SEQUENCE DIVERSITY OF SERINE REPEAT ANTIGEN GENE EXON II OF *PLASMODIUM FALCIPARUM* IN WORLDWIDE COLLECTED WILD ISOLATES

Qing Liu[1,7,a], Marcelo U Ferreira[2], Benedict T Ndawi[3], Hiroshi Ohmae[4], Ipemida S Adagu[5], Tomohiro Morikawa[6], Toshihiro Horii[6], Shin Isomura[1] and Fumihiko Kawamoto[1,b]

[1]Department of International Health, Nagoya University School of Medicine, Nagoya, Japan; [2]Department of Parasitology, Institute of Biomedical Sciences, University of San Paulo, San Paulo, Brazil; [3]Primary Health Care Institute, Iringa, Tanzania; [4]Department of Medical Zoology, Institute of Basic Medical Sciences, University of Tsukuba, Tsukuba, Japan; [5]London School of Hygiene and Tropical Medicine, London, UK; [6]Department of Molecular Protozoology, Research Institute of Microbial Diseases, Osaka University, Osaka, Japan; [7]Department of Parasitology, West China University of Medical Sciences, Chengdu, China

Abstract. Field isolates of *Plasmodium falciparum* collected from endemic areas of Southeast Asia, Solomon Islands, tropical African countries and Brazil were analyzed for the genetic diversity of the exon II of serine repeat antigen gene (SERA) by sequencing of genomic DNA. Of sixty-nine isolates, as compared to the reported FCR3, K1 and Honduras-1 types of exon II sequences, 5, 9 and 20 new allelic forms were found in 23 isolates of the FCR3 type, 36 of the K1 type and 10 of the Honduras-1 type. A group of novel non-synonymous substitutions, 4 new insertions and 3 new deletions of octamer units were found in the octamer repeat region (OR) of the exon II, and most of them clustered within a 40-residues domain. An octamer "SNPVSSEP" revealed in the OR was confirmed as a new repeat unit. Based on the sequences of the serine repeat region (SR) of the exon II, the allelic forms of the Honduras-1 type were conjectured to be the recombinant forms between the K1 type and FCR3 type. The allelic forms of K1 type with less or more repeat serine residues in the serine stretch of the SR than the reported 21 serine residues had most of the variations in the OR. Moreover, a biased geographical distribution of allelic forms was observed. Isolates from African and Southeast Asian countries accounted for most of the new allelic forms (29/33). All of the three types were detected in Southeast Asia but none of the FCR3 type in Africa. One of two groups of FCR3 new allelic forms was found solely in Brazil while another was mainly in Solomon Islands.

INTRODUCTION

Control of malaria represents one of the world's greatest health challenges and never before has a malaria vaccine been more urgently required, especially as an estimated 300-500 million people are infected with 2.3 million deaths per year (WHO, 1995). Among the tens of target antigens investigated for malaria vaccine development, serine repeat antigen (SERA), a protein located in the parasitophorous vacuole of mature schizont stage parasites (Delplace *et al*, 1987, Knapp *et al*, 1989) and on the surface of free merozoites (Perkins and Ziefer, 1994), is highlighted as one of the leading vaccine candidates by the World Health Organization for introduction to human immunization trials (Engers and Godal, 1998). SERA protein is synthesized in the schizont stage, and processed into 56- and 73-kDa fragments with removal of a 23-amino acid signal sequence and secreted into parasitophorous vacuole lumen (Delplace *et al*, 1987, 1988; Ragge *et al*, 1990; Debrabant *et al*, 1992). N-terminal sequencing showed that the 73-kDa fragment is further processed, at the merozoite release and reinvasion stage, into a 47-kDa amino-terminal portion and a disulfide linked 18-kDa carboxy-terminus (Debrabant and Delplace, 1989; Debrabant *et al*, 1992) (Fig 1). As a potential vaccine antigen, antibodies against SERA were capable of inhibiting the invasion of erythrocytes by *P. falciparum in vitro* (Barr *et al*, 1991) or inhibiting parasite growth in culture (Sugiyama *et al*, 1996, Pang and Horii, 1998), and protective immunity in monkeys (Inselburg *et al*, 1991; 1993a,b; Suzue *et al*, 1997) and mice (Gor *et al*, 1998) were

[a]Current address: Dr Qing Liu, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT 06520, USA.
Tel: 1-203-787-7974
E-mail: qing.liu@yale.edu
[b]Correspondence: Dr F Kawamoto, Department of International Health, Nagoya University School of Medicine, Nagoya, Japan.
Tel: 81-52-7442113; Fax: 81-52-7442114
E-mail: hiko@med.nagoya-u.ac.jp

induced by immunization with intact SERA and recombinant amino-terminal SERA fragments. A poxvirus-vectored, multi-antigen, multistage vaccine candidate containing the recombinant SERA gene (Tine *et al*, 1996) has been reported recently.

Numerous studies have demonstrated that many vaccine candidate antigens of *P. falciparum*, such as circumsporozoite protein (CSP) and merozoite surface antigen 1 (MSA-1) are genetically polymorphic and that their allelic forms differ in the ability to abrogate recognition of the host's immune response (Udhayakumar *et al*, 1994; Zevering *et al*, 1994; Babiker and Walliker, 1997). Sequences of *P. falciparum* SERA gene have been reported from several laboratory-maintained strains (Weber *et al*; 1987; Bzik *et al*; 1988; Horii *et al*, 1988; Knapp *et al*, 1989; Li *et al*, 1989; Fox and Bzik, 1994; Morimatsu *et al*, 1997). But there is no reported data on the genetic diversity of wild isolates. In a comparison of three complete and two partial sequences of the SERA gene derived from cultivated strains of *P. falciparum*, Fox *et al* (1997) demonstrated that this gene is extremely well conserved, especially in the central and C-terminal portions (Fig 1). Morimatsu *et al* (1997) analyzed genetic diversity of the N-terminal 47-kDa fragment of the SERA gene from 7 cultivated strains and described that the deletion/insertion events, rather than point mutations, account for the majority of its limited diversity. Recently, a novel epitope within a highly conserved region of the N-terminus of the 47-kDa fragment has been reported as a target of parasite-inhibitory antibodies (Fox *et al*, 1997). It is presumed that the genetic variations in natural parasite populations may occur more frequently than those in the culture-adapted strains because of immunological selective pressure in endemic areas. In this study, field isolates of *P. falciparum* collected worldwide were analyzed for the genetic diversity of SERA gene in natural populations by PCR amplification and direct sequencing the SERA exon II which encodes the major part of the N-terminal 47-kDa fragment.

