nucleotide differences between pairs of sequences.

**Nucleotide sequence accession numbers.** The sequences obtained in this study have been submitted to GenBank with accession numbers DQ193573 to DQ193595.

#### RESULTS

We successfully obtained the DNA sequences of the 3' region spanning the Th2R and Th3R epitopes at the C-terminal portion of the circumsporozoite protein of *Plasmodium falciparum* isolates from 142 symptomatic malaria patients present-

TABLE 2. Ages of patients and nucleotide diversity for Th2R and Th3R epitopes of Vietnamese isolates over the years 1998 to 2003

Group	No. of isolates	Age (yr) of patients		Sequence diversity parameters ( $\pi$ [Hd]) <sup>a</sup>		
		Range	Mean $\pm$ SD	Th2R epitope	Th3R epitope	Whole 3' end
BL1998	43	11–68	30.7 $\pm$ 14.1	0.03482 (0.758)	0.02566 (0.506)	0.01383 (0.793)
BL2001	34	11–71	29.2 $\pm$ 14.2	0.02063 (0.613)	0.00878 (0.114)	0.00699 (0.613)
BL2002	33	11–66	30.9 $\pm$ 13.8	0.01649 (0.578)	0.01953 (0.411)	0.00798 (0.587)
BL2003	32	6–60	29.8 $\pm$ 13.6	0.02143 (0.696)	0.01327 (0.339)	0.00808 (0.725)
All	142	6–71	30.2 $\pm$ 13.8	0.02391 (0.675)	0.01697 (0.359)	0.00960 (0.697)

<sup>a</sup> Nucleotide diversity ( $\pi$ ) is calculated with the Jukes and Cantor correction (28), and haplotype diversity (Hd) is calculated by Nei's method (30).

ing at the Bao Loc General Hospital in southern Vietnam. The age distributions as well as the DNA polymorphism parameters for the respective groups in Bao Loc are depicted in Table 2. Among the 142 isolates sequenced from Bao Loc, we characterized 20 variants (Fig. 1) of the 3' region of the CSP gene of *P. falciparum*, of which 16 are newly identified. Fifteen haplotypes were found in BL1998 alone, during the peak transmission period, whereas six, eight, and nine alleles were found in BL2001, BL2002, and BL2003, respectively. Three variants, i.e., alleles 1 to 3 in Fig. 1, accounted for the majority of the sequences found in Bao Loc, representing 52.8%, 10.6%, and 10.6% of the Vietnamese sequences, respectively, and they appeared to be stable over our study period, as no significant difference in allele frequencies was seen between the years ( $\chi^2 = 70.003$ , degrees of freedom = 57,  $P = 0.116$ ). Of note also is that most alleles (75%) identified in Bao Loc were predominant over the years, as evidenced by the fact that only five actually “new” variants were noted. Furthermore, most of the variants are novel and have not been reported previously. These results prompted us to investigate samples from Myanmar and Indonesia, and a similar trend of allele predominance was noted, with three more variants, i.e., alleles 21 to 23, detected (Fig. 1).

In reference to the 7G8 sequence, we observed eight mutation points at residues 332, 333, 336 to 340, and 342 in the

Th2R epitope and four mutation points at residues 367, 369, 372, and 374 in the Th3R epitope (Fig. 2a and b, respectively). All polymorphisms were nonsynonymous point mutations, and at the nucleotide level, most mutations were at the first or second position of the codons. However, two third-position substitutions at residues 337 and 338 that led to K→N and I→M, respectively, were also seen, and both were within the Th2R epitope region. It is also of interest that the I→M mutation seems to have been described only in Southeast Asia (26; this study). Bhattacharya (5) reported unique substitutions at residues 331 (I→K) and 368 (K→Q) common to many clones and isolates from India, but we did not find these substitutions.

To evaluate the nature and the extent of the polymorphism patterns for parasites over the study period, we compared the sequences of the respective epitopes for each group of samples. Overall, 15 variants for the Th2R epitope (Fig. 2a) were identified, of which 11 were recorded in BL1998 alone, followed by 6 each for the BL2001 and BL2002 samples and 8 for the BL2003 samples. Excluding the 11 alleles found at the study start point (BL1998), one realizes that, in essence, only four variants were actually “new” or introduced into the population over the study period. A similar pattern was seen in the Th3R epitope (Fig. 2); again, BL1998 had the highest number of variants, i.e., seven, whereas two, four, and five alleles were

	BL 1998	BL 2001	BL 2002	BL 2003	FL 2003	FL 2004	TT 1999	Total
7G8 PSDKHIEQYLKIKNSISTEWSF CSVTCGNGIQVRIKP GSANKPKDELVDYENDIEKKICKMEK	19 (44.2%)	20 (58.8%)	21 (63.6%)	15 (46.9%)	10 (71.4%)	21 (95.5%)	15 (78.9%)	121 (61.4%)
1 .....Q.....L.....	1 (2.3%)	4 (11.8%)	2 (6.1%)	8 (25.0%)	1 (7.1%)			16 (8.1%)
2 .....MD.....L.....	4 (9.3%)	7 (20.6%)	1 (3.0%)	3 (9.4%)				15 (7.6%)
3 .....N.Q.....L.....	2 (4.7%)		4 (12.1%)	1 (3.1%)			3 (15.8%)	10 (5.1%)
4 .....QY.....L.....	4 (9.3%)	1 (2.9%)			2 (14.3%)			7 (3.6%)
5 .....TE.....Q.....L.....	2 (4.7%)	1 (2.9%)	1 (3.0%)	1 (3.1%)				5 (2.5%)
6 .....KE.....T.....Q.....L.....	1 (2.3%)		2 (6.1%)					3 (1.5%)
7 .....N.Q.....L.....	2 (4.7%)			1 (3.1%)				3 (1.5%)
8 .....E.Q.....L.....	2 (4.7%)							2 (1.0%)
9 .....TE.....Q.....L.....	1 (2.3%)			1 (3.1%)				2 (1.0%)
10 .....K.....T.....Q.....L.....		1 (2.9%)						1 (0.5%)
11 .....T.....Q.....L.....	1 (2.3%)							1 (0.5%)
12 .....T.....N.Q.....L.....	1 (2.3%)							1 (0.5%)
13 .....K.....E.Q.....L.....				1 (3.1%)				1 (0.5%)
14 .....N.Q.....L.....			1 (3.0%)					1 (0.5%)
15 .....NMQ.....L.....			1 (3.0%)					1 (0.5%)
16 .....TN.Q.....L.....				1 (3.1%)				1 (0.5%)
17 .....R.MQ.....L.....	1 (2.3%)							1 (0.5%)
18 .....RN.Q.....L.....	1 (2.3%)							1 (0.5%)
19 .....RN.Q.....L.....	1 (2.3%)							1 (0.5%)
20 .....Q.....L.....					1 (7.1%)			1 (0.5%)
21 .....Q.....L.....						1 (4.5%)		1 (0.5%)
22 .....E.....Q.....L.....							1 (5.3%)	1 (0.5%)
23 .....E.....Q.....L.....								1 (0.5%)
<b>Total</b>	<b>43</b>	<b>34</b>	<b>33</b>	<b>32</b>	<b>14</b>	<b>22</b>	<b>19</b>	<b>197</b>

FIG. 1. Proportions and nonsilent mutations observed in CSP sequences of alleles found in Bao Loc, Vietnam (1998 to 2003), as well as in Flores, Indonesia, and Thaton, Myanmar. No significant difference in allele frequencies was noted over the years ( $\chi^2 = 70.003$ , degrees of freedom = 57,  $P = 0.116$ ). The 7G8 strain (10) is used as the reference; BL, FL, and TT, Bao Loc, Flores, and Thaton, respectively. The figures adjacent to these letters representing the cities indicate the year of sample collection. The (I→M) mutation appears to be unique to Southeast Asian populations (26; this study).