MATERIALS AND METHODS

Samples collection and DNA preparation

Samples were collected from different endemic areas scattered in the world, *ie* 11 isolates from Rondonia, Porto Velho, southwestern Brazilian Amazon; 23 isolates from hospitals in Bao Loc Province and Ho Chimin City, and from a survey conducted in Gia Lai Province, Vietnam; 6 from the Solomon Islands; 14 from health centers and hospitals in Tanga, Tanzania; and 15 from endemic areas in Nigeria, Gabon, Guinea Bissau, Indonesia, Thailand and Myanmar. Patients in local hospitals and villagers in surveyed villages were diagnosed by microscopic examination of Giemsa-stained thick smears or acridine orange-stained thin smears (Kawamoto, 1991). The *P. falciparum* infected patients were administered anti-malaria medicine under the direction of local hospitals. For most of the samples, venous blood (0.2-1.0 ml) was collected from patient volunteers after informed consent, while finger-prick blood samples on filter paper were obtained in west African countries. All of the venous blood samples were stored and transported at -30ºC or with ice pack until DNA extraction. Each of the dried blood spot on filter paper was kept in a separate plastic bag and transported at room temperature.

After thawing, 200 μl of anti-coagulated blood were lysed in 50 μl of 0.15% saponin in phosphate-buffered saline (PBS, pH 7.4), washed with PBS, and re-suspended in 200 μl of TEN buffer (10 mM Tri-HCl pH 8.0, 5 mM EDTA and 100 mM NaCl) with 0.5% SDS-200 μg/ml of Proteinase K (Boehringer Mannheim, Germany) for an overnight incubation at 56ºC. Genomic DNA was extracted by standard phenol-chloroform method, re-dissolved in TE buffer (10 mM Tris-HCl, 5 mM EDTA pH 8.0) and stored at -30ºC until use. Blood spots on the filter paper were well excised and mixed by vortex with 150 μl of 30% (W/V) Chelex-100 beads (Bio-Rad, Richimond, CA, USA) at pH 10, then subjected to boiling in water bath for 10 minutes and centrifuged at 20,000*g* for 10 minutes at 4ºC. The supernatant was transferred to a new tube and used as PCR template (Kimura *et al*, 1995).

PCR amplification and purification of the PCR products

Three sets of oligo-nucleotide primers were designed from the reported sequences of the SERA gene, flanking two variable regions and the conserved region between them (Fig 1). Amplification was performed in a 25 μl of reaction mixture containing 0.5 μM each of forward and reverse primers, 1-3 μl of template DNA, 2.5 μl of 10 x buffer, 200 μM of each dNTP, and 0.5 units of *Taq* DNA polymerase (Pharmacia Biotech, USA). The thermal cycling profile was 3 minutes at 94ºC for the first cycle, 45 seconds at 94ºC, 1 minute at 57ºC and

3 minutes at 72ºC for 38 cycles, and 10 minutes at 72ºC for the last cycle. The anticipated sizes of PCR products with reference to the sequence of Honduras-1 strain are 227 bp, 310 bp, and 304 bp corresponding to the octamer repeat region (OR), conserved region and serine repeat region (SR), respectively (Fig 1). Genomic DNA from HB3, K1 and FCR3 strains were used as positive controls. Distilled water and genomic DNA from the blood of *P. vivax*-infected patient were used as the negative controls. Single band of PCR products checked by electrophoresis on 0.65% SynerGel - 7% agarose gel (Diversified Biotech, and Gibco BRL, Grand Island, NY, USA) was directly transferred to a filter tube for purification by High Pure PCR Product Purification Kit (Boehringer Mannheim, Germany), following the protocol of supplier. DNA concentration was measured by spectrophotometer (A_{260}).

DNA sequencing and data analysis

0.2-0.4 μg of purified DNA were directly sequenced in a 373 DNA sequencing system (PEApplied Biosystem, USA) by using a ABI PRISM™ Dye Terminator Cycle Sequencing Ready Kit and following the protocol of the supplier. Sequencing primers were the same as those for PCR amplification, and both forward and reverse strands of the three fragments were sequenced.

Nucleotide sequences were read out by ABI PRISM™ DNA Sequencing Software version 2.1.2 (PE Applied Biosystem, USA) and compared with the known data from the GenBank/EMBL/DDBJ database for identifying the type of the SERA gene exon II. For diversity analysis, sequences were aligned based on maximum matching of nucleotides by using the GENETYX-MAC/Homology/ Malign program (Software Development Co, Ltd), with corrections made by eye. The sequences were aligned by creating gaps as needed so that identical amino acid motifs align with one another, while the corresponding nucleotides are the characters. The nucleotide sequences data reported in this paper will appear in the DDBJ/EMBL/ GenBank nucleotide sequence database with the accession numbers AB021821-AB021856.