(a)														Vietnam	Indonesia	Myanmar	Total				
Allele	326	326	326	326	326	326	332	333	336	337	338	339	340					341	342		
	P	S	D	K	H	I	E	Q	Y	L	K	K	I	Q	N	S	L				
7G8	CCA	AGT	GAT	AAG	CAC	ATA	GAA	CAA	TAT	TTA	AAG	AAA	ATA	AAA	AAT	TCT	ATT				
1	A	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	77 (54.2%)	33 (91.7%)	15 (78.9%)	125 (63.5%)
2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	19 (13.4%)			19 (9.6%)
3	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	15 (10.6%)	1 (2.8%)		16 (8.1%)
4	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	7 (4.9%)		3 (15.8%)	10 (5.1%)
5	...	...	...	...	...	...	AC	G	...	...	...	...	...	...	...	...	...	7 (4.9%)	2 (5.6%)		9 (4.6%)
6	...	...	...	...	...	...	A	G	...	...	...	...	...	...	...	...	...	5 (3.5%)			5 (2.5%)
7	...	...	...	...	...	...	...	...	...	...	...	G	...	...	...	...	...	3 (2.1%)			3 (1.5%)
8	...	...	...	...	...	...	...	...	...	...	...	G	...	...	...	...	...	2 (1.4%)			2 (1.0%)
9	...	...	...	...	...	...	A	...	...	...	...	...	...	...	...	...	...	1 (0.7%)			1 (0.5%)
10	...	...	...	...	...	...	AC	...	...	...	...	...	...	...	...	...	...	1 (0.7%)			1 (0.5%)
11	...	...	...	...	...	...	AC	...	...	...	...	...	...	...	...	...	...	1 (0.7%)			1 (0.5%)
12	...	...	...	...	...	...	...	A	...	...	...	G	...	...	...	...	...	1 (0.7%)			1 (0.5%)
13	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1 (0.7%)			1 (0.5%)
14	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1 (0.7%)			1 (0.5%)
15	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1 (0.7%)			1 (0.5%)
16	...	...	...	...	...	...	...	G	...	...	...	...	...	...	...	...	...			1 (5.3%)	1 (0.5%)

(b)														Vietnam	Indonesia	Myanmar	Total			
Allele	366	367	369	369	369	369	372	374	374	374	374	374	380							
	A	N	K	P	K	D	E	L	D	Y	E	N	D	I	E					
7G8	GCT	AAT	AAA	CCT	AAA	GAC	GAA	TTA	GAT	TAT	GAA	AAT	GAT	ATT	GAA					
1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	113 (79.6%)	34 (94.4%)	15 (78.9%)	162 (82.2%)	
2	...	G	...	...	...	...	C	...	...	...	...	...	...	...	...	9 (6.3%)	1 (2.8%)	3 (15.8%)	13 (6.6%)	
3	...	...	...	...	...	...	C	...	...	...	...	...	...	...	...	9 (6.3%)			9 (4.6%)	
4	...	GG	...	...	...	...	...	A	...	...	...	...	...	...	...	6 (4.2%)			6 (3.0%)	
5	...	GG	...	T	...	...	...	...	...	...	...	...	...	...	...	2 (1.4%)	1 (2.8%)		3 (1.5%)	
6	...	G	...	...	...	...	...	...	...	...	...	...	...	...	...	2 (1.4%)			2 (1.0%)	
7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...			1 (0.5%)
8	...	GG	...	...	...	...	...	...	...	...	...	...	...	...	...	1 (0.7%)		1 (5.3%)	1 (0.5%)	

FIG. 2. Nucleotide sequence variation in the Th2R (a) and Th3R (b) epitope regions of the circumsporozoite gene among *P. falciparum* field isolates collected from Southeast Asia. The numbers above indicate amino residue positions in reference to the sequence of the 7G8 strain (10).

found in BL2001, BL2002, and BL2003, respectively. However, unlike for the Th2R epitope, no new variant was noted. As shown in Fig. 2, the most common variants in Bao Loc for both epitopes appear to be the “universal” variants described by Escalante et al. (14), i.e., PSDKHIEQYLKKIQNSL for the Th2R epitope (54.2%) and GSANKPKDEL DYENDI for the Th3R epitope (79.6%). These same sequences were recently reported to be predominant alleles in western Thailand (26); however, they contrast with the sequences described in The Gambia (1). Consistent with most previous reports, we noted that the Th2R epitope appears to be more polymorphic than the Th3R epitope, a finding that is reflected both in the number of allelic types and in the diversity parameters (Table 2; Fig. 2). For instance, in each group of isolates, the diversity in Th2R is almost twice that in Th3R except for BL2001. With the exception of the Th2R and Th3R epitopes, no mutation was found in the other regions of the gene studied (Fig. 1).

**DISCUSSION**

Because CSP is one of the most widely characterized malaria vaccine candidate antigens, and the only one whose components have gone so far as completion of a successful phase IIb clinical trial (3, 6), the generation of relevant epidemiologic and immunologic data for the CSP gene is of crucial significance, but such data are lacking, particularly for regions of low malaria endemicity. The first longitudinal study on the sequence polymorphisms of the T-cell epitopes of CSP for a setting of low to moderate malaria endemicity is presented here for a 5-year observation period in southern Vietnam. The region of the gene was chosen since it has been established that most sequence variation within the CSP gene has largely been

restricted to the B- and T-cell epitopes, of which the 3’ portion is the most polymorphic (14, 22, 34). Like Doolan et al. (13) and Jongwutiwes et al. (24), we have noted that the genetic diversity within the circumsporozoite protein of *P. falciparum* isolates appears to be regionally restricted, as portrayed by both the many unique, previously undescribed sequences found in this study and the apparent temporal stability of the predominant alleles noted here. For regions of low to moderate malaria endemicity (like Bao Loc), there seems to be a major allelic type representative of the population, together with minor variants present in different proportions (26, 36; this study). This is in contrast to the situation in regions of high endemicity, like Africa, where genetic polymorphisms are reportedly high (1, 4, 27). For instance, the most common allele in Vietnam (allele 1 in Fig. 1) appears to be the major sequence reported among isolates from India (14), Thailand (24), Myanmar (12), and Vanuatu (36) and has a sequence identical to those of the T4 and T9-101 strains, both of which originated from Thailand. Most of the Brazilian sequences reported by De Stricker et al. (12), including the B1 strain, were identical to those of the 7G8 strain and three Venezuelan isolates reported by Escalante et al. (14). Our results thus seem to suggest that variation in the Th2R and Th3R epitopes of CSP in Vietnam (or presumably in Southeast Asia) may not be due to frequency-dependent selection, since the frequencies of predominant alleles were stable over the period of study, while rare variants remained at a lower frequency. Similar findings have been documented for the blood-stage malaria merozoite surface proteins in Vietnam (15) and The Gambia (9). However, the absence of variation in allele frequencies over this period of a few years alone may not completely rule out the possibility of balancing selection operating over a different

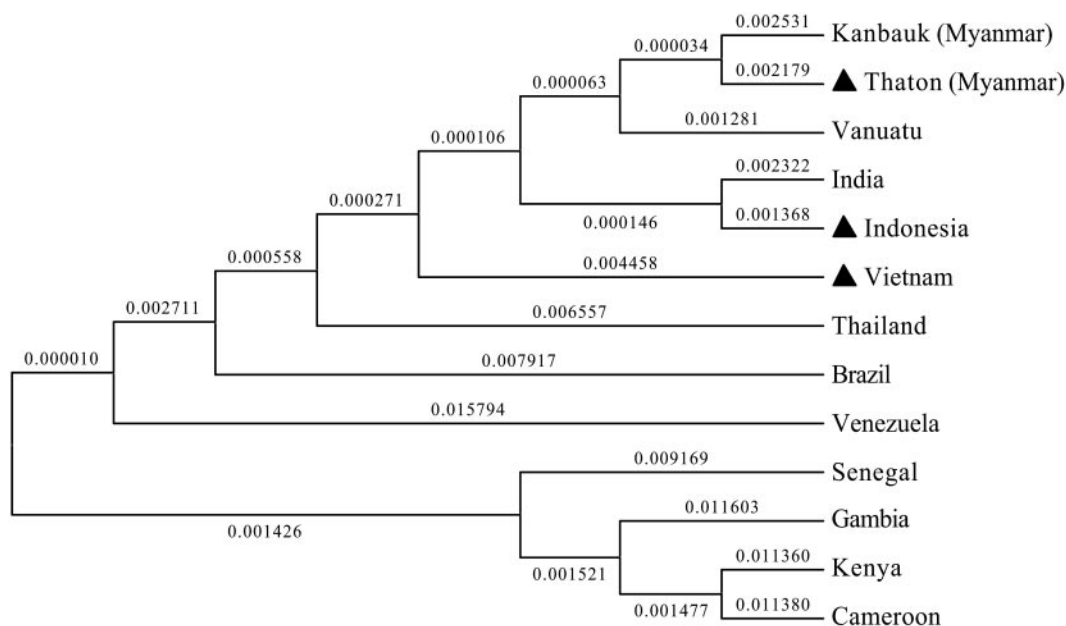


FIG. 3. A neighbor-joining tree based on population pairwise distances. Dxy is based on the average number of pairwise nucleotide substitutions per site between populations (30). Calculations were done with DnaSP (version 4.0) by using the partial *P. falciparum* CSP sequences from this study (shaded triangles) as well as from GenBank, and values were exported to MEGA2.1 for reconstruction of this tree. Note the distinct clusters among African and Southeast Asian sequences. Kanbauk and Thaton are both in Myanmar.

time scale. Moreover, it would be interesting to see temporal patterns of genetic diversity for CSP in an African setting.