RESULTS

Based on the sequences derived from cultured strains (Weber *et al*, 1987; Bzik *et al*, 1988; Horii *et al*, 1988; Knapp *et al*, 1989; Li *et al*, 1989; Fox *et al*, 1994, 1997; Morimatsu *et al*, 1997), the variable regions of the SERA exon II can be subdivided into an octamer repeat region (OR) and a serine repeat region (SR)(Fig 1). The OR consist of a "repeat 1" which has 3 to 4 octamer units and a "repeat 2" with 1 or 3 octamer units (Table 1, Fig 1). The SR is characterized by a stretch of repeat serine residues, and in some strains, a 13-amino acid insertion (insert E in Table 1) in its upstream and a cluster of 12 non-synonymous substitutions (residue 230 to 246 in Table 1) in downstream of the stretch. By the structural features in the SR, the N-terminal of 47-kDa fragment of the *P. falciparum* SERA from cultured strains was classified as three types: FCR3 type (HB3, SL3 and FCR3 strains) without both the 13-amino acid insertion and the cluster of 12 non-synonymous substitutions; K1 type (Camp, T9-102, PA7 and K1 strains) with both of them; and Honduras-1 type (T9-96, 3D7 and Honduras-1 strains) with the 13-amino acid insert but without the clustered non-synonymous substitutions (Table 1 and Morimatsu *et al*, 1997). In this study, the sequence of T9-102 strain was listed under the K1 type for the analysis because three isolates had sequences identical to that of the T9-102 strain but no sequence identical to that of K1 strain was detected in wild isolates.

Overall diversity of the SERA exon II

Nucleotide sequences derived from 69 wild isolates from global endemic areas showed that the genetic diversity of the SERA exon II mainly gathered in the OR and the SR, while several nucleotide mutations were detected in the conserved region between them. In Table 1, analysis of expected amino acid sequences revealed that the insertion/ deletion of octamer units frequently occurred in the OR, and that point mutations of nucleotides resulted in non-synonymous substitutions in the OR and the conserved region. 14 non-synonymous substitutions at amino acid residues 29, 31, 42, 48, 50, 52, 54, 57, 58, 60, 65, 67, 74 and 246, 4 insertions of the octamer units between residue 41-42, 65-66, and 73-74 (Insert A, B and D in Table 1) and 3 deletions of the octamer units between residue 34-49, 42-49 and 58-65 (Table 1) were identified as novel variations never reported before. The insertion of octamer unit "SNPVSSEP" (Insert D), as a first report, revealed the presence of a new octamer unit which forms "repeat 3" by two units in the OR (Fig 1 and Table 1).

Length polymorphism in which the number of the repeat serine residues varied from 15 to 53

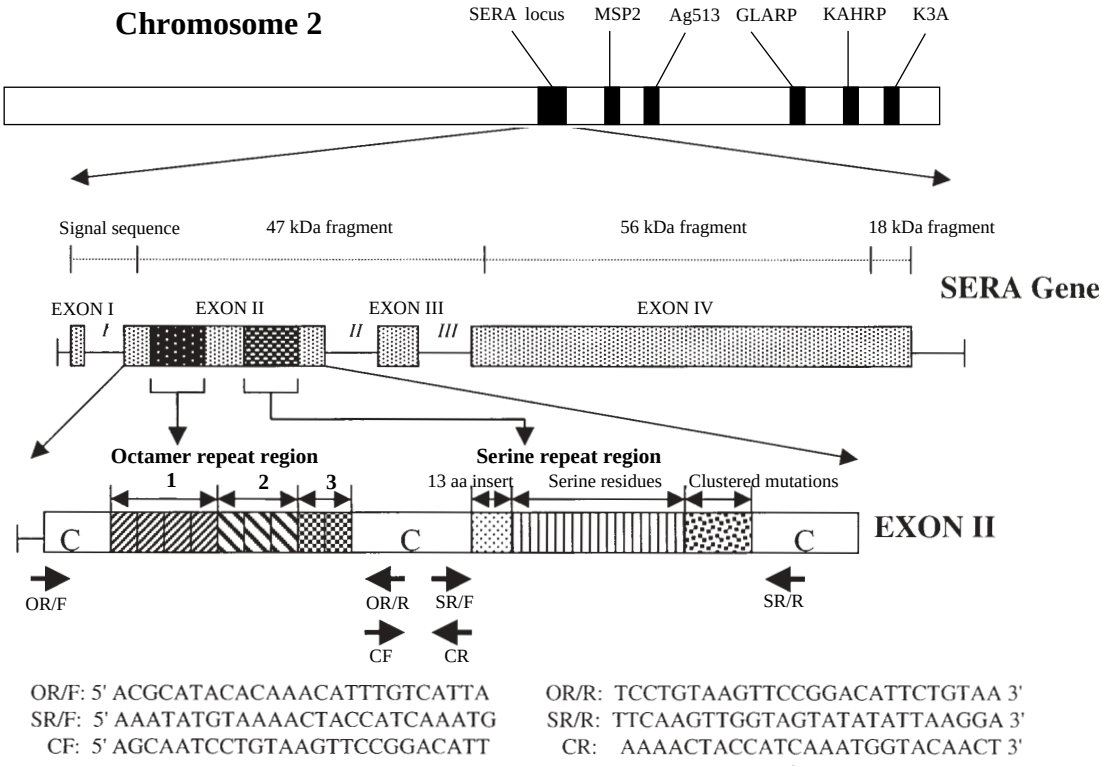

Fig 1–Schematic representation of *Plasmodium falciparum* SERA gene exon II. In SERA gene strip, shaded bars are exons and the lines marked with italic roman letters are the introns of the gene. In exon II strip, shaded areas between either two of the three blank bars are the variable regions, including the octamer repeat region (OR) in dark or light slant lined bars, and the serine repeat region (SR) in pointed or vertical lined bars, respectively. Slant lined bars marked with 1 and 2 indicate the units of the repeat 1 and repeat 2 in OR. The bars marked with 3 are the units of repeat 3 reported by the present study. The sequences marked with OR/F, OR/R, SR/F, SR/R, CF and CR are primers for PCR amplification and sequencing. Short arrows show the annealing sites of the primers in conservative regions marked with italic C.

was observed in the serine stretch of the SR (Table 1). Nucleotide alternations in the serine stretch induced only synonymous substitutions with an exception of residue 193 (Asn→Ser) (Table 1). However, when the variations of nucleotide sequences cording the SS (Ser-Ser) motif and the SN (Ser-Asn) motif in the serine stretch were abbreviated to A, B, C and D, new sequence patterns which we refer to as allotype patterns of isolates resulted (Fig 2). The three types of the SERA Exon II are distinct from each other by their unique allotype patterns in this domain. As the allotype patterns of the K1 type are characterized by the combination of "CAAC", "BCA" and "BCC. . . . . ." structures and those of the FCR3 type are characterized by the "BA. . . . . ." structure, allotype patterns of Honduras-1 type were conjectured to be hybrids between the K1 and FCR3 types because they resemble the K1 type in the 3′ region and the FCR3 type in the 5′ region, respectively (Fig 2). Two positions within the serine stretch and one position immediately adjacent to the serine stretch were conjectured to be recombinant sites (Fig 2). 5, 8 and 9 allotype patterns were detected in the FCR3, Honduras-1 and K1 types, respectively. But, sequences of the isolates with same allotype pattern of the serine stretch may be different in the OR and the conserved regions (Table 1, Fig 2).