Numerous reports have documented that in areas with higher rates of malaria transmission intensity the malaria isolates generally exhibit a greater amount of genetic diversity and that residents of these high-transmission areas tend to acquire antimalarial immunity at an earlier age than in areas where transmission is less intense. Also, transmission intensity is correlated with the prevalence of malaria and is a direct measure of parasite reproduction. This scenario is somewhat exemplified in our study, since most of the malaria patients presenting at the hospital were adults (Table 2), and the sequence heterogeneity observed in BL1998 alone (within a period of 4 months of the high-transmission season) was marked compared to that in the following years (Fig. 1). Although the effects of sampling or chance cannot be completely ruled out, this finding is consistent with the notion that the transmission dynamics together with other factors may be crucial in the generation and maintenance of genetic diversity within this antigen.

From this analysis of the most polymorphic part of the CSP gene, we have noted that the observed polymorphisms could be a result of host selective pressures, possibly the immune system (11, 22, 23), and importantly, there appear to be constraints on the parasite's ability to change. This probably reflects a balance between the parasite's biological and survival requirements. Within the Th2R epitope, mutations appear to be restricted both to certain positions, e.g., at residues 332 to 333 and 336 to 340, and to the nature of the replacing residue, as evidenced by the fact that all but one of the replacements were hydrophilic. The exception is the replacement with a neutral residue (I→M), which is possibly a physiologically acceptable substitution within the protein's conformational context. To date,

this residue change appears to be restricted solely to Southeast Asian populations (26; this study), suggesting a possible role for host factors. A similar pattern was noted in Th3R, although both hydrophobic and hydrophilic replacements were present. Rathore and McCutchan (31) demonstrated that the region of CSP containing the cytotoxic T-lymphocyte epitopes plays a significant role in hepatocyte binding and that an intact carboxyl end is essential for this binding process. Thus, the host immune system may recognize a sequence motif that is essential for parasite survival, implying that only mutations that could at least satisfy the dual requirements of immune evasion and parasite survival would be favored. Some studies (7, 18, 37) have indicated that amino acid variations in T-cell epitope sequences affected HLA binding, whereas others more directly affected T-cell receptor residues by inducing antagonistic peptides. Taken together, these findings suggest that polymorphisms in CSP are not merely an arms race but are part of a complex process that may be driven and maintained by a balance of opposing forces resulting from interaction between the parasite, the host, and/or other environmental factors. It would be interesting to investigate the distribution and frequency of MHC alleles and their association with local epitope variants in these populations to better understand some of these mechanisms.

Table 1 shows the sequence diversity parameters for partial sequences of the CSP gene of *Plasmodium falciparum* field isolates from different countries. Apart from the sequences from Venezuela, there seem to be distinct differences among isolates from different regions, namely, Africa, Latin America, and Asia. This is reflected in the number of haplotypes as well as average genetic distances. Grouping of the sequences into the regional categories described above (Brazil and Venezuela as Latin American; Kenya, Cameroon, The Gambia, and Sen-

gal as African; and Vietnam, Myanmar, Indonesia, Thailand, Vanuatu, and India as Asian) and independent estimates of average Wright's fixation index ( $F_{ST}$ ) values using the method of Hudson et al. (21) for each group revealed values of 0.428, 0.023, and 0.025 for the Latin American, African, and Asian parasite populations, respectively. Again, the high  $F_{ST}$  value noted for the American isolates is because of the marked diversity of the Venezuelan isolate sequences relative to that of isolates from Brazil. The computation of  $F_{ST}$  for, say, African versus Asian isolates or Asian versus Latin American isolates resulted in values up to 10 times higher than those shown above (data not shown). The average  $F_{ST}$  value for all populations combined was 0.257, a value suggestive of distinct differentiation among the subgroups. A similar result was obtained by phylogenetic analyses based on exported Dxy genetic distances (Fig. 3). Furthermore, tests of neutrality (Tajima's [35]  $D$  as well as Fu and Li's [17]  $F^*$ ) also reveal distinct patterns among African and Asian parasite populations, with the implication that the former are under balancing selection, whereas the Asian sequences all had negative values. However, caution in the interpretation of the results of these tests is necessary, since most lack statistical significance (Table 1), and because the power of these tests is affected by the number of mutations in the sample (more segregating sites have more power), the results could have also been influenced by the length of the DNA region sequenced.

In conclusion, it is apparent from this study that diversity in the *Plasmodium falciparum* CSP gene appears to be regionally restricted, as shown by the distinct differences in Asian and African parasite populations. Moreover, within different countries there seem to be unique mutation patterns. Taken together, these findings seem to suggest that regional and environmental factors, together with host genetic factors, could be crucial in generating polymorphisms, at least in the carboxy-terminal T-cell epitopes of *P. falciparum* CSP. Irrespective of what governs these polymorphisms, the absence of variation in T-cell epitope haplotype frequencies of CSP over relatively short periods of time is a very interesting finding, with major implications for malaria vaccine development (since no new variants seem to emerge within a few years of exposure to locally prevalent variants). This is probably reflected in the recent apparent success in clinical trials of a CSP-based vaccine, RTS,S/AS02 (3, 6), and the observation that protection was not strain specific (2).

#### ACKNOWLEDGMENTS

This study was financed by Grants-in-Aid for Scientific Research B2 (13570006, 15406014, 17406010) and C (14570213, 16590341) from the Japan Society for Promotion of Science to F.K.

We thank K. Lin (Myanmar) and I. S. Tantular, S. Puserawati, Y. P. Dachlan, and H. Kerong (Indonesia) for their help during surveys of malaria and T. Isomoto (Japan) for helpful suggestions and support.