Among the 69 isolates, according to the differences of non-synonymous substitutions, the insertion/deletion of the octamer units and the length of the serine stretch in their sequences, 5, 9 and 20 new allelic forms of FCR3 type, Honduras-1 type and K1 type were detected in 23, 10, and 36 isolates, respectively. However, 6 and 3 isolates

Table 1

Genetic diversity of *Plasmodium falciparum* SERA exon II in wild isolates worldwide.

Allelic forms[b]	Octamer repeat region (OR): Repeat 1 (29[a] 31 / 40 / Insert A / 42 / 48 50 52 54 / 57)	Repeat 2 (58 60 / 65 / Insert B / Insert C)	Repeat 3 (67 / Insert D)	Serine repeat region (SR): 74	151	180	183	Insert E	193	S#	230-32 / 237-8 / 240 - 246
FCR3 (6)	SQTGNTGGGQAGNTGG--------GQAGNTVGDQAGSTGG	SPQGSTGA----------------	SQPGSSEP--------	S	E	G	S	-------------	N	35	ESLPANGPDSPTVKPPR
F1 (3)	................--------................	........----------------	........--------	.	.	.	.	-------------	.	39	....
F2 (3)	..............V.--------D...............	........----------------	........--------	.	.	.	.	-------------	.	37	....
F3 (3)	..............V.--------D...............	........----------------	........--------	.	.	.	.	-------------	.	51	....
F4 (5)	..............V.--------D...............	........----------------	........--------	.	K	.	.	-------------	.	51	....
F5 (3)	..............V.--------D...............	........----------------	........--------	.	K	.	.	-------------	.	53	....
Hond-1 (0)	........----------------D...............	........----------------	........--------	.	.	.	G	TGTVRGDTEPISD	S	35	....
H1 (2)	........------------------------........	........----------------	........--------	.	.	.	G	-------------	S	33	....
H2 (1)	................GQAGNTGG......G...P.....	------------------------	........--------	I	.	.	G	....	S	35	....
H3 (1)	..............V.--------................	........----------------	........--------	.	.	.	G	....	S	37	....
H4 (1)	................--------................	.......E----------------	........--------	.	K	.	G	....	S	39	....
H5 (1)	................--------................	........----------------	........--------	.	K	.	G	.........S...	S	31	....
H6 (1)	..............V.--------................	........----------------	........--------	.	.	.	G	.........S...	S	31	....
H7 (1)	..............V.--------................	------------------------	........--------	I	.	.	G	.........S...	S	31	....
H8 (1)	................--------................	........----------------	........--------	.	K	.	G	.........S...	S	31	VNP..............
H9 (1)	................----------------........	........SPQGSTGASPQGSTGA	........--------	.	K	.	G	.........S...	S	15	VNP..............
K1 (0)	..............V.--------V...............	........----------------	........--------	.	.	S	G	....	S	21	VNP....AG.TPDAKKK
T9-102 (3)	..............V.--------................	........----------------	........--------	.	.	.	G	....	S	21	VNP....AG.TPDAKKK
T1 (1)	..............V.----------------........	........----------------	........--------	.	.	.	G	....	S	21	VNP....AG.TPDAKKK
T2 (1)	................----------------........	........----------------	........--------	.	.	.	G	....	S	21	VNP....AG.TPDAKKK
T3 (5)	................----------------........	........----------------	........--------	.	.	.	G	.........S...	S	21	VNP....AG.TPDAKKK
T4 (5)	..............V.--------................	........----------------	........--------	.	.	.	G	.........S...	S	21	VNP....AG.TPDAKKK
T5 (2)	..............V.--------................	------------------------	........--------	.	.	.	G	.........S...	S	21	VNP....AG.TPDAKKK
T6 (1)	................--------................	........----------------	........--------	.	.	.	G	.........S...	S	21	VNP....AG.TPDAKKK
T7 (4)	................--------................	........----------------	........--------	.	K	.	G	.........S...	S	21	VNP....AG.TPDAKKK
T8 (1)	................--------................	........----------------	........--------	.	K	.	G	.........S...	S	21	VNP....AG.TPDAKK.
T9 (1)	................--------................	........----------------	........--------	.	K	.	G	.........S...	S	15	VNP....AG.TPDAKKK
T10 (2)	..............V.--------D...............	........----------------	........--------	.	K	.	G	.........S...	S	15	VNP....AG.TPDAKKK
T11 (1)	................----------------........	........SPQGSTGA--------	........--------	.	K	.	G	....	S	15	VNP....AG.TPDAKKK
T12 (1)	................----------------........	..P.....SPQGSTGA--------	........--------	.	K	.	G	....	S	23	VNP....AG.TPDAKKK
T13 (1)	................----------------G.......	........----------------	........SNPVSSEP	.	K	.	G	....	S	23	VNP....AG.TPDAKKK
T14 (1)	..............V.--------D...............	........----------------	........--------	.	K	.	G	.........S...	S	23	VNP....AG.TPDAKKK
T15 (1)	...R..........V.DQAGNTVGD...............	........----------------	........--------	.	K	.	G	.........S...	S	23	VNP....AG.TPDAKKK
T16 (1)	................------------------...N..	N......G----------------	.P......--------	.	K	.	G	.........S...	S	25	VNP....AG.TPDAKKK
T17 (1)	................--------................	........----------------	........--------	.	K	.	G	.........S...	S	31	VNP....AG.TPDAKKK
T18 (1)	..............V.--------D....G..G.......	........----------------	........--------	.	.	.	G	.........S...	S	31	VNP....AG.TPDAKK.
T19 (1)	................GQAGNTGG................	........----------------	........--------	.	.	.	G	.........S...	S	31	VNP....AG.TPDAKKK
T20 (1)	.....A........V.--------D...............	........----------------	........--------	.	K	.	G	.........S...	S	33	VNP....AG.TPDAKKK

[a] Numerals, with reference to the sequence of FCR3 strain, indicate the locations of non-synonymous substitutions of amino acid residues (in italic boldface letters).

[b] Sequences of FCR3, Honduras-1 (Hond-1) and K1 are used as representatives of three types of SERA exon II, while that of T9-102 strain is listed for analysis. F-, H- and T- are used for new allelic forms of each type. Numerals in parentheses are the number of isolate(s) with identical sequences. Amino acid residue identity is indicated by a dot; the lack of a corresponding residue is indicated by a dash.

S# Numerals in this column are the numbers of repeat serine residues in the serine stretch.

were identical to the reported sequences of the FCR3 and T9-102 strains, respectively (Table 1).

Diversity within each of the three types

Five new allelic forms of the FCR3 type were detected from 17 isolates, having longer serine stretches with 37, 39, 51 and 53 repeat serine residues than the reported 35 residues in the SR, two non-synonymous substitutions at residue 40 (Gly→Val) and 42 (Gly→Asp) in the OR, and one non-synonymous substitution at residue 151 (Glu→Lys) (F1-F5 in Table 1). Other 6 isolates had sequences identical to that of the FCR3 strain.

Higher diversity was found in allelic forms of the K1 type. Of 69 wild isolates, 36 were identified as the K1 type. Sequence analysis showed that, except 3 isolates with identical sequence to that of the T9-102 strain, 33 of the 36 isolates could be classified into 20 new allelic forms of the K1 type (Table 1). Among them, T1-T8 (Table 1) from 20 isolates with 21 repeat serine residues in the SR the same as those of the T9-102 and K1 strains were found to have only four non-synonymous substitutions separately locating in the OR (residue 40, Gly→Asp), the conserved region (residue 151, Glu→Lys) and the SR (a residue, Pro→Ser, inside the Insert E and residue 246, Arg→Lys), and two deletions of the octamer units (residues 42-49 and residues 58-65 in Table 1). However, the other 13 isolates with 15, 23, 25, 31 and 33 repeat serine residues in the serine stretch were found to have more variations in the OR, including 4 new insertions of octamer units (Insert A, B and D in Table 1), 10 novel non-synonymous and one deletion (Table 1). These results suggest a likely linkage of genetic variations between the OR and the SR. Moreover, frequencies of either insertion/deletion events (13/36) or non-synonymous mutations (29/36) occurred in the "repeat 1" domain were significantly higher than those of the "repeat 2" domain (4/36 or 3/36) (Table 1).

Nine new allelic forms of the Honduras-1 type were detected from 10 of the 69 isolates, but no wild isolates had identical sequence to those of the reported strains of this type (Table 1). It seems that this type was less common in the surveyed wild populations as compared to the other two types. The numbers of the repeat serine residues in the serine stretch varied from 15 to 39 among the allelic forms. Four novel non-synonymous substitutions at residues 48, 52, 57 and 74, and a new insertion (Insert A) and three deletions (residues 34-49, residues 42-49 and residues 58-65 in Table 1) of octamer units were detected in the OR. Since Honduras-1 type might be a hybrid between the K1 and the FCR3 types, the allelic forms of this type (Table 1) could be subdivided into H1-H4 group and H5-H9 group by their distinct allotype patterns HI-IV and HV-VII (Fig 2) related to those of the FCR3 type and K1 type, respectively.

Geographical distribution of the three types

Table 2 shows that the allelic forms of three types of *P. falciparum* SERA exon II have a biased

Table 2
Geographic distributions of allelic forms of *Plasmodium falciparum* SERA exon II.

Allotypes[a]	SEA[b]	Solomon Islands	African countries[c]	Brazil	Total
FCR3	6	0	0	0	6
FI-II	1(F2)[d]	5(F1-2)	0	0	6
FIII-IV	0	0	0	11(F3-5)	11
Honduras-1					
HI-IV	4(H1-3)	0	1(H4)	0	5
HV-VII	2(H6-7)	0	3(H5, H8-9)	0	5
K1	17(T9-102, T1-7)	1(T9-102)	5(T4, T7-8)	0	23
TI-VIII	1(T10)	0	12(T9-20)	0	13
Total	31	6	21	11	69

[a]Representative strains of each type are in boldface letters. Roman numerals marked with F, H and T are different allotypes in the serine repeat stretch of SR (Fig 2).
[b]Isolates collected from Southeast Asia endemic areas in Vietnam, Thailand, Indonesia and Myanmar.
[c]Isolates collected from endemic areas in Tanzania, Gabon, Guinea Bissau and Nigeria.
[d]In parentheses, numerals marked with F, H and T, and T9-102, are the allelic forms of SERA exon II (Table 1).