#### REFERENCES

- Allouche, A., H. Silveira, D. J. Conway, K. Bojang, T. Doherty, J. Cohen, M. Pinder, and B. M. Greenwood. 2000. High-throughput sequence typing of T-cell epitope polymorphisms in *Plasmodium falciparum* circumsporozoite protein. *Mol. Biochem. Parasitol.* **106**:273–282.
- Allouche, A., P. Milligan, D. J. Conway, M. Pinder, K. Bojang, T. Doherty, N. Tornieporth, J. Cohen, and B. M. Greenwood. 2003. Protective efficacy of the RTS,S/AS02 *Plasmodium falciparum* malaria vaccine is not strain specific. *Am. J. Trop. Med. Hyg.* **68**:97–101.
- Alonso, P. L., J. Sacarlal, J. J. Aponte, A. Leach, E. Macete, J. Milman, I. Mandomando, B. Spiessens, C. Guinovart, M. Espasa, Q. Bassat, P. Aide, O. Ofori-Anyanam, M. M. Navia, S. Corachan, M. Ceuppens, M. C. Dubois, M. A. Demoitie, F. Dubovsky, C. Menendez, N. Tornieporth, W. R. Ballou, R. Thompson, and J. Cohen. 2004. Efficacy of the RTS,S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial. *Lancet* **364**:1411–1420.
- Anderson, T. J., B. Haubold, J. T. Williams, J. G. Estrada-Franco, L. Richardson, R. Mollinedo, M. Bockarie, J. Mokili, S. Mharakurwa, N. French, J. Whitworth, I. D. Velez, A. H. Brockman, F. Nosten, M. U. Ferreira, and K. P. Day. 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**:1467–1482.
- Bhattacharya, P. R. 1999. Genetic polymorphism in T-cell epitopes of the circumsporozoite protein of *Plasmodium falciparum* clones and isolates from India. *Trans. R. Soc. Trop. Med. Hyg.* **93**:204–207.
- Bojang, K. A., P. J. Milligan, M. Pinder, L. Vigneron, A. Allouche, K. E. Kester, W. R. Ballou, D. J. Conway, W. H. Reece, P. Gothard, L. Yamuah, M. Delchambre, G. Voss, B. M. Greenwood, A. Hill, K. P. McAdam, N. Tornieporth, J. D. Cohen, T. Doherty, and RTS, S Malaria Vaccine Trial Team. 2001. Efficacy of RTS,S/AS02 malaria vaccine against *Plasmodium falciparum* infection in semi-immune adult men in The Gambia: a randomised trial. *Lancet* **358**:1927–1934.
- Bonelo, A., D. Valmori, F. Triponez, J. M. Tiercy, G. Mentha, J. Oberholzer, P. Champagne, J. F. Romero, F. Esposito, I. Nebié, C. Barbey, P. Romero, S. Herrera, G. Corradin, and J. A. López. 2000. Generation and characterization of malaria-specific human CD8<sup>+</sup> lymphocyte clones: effect of natural polymorphism on T cell recognition and endogenous cognate antigen presentation by liver cells. *Eur. J. Immunol.* **30**:3079–3088.
- Breman, J. G. 2001. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **64**:1–11.
- Conway, D. J., B. M. Greenwood, and J. S. McBride. 1992. Longitudinal study of *Plasmodium falciparum* polymorphic antigens in a malaria-endemic population. *Infect. Immun.* **60**:1122–1127.
- Dame, J. B., J. L. Williams, T. F. McCutchan, J. L. Weber, R. A. Wirtz, W. T. Hockmeyer, W. L. Maloy, J. D. Haynes, I. Schneider, D. Roberts, G. S. Sanders, E. P. Reddy, C. L. Diggs, and L. H. Miller. 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* **225**:593–599.
- De La Cruz, V. F., W. L. Maloy, L. H. Miller, M. F. Good, and T. F. McCutchan. 1989. The immunologic significance of variation within malaria circumsporozoite protein sequences. *J. Immunol.* **142**:3568–3575.
- De Stricker, K., J. Vuust, S. Jepsen, C. Oeuvray, and M. Theisen. 2000. Conservation and heterogeneity of the glutamate-rich protein (GLURP) among field isolates and laboratory lines of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **111**:123–130.
- Doolan, D. L., A. J. Saul, and M. F. Good. 1992. Geographically restricted heterogeneity of the *Plasmodium falciparum* circumsporozoite protein: relevance for vaccine development. *Infect. Immun.* **60**:675–682.
- Escalante, A. A., H. M. Grebert, R. Isea, I. F. Goldman, L. Basco, M. Magris, S. Biswas, S. Kariuki, and A. A. Lal. 2002. A study of genetic diversity in the gene encoding the circumsporozoite protein (CSP) of *Plasmodium falciparum* from different transmission areas. XVI. Asembo Bay Cohort Project. *Mol. Biochem. Parasitol.* **125**:83–90.
- Ferreira, M. U., Q. Liu, M. Zhou, M. Kimura, O. Kaneko, H. Van Thien, S. Isomura, K. Tanabe, and F. Kawamoto. 1998. Stable patterns of allelic diversity at the merozoite surface protein-1 locus of *Plasmodium falciparum* in clinical isolates from southern Vietnam. *J. Eukaryot. Microbiol.* **45**:131–136.
- Ferreira, M. U., W. L. Ribeiro, A. P. Tonon, F. Kawamoto, and S. M. Rich. 2003. Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*. *Gene* **304**:65–75.
- Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- Gilbert, S. C., M. Plebanski, S. Gupta, J. Morris, M. Cox, M. Aidoo, D. Kwiatkowski, B. M. Greenwood, H. C. Whittle, and A. V. Hill. 1998. Association of malaria parasite population structure, HLA, and immunological antagonism. *Science* **279**:1173–1177.
- Gonzalez, J. M., K. Peter, F. Esposito, I. Nebie, J. M. Tiercy, A. Bonelo, M. Arevalo-Herrera, D. Valmori, P. Romero, S. Herrera, G. Corradin, and J. A. Lopez. 2000. HLA-A\*0201 restricted CD8<sup>+</sup> T-lymphocyte responses to malaria: identification of new *Plasmodium falciparum* epitopes by IFN- $\gamma$  ELISPOT. *Parasite Immunol.* **22**:501–514.
- Hudson, R. R., D. D. Boos, and N. L. Kaplan. 1992. A statistical test for detecting population subdivision. *Mol. Biol. Evol.* **9**:138–151.
- Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**:583–589.
- Hughes, A. L. 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* **127**:345–353.