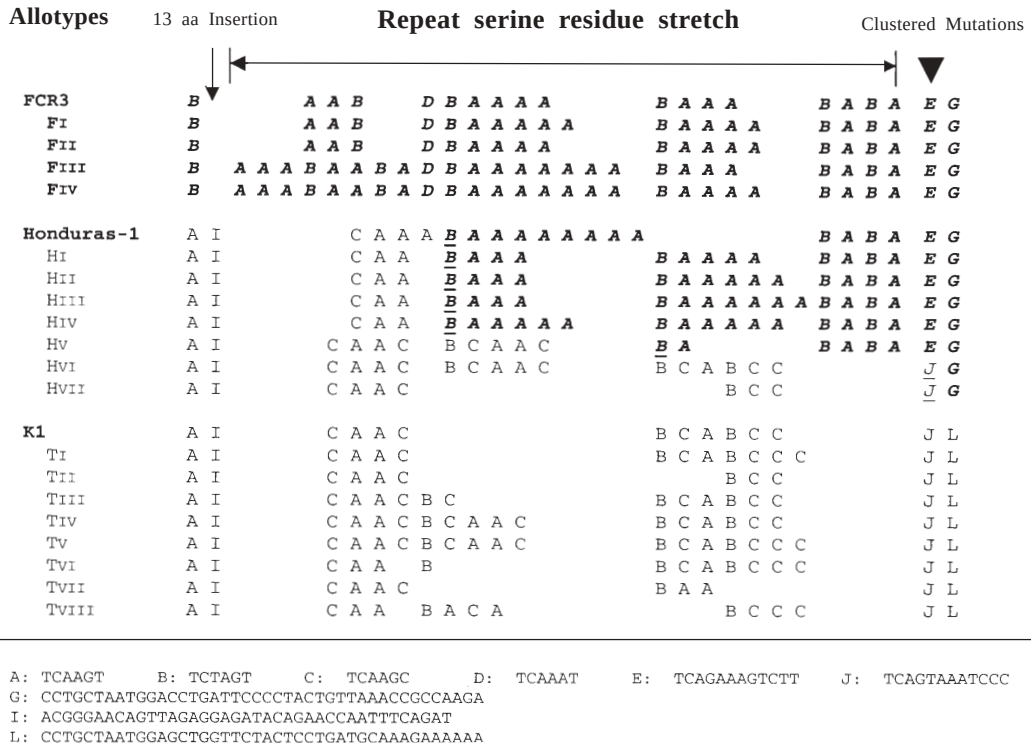

Fig 2–Allotype patterns in the serine repeat region of *Plasmodium falciparum* SERA gene. Under the allotype column, the representative strains of each type are in boldface letters. Roman numerals marked with F-, H- and T- are different allotypes in the SR. Capital letters A, B and C are abbreviations of the sequences coding three different serine-serine motifs, and D is for the serine-asparagine motif. Letters E, G, I, J, and L are abbreviations for different sequence fragments. Allotype patterns of FCR3 type and the posterior portion, homologous to those of FCR3 type, of some Honduras-1 allotypes are in boldface italic letters. Allotypes of K1 type and the homologous anterior portion of Honduras-1 allotypes are in normal letters. Underlined letters indicate the conjectured recombinant sites.

geographical distribution. A group of FCR3 allelic forms (F3-F5) with the longest serine stretch in the SR were allelic forms solely detected in Brazilian isolates, while two other allelic forms (F1-F2) were detected in the isolates from Solomon Islands and Indonesia. Six isolates from Vietnam were identical to the reported FCR3 strain. No allelic form of this type was detected in the African isolates.

The allelic forms of K1 type were detected in 17 of 21 African isolates and 18 of 31 isolates from Southeast Asia (Table 2). Most of the K1 allelic forms in the African isolates (T9-T20 in Table 1, TI-TVIII in Fig 2) were found to have most of the novel non-synonymous substitutions and new insertions in the OR and longer or shorter serine stretch than the reported. On the other hand, most of the K1 type isolates (17/18) from Southeast Asia (T1-T7, T9-102 in Table 1) had the same length of serine stretch and the allotype patterns of the SR (K1 in Fig 2) as the reported strains, and a few non-synonymous substitutions and insertion/deletions of octamer units in the OR (Table 1), suggesting a less diversity of the K1 type in Southeast Asia.

The allelic forms of the Honduras-1 type were detected in either 6 isolates from Southeast Asia or 4 isolates from African countries (Table 2). Thus, the allelic forms of the K1 and Honduras-1 type were widely distributed in the Old World and all of the three types were found in Southeast Asia.

DISCUSSION

Protective immunity against *P. falciparum* was demonstrated in the 1970s by immunization of human patients with irradiated sporozoites (Clyde

et al, 1973). Nowadays, many genes of malarial antigens, such as merozoite surface proteins (MSP-1, MSP-2), circumsporozoite protein (CSP), serine repeat antigen (SERA), apical membrane protein (AMA), ring-infected erythrocyte surface antigen (RESA) etc have been studied as candidates of malaria vaccine (Doolan and Hoffman, 1997; Engers *et al*, 1998). Generally, determining the extent of diversity of antigenic genes and identifying the mechanisms by which antigen variation is generated and persists in populations of *P. falciparum* is important for the successive efforts of developing an effective malaria vaccine (Rich *et al*, 1997). In this study, we focused on the genetic diversity of the SERA gene in wild *P. falciparum* isolates collected from global endemic areas, and found that this gene in natural populations has comparatively higher diversity than that in cultured strains. From sequence analysis of the SERA exon II, it is suggested that the recombination events and the length polymorphism of the serine stretch of the SR, and the deletion/insertion of the octamer units and non-synonymous substitutions in the OR are possible mechanisms of genetic variation of this gene in natural populations of this parasite. Furthermore, these mechanisms, together with the presence of more than one allelic forms in same host (multiple-clonal infection, data not shown), suggest a potential which may act as the genetic base of adapting the parasite to the host immune system and evading the immune defenses induced by the SERA based vaccine, like those found in the CSP and MSA-1 genes (Jongwutiwes *et al*, 1994; Kaneko *et al*, 1997; Ferreira *et al*, 1998).

There were a few studies on the immunological significance of the OR of the SERA but no data about that of the SR. In this study, frequent non-synonymous substitutions and the insertion/ deletion of the octamer units were found in the OR (Table 1), suggesting that the OR domain is subjected to immunological selective pressure, and presumably have immuno-dominant epitopes as the target of host immunoresponse in *P. falciparum* infection. This might have been supported by a recent report (Fox *et al*, 1997) that a fusion peptide encoded by the OR domain (amino acid residue 17-58 in Table 1) was recognized strongly by the pooled immune sera from Nigerian patients. Also, a 14-mer peptide (amino acid residue 59-72), flanking the most posterior octamer unit and its adjacent conserved segment, was identified as a target epitope of a parasite-inhibitory monoclonal antibody *in vitro* (Fox *et al*, 1997). It is important that the adjacent conserved segment was confirmed to be an epitope, because this segment has been found, in this study, to be the "repeat 3" of octamer units, which had never been reported before and thus no more conserved. Further studies are necessary to ascertain the presence of other epitope(s) encoded by the OR of the SERA.

It is yet unclear about the biological significance of the length polymorphism of the serine stretch, synonymous substitutions and recombination in the SR. However, if the 21 repeat serine residues in the serine stretch of the reported strains of K1 type can be considered as a stable structure due to the culture adaptation, it is worth to note that those K1 allelic forms with less or more repeat serine residues than the 21 have higher frequency of variations in the OR (Table 1). This may suggest a linkage of genetic variations between the SR and the OR in wild populations under the host immune pressure. Rich *et al* (1997) analyzed 25 DNA sequences of *P. falciparum* CSP gene and discussed the silent substitutions among 14 repeat allotypes encoding two amino acid motifs. They inferred that the variations of the repeat numbers and sequences within the central region of tandem repeats of CSP gene are the consequence of mitotic intragenic recombination between repeat allotypes (Rich *et al*, 1998). The different allotype patterns of the SERA types are likely to be generated similarly but the allelic forms of Honduras-1 type should be the consequence of meiotic (sexual) intragenic recombination between the other two types.

Although the SERA gene exon II in natural *P. falciparum* populations seems not as highly conservative as in culture strains, genetic variations possibly linking to immunogenicity are likely to be confined in the OR. It seems that the diversity of this gene tends to be restricted because 20 allelic forms with differences of only one to two amino acids in the OR constitute as many as 68.1% (47/69) of the isolates (FCR3, F1-F5, H3-H6, H8, T9-102, T4, T6-T10, T14, T17 in Table 1) in spite of their differences in the SR, and 18 allelic forms (H2-H4, H6-H7, H9, T1-T2, T11-T20) were singletons (those in which only a single sequence contains the variant) in the OR. This fact may imply a relatively slow and steady variation of this gene in nature and thus, developing the SERA as a portion of malaria vaccine cocktail can be reasonably confined to the major allelic forms with high prevalence in endemic region.

The geographical distribution of the allelic

forms of the three SERA types is appearently biased (Table 2). Unlike the quite homologous allelic forms of FCR3 type, allelic forms of the K1 type in African isolates have higher diversity in both of the OR and SR as compared with isolates from Southeast Asia (Table 2), suggesting different pressures of host immune response between endemic regions of these continents. It is worth to investigate whether or not the human hosts, vectors and even environmental factors between endemic regions of Africa and Southeast Asia influence the K1 type of *P. falciparum* SERA in different ways. The allelic forms of the Honduras-1 type scattered over the Old World with a low proportion in the tested isolates (10/69, Table 2). As allelic forms of the Honduras-1 type might be recombinants of the other two types, whether or not the rareness of the FCR3 type in Africa isolates (Table 2) is true and thus accounts for limited recombination events for the Honduras-1 allelic forms needs to be further studied. Although the analysis of many more isolates is necessary for elucidating details of the global distribution of the types of the SERA gene exon II, our results may serve as a baseline data for the vaccine development utilizing epitopes from the SERA gene of *P. falciparum*.

ACKNOWLEDGEMENTS

The authors thank Dr David Warhurst, London School of Hygiene and Tropical Medicine, London, UK, for providing the samples collected from west African countries. This study was supported by Grant-in-Aid for International Scientific Research: Field Survey (07041159 and 09041179) and Grant-in-Aid for Scientific Research on Priority Areas (08281207, 09270210 and 10166209) from the Ministry of Education, Science, Culture and Sports, Japan, and by a grant from the Toyota Foundation (96B3-011), Japan.

REFERENCES

Babiker HA, Walliker D. Current views on the population structure of *Plasmodium falciparum*: implications for control. *Parasitol Today* 1997; 13: 262-7.

Barr PJ, Inselburg J, Green JM, *et al*. Immunogenicity of recombinant *Plasmodium falciparum* SERA protein in rodents. *Mol Biochem Parasitol* 1991; 45: 159-79.

Bzik DJ, Li W-b, Horii T, Inselburg J. Amino acid sequence of the serine repeat antigen (SERA) of *Plasmodium falciparum* determined from cloned cDNA. *Mol Biochem Parasitol* 1988; 30: 279-88.

Clyde DF, McCarthy VC, Miler RM, Hornick RB. Specificity of protection of man immunized against sporozoite-induced falciparum malaria. *Am J Med Sci* 1973; 266: 398-403.

Debrabant JA, Delplace P. Leupeptin alters the proteolytic processing of p126, the major parasitophorous vacuole antigen of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1989; 33: 151-8.

Debrabant JA, Maes P, Delplace P, Dubremetz F, Tartar A, Camus D. Intramolecular mapping of *Plasmodium falciparum* p126 proteolytic fragments by N-terminal amino acid sequencing. *Mol Biochem Parasitol* 1992; 53: 89-95.

Delplace P, Fortier B, Tronchin G, Dubermetz JF, Vernes A. Localization, biosynthesis, processing and isolation of a major 126 kDa antigen of the parasitophorous vacuole of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1987; 33: 151-18.

Delplace P, Bhatai A, Cagnard M, *et al*. Protein p126: A parasitophorous vacuole antigen associated with the release of *Plasmodium falciparum* merozoites. *Biol Cell* 1988; 64: 215-21.

Doolan DL, Hoffman SL. Multi-gene vaccination against malaria: A multistage, multi-immune response approach. *Parasitol Today* 1997; 13: 171-8.

Engers HD, Godal T. Malaria vaccine development: current status. *Parasitol Today* 1998; 14: 56-64.