23. Hughes, M. K., and A. L. Hughes. 1995. Natural selection on *Plasmodium* surface proteins. *Mol. Biochem. Parasitol.* **71**:99–113.
24. Jongvutives, S., K. Tanabe, M. K. Hughes, H. Kanbara, and A. L. Hughes. 1994. Allelic variation in the circumsporozoite protein of *Plasmodium falciparum* of Thai field isolates. *Am. J. Trop. Med. Hyg.* **51**:659–668.
25. Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244–1245.
26. Kumkhaek, C., K. Phra-ek, P. Singhasivanon, S. Looareesuwan, C. Hirunpetcharat, A. Brockman, A. C. Gruner, N. Lebrun, L. Renia, F. Nosten, G. Snounou, and S. Khumsmith. 2004. A survey of the Th2R and Th3R allelic variants in the circumsporozoite protein gene of *Plasmodium falciparum* parasites from western Thailand. *Southeast Asian J. Trop. Med. Public Health* **35**:281–287.
27. Lockyer, M. J., K. Marsh, and C. I. Newbold. 1989. Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epitopes of the circumsporozoite protein. *Mol. Biochem. Parasitol.* **37**:275–280.
28. Lynch, M., and T. J. Crease. 1990. The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* **7**:377–394.
29. Nardin, E. H., G. A. Oliveira, J. M. Calvo-Calle, Z. R. Castro, R. S. Nussenzweig, B. Schmeckpeper, B. F. Hall, C. Diggs, S. Bodison, and R. Edelman. 2000. Synthetic malaria peptide vaccine elicits high levels of antibodies in vaccinees of defined HLA genotypes. *J. Infect. Dis.* **182**:1486–1496.
30. Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York, N.Y.
31. Rathore, D., and T. F. McCutchan. 2000. The cytotoxic T-lymphocyte epitope of the *Plasmodium falciparum* circumsporozoite protein also modulates the efficiency of receptor-ligand interaction with hepatocytes. *Infect. Immun.* **68**:740–743.
32. Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**:2496–2497.
33. Saitfri, I., A. Jalloh, I. S. Tantular, S. Puserawati, T. T. Win, Q. Liu, M. U. Ferreira, Y. P. Dachlan, T. Horii, and F. Kawamoto. 2003. Sequence diversity in the amino-terminal region of the malaria-vaccine candidate serine repeat antigen in natural *Plasmodium falciparum* populations. *Parasitol. Int.* **52**:117–131.
34. Shi, Y. P., M. P. Alpers, M. M. Pova, and A. A. Lal. 1992. Diversity in 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.
35. Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
36. Tanabe, K., N. Sakihama, and A. Kaneko. 2004. Stable SNPs in malaria antigen genes in isolated populations. *Science* **303**:493.
37. Udhayakumar, V., J. M. Ongecha, Y. P. Shi, M. Aidoo, A. S. Orago, A. J. Oloo, W. A. Hawley, B. L. Nahlen, S. L. Hoffman, W. R. Weiss, and A. A. Lal. 1997. Cytotoxic T cell reactivity and HLA-B35 binding of the variant *Plasmodium falciparum* circumsporozoite protein CD8<sup>+</sup> CTL epitope in naturally exposed Kenyan adults. *Eur. J. Immunol.* **27**:1952–1957.
38. Verle, P., T. C. Tuy, T. T. Lieu, A. Kongs, and M. Coosemans. 1998. New challenges for malaria control in northern Vietnam. *Res. Rev. Parasitol.* **58**:169–174.
39. Wang, R., D. L. Doolan, T. P. Le, R. C. Hedstrom, K. M. Coonan, Y. Charoenvit, T. R. Jones, P. Hobart, M. Margalith, J. Ng, W. R. Weiss, M. Sedegah, C. de Taisne, J. A. Norman, and S. L. Hoffman. 1998. Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine. *Science* **282**:476–480.
40. Yoshida, N., R. S. Nussenzweig, P. Potocnjak, V. Nussenzweig, and M. Aikawa. 1980. Hybridoma produces protective antibodies directed against the sporozoite stage of malaria parasite. *Science* **207**:71–73.
41. Zavala, F., A. H. Cochrane, E. H. Nardin, R. S. Nussenzweig, and V. Nussenzweig. 1983. Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. *J. Exp. Med.* **157**:1947–1957.