Ferreira MU, Liu Q, Zhou M, *et al*. Stable patterns of allelic diversity at the merozoite surface protein-1 locus of *Plasmodium falciparum* in clinical isolates from southern Vietnam. *J Euk Microbiol* 1998; 45: 131-6.

Fox BA, Bzik DJ. Analysis of stage-specific transcripts of the *Plasmodium falciparum* serine repeat antigen (SERA) gene and transcription from the SERA locus. *Mol Biochem Parasitol* 1994; 68: 133-44.

Fox BA, Pang XL, Suzue K, Horii T, Bzik D. *Plasmodium falciparum*: An epitope within a highly conserved region of the 47-kDa aminoterminal domain of the serine repeat antigen is a target of parasite-inhibitory antibodies. *Exp Parasitol* 1997; 85: 121-34.

Gor DO, Li AC, Rosenthal PJ. Protective immune responses against protease-like antigens of the murine malaria parasite *Plasmodium vinckei*. *Vaccine* 1998; 16: 1193-202.

Horii T, Bzik DJ, Inselburg J. Characterization of antigen-expressing *Plasmodium falciparum* cDNA clone that are reactive with parasite inhibitory antibodies. *Mol Biochem Parasitol* 1988; 30: 9-18.

Inselburg J, Bzik DJ, Li W-b, *et al*. Protective immunity induced in Aotus monkeys by recombinant SERA

proteins of *Plasmodium falcipaum*. *Infect Immun* 1991; 59: 1247-50.

Inselburg J, Bathurst IC, Kansopon J, Barchfeld GL, Barr PJ, Rossan RN. Protective immunity induced in Aotus monkeys by a recombinant SERA protein of *Plasmodium falciparum*: Adjuvant effects on induction of protective immunity. *Infect Immun* 1993a; 61: 2041-47.

Inselburg J, Bathurst IC, Kansopon J, Barr PJ, Rossan RN. Protective immunity induced in Aotus monkeys by a recombinant SERA protein of *Plasmodium falciparum* Further studies using SERA 1 and MF75.2 adjuvant. *Infect Immun* 1993b; 61: 2048-52.

Jongwutiwes S, Tanabe K, Hughes MA, Kanbara H, Hughes AL. Allelic variation in the circumsporozoite protein of *Plasmodium falciparum* from Thai field isolates. *Am J Trop Med Hyg* 1994; 51: 659-68.

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 Pasasitol* 1997; 86: 45-57.

Kawamoto F. Rapid diagnosis of malaria by fluorescence microscopy with light microscope and interference filter. *Lancet* 1991; 337: 200-2.

Kimura K, Kaneko O, Inoue A, Ishii A, Tanabe K. Amplification by polymerase chain reaction of *Plasmodium falciparum* DNA from Giemsa stained thin blood smears. *Mol Biochem Parasitol* 1995; 70: 193-7.

Knapp B, Hundt E, Nau U, Kupper HA. Molecular cloning, gene structure and localization in a blood stage antigen of *Plasmodium falciparum* characterized by a serine stretch. *Mol Biochem Parasitol* 1989; 32: 73-88.

Li W-b, Bizk DJ, Horii T, Iselburg J. Structure and expression of the *Plasmodium falciparum* SERA gene. *Mol Biochem Parasitol* 1989; 33: 13-26.

Morimatsu K, Morikawa T, Tanabe K, Bzik DJ, Horii T. Sequence diversity in the amino-terminal 47-kDa fragment of the *Plasmodium falciparum* serine repeat antigen. *Mol Biochem Parasitol* 1997; 86: 249-52.

Pang XL, Horii T. Complement-mediated killing of *Plasmodium falciparum* erythrocytic schizont with antibodies to the recombinant serine repeat antigen (SERA). *Vaccine* 1998; 16: 1299-305.

Perkins ME, Ziefer A. Preferential binding of *Plasmodium falciparum* SERA and rhoptry proteins to erythrocyte membrane inner leaflet phospholipids. *Infect Immun* 1994; 62: 1207-12.

Ragge K, Aronold HH, Tummler M, Knapp B, Hundt E, Lingelbach K. *In vitro* biosynthesis and membrane translocation of the serine rich protein of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1990; 42: 93-100.

Rich SM, Hudson RR, Ayala FJ. *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. *Proc Natl Acad Sci USA* 1997; 94: 13040-5.

Rich SM, Light MC, Hudson RR, Ayala FJ. Malaria's eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1998; 95: 4425-30.

Sugiyama T, Suzue K, Okamoto M, Inselburg J, Tai K, Horii T. Production of recombinant SERA proteins of *Plasmodium falciparum* in *E. coli* by using synthetic genes. *Vaccine* 1996; 14: 1068-76.

Suzue K, Ito M, Matsumoto Y, Tanioka Y, Horii T. Protective immunity induced in squirrel monkeys with recombinant serine repeat antigen (SERA) of *Plasmodium falciparum*. *Parasitol Int* 1997; 46: 1-25.

Tine JA, Lanar DE, Smith DM, *et al*. NYVAC-Pf7: a poxvirus-vectored, multiantigen multistage vaccine candidate for *Plasmodium falciparum* malaria. *Infect Immun* 1996; 64: 3833-44.

Udhayakumar V, Shi Y-P, Kumar S, Jue DL, Wohlhueter RM, Lal AA. Antigenic diversity in the circumsporozoite protein of *Plasmodium falciparum* abrogates cytotoxic-T-cell recognition. *Infect Immun* 1994; 62: 1410-3.

Weber JL, Lyon JA, Camus D. Blood stage antigen genes of *Plasmodium falciparum.* In: Agabian N, Goodman H, Noguiera N, eds. Molecular Strategies of Parasitic Invasion. New York: Liss, 1987: 379-88.

World Health Organization. Tropic disease report, Twelfth Program Report. Geneva: World Health Organization, 1995.

Zevering Y, Khamboonruang C, Good MF. Natural amino acid polymorphisms of the circumsporozoite protein of *Plasmodium falciparum* abrogate specific human CD4+ T cell responsiveness. *Eur J Immunol* 1994; 24: 1418-25.