Genetic Polymorphism of *Plasmodium vivax* *msp1p*, a Paralog of Merozoite Surface Protein 1, from Worldwide Isolates

Yue Wang,[†] Osamu Kaneko,[†] Jetsumon Sattabongkot, Jun-Hu Chen, Feng Lu, Jong-Yil Chai, Satoru Takeo, Takafumi Tsuboi, Francisco J. Ayala, Yong Chen, Chae Seung Lim, and Eun-Taek Han*

Department of Parasitology, Kangwon National University College of Medicine, Chuncheon, Gangwon-do, Republic of Korea; Institute of Parasitic Diseases, Zhejiang Academy of Medical Sciences, Hangzhou, People's Republic of China; Department of Protozoology, Institute of Tropical Medicine and the Global Center of Excellence Program, Nagasaki University, Nagasaki, Japan; Department of Entomology, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand; Jiangsu Institute of Parasitic Diseases, Wuxi, People's Republic of China; Department of Parasitology and Tropical Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea; Cell-Free Science and Technology Research Center, Ehime University, Ehime, Japan; Department of Ecology and Evolutionary Biology, University of California, Irvine, California; Zhejiang Medical College and Zhejiang Academy of Medical Sciences, Hangzhou, People's Republic of China; Department of Laboratory Medicine, College of Medicine, Korea University, Seoul, Republic of Korea

Abstract. *Plasmodium vivax msp1p*, a paralog of the candidate vaccine antigen *P. vivax* merozoite surface protein 1, possesses a signal peptide at its N-terminus and two epidermal growth factor-like domains at its C-terminus with a glycosylphosphatidylinositol attachment site. The *msp1p* gene locus may have originated by a duplication of the *msp1* gene locus in a common ancestor of the analyzed *Plasmodium* species and lost from *P. yoelii*, *P. berghei*, and *P. falciparum* during their evolutionary history. Full-length sequences of the *msp1p* gene were generally highly conserved; they had a few amino acid substitutions, one highly polymorphic E/Q-rich region, and a single-to-triple hepta-peptide repeat motif. Twenty-one distinguishable allelic types (A1–A21) of the E/Q-rich region were identified from worldwide isolates. Among them, four types were detected in isolates from South Korea. The length polymorphism of the E/Q-rich region might be useful as a genetic marker for population structure studies in malaria-endemic areas.

INTRODUCTION

Among the species of malarial parasite that infect humans, *Plasmodium vivax* is the most globally prevalent and threatens almost 40% of the world's population, resulting in approximately 250 million clinical infections each year.¹ Although *P. vivax* malaria had been considered relatively benign, compared with that of *P. falciparum*, this view is now being challenged. Additionally, resistance to chloroquine is appearing in countries where malaria is endemic.² Thus, there are good reasons to pursue an effective *P. vivax* vaccine.³ In this regard and in contrast to *P. falciparum*, research into *P. vivax* is limited, due in part to difficulties in culturing blood-stage parasites *in vitro*.⁴ Nevertheless, several *P. vivax* vaccine candidates from different parasitic stages have been characterized.⁵ Among them, various merozoite surface proteins (MSPs), apical membrane protein-1, duffy binding protein, Pvs25, Pvs28, circumsporozoite protein, and thrombospondin-related anonymous protein have been studied.⁵

Merozoite surface proteins have been characterized and are highly immunogenic in natural infection.^{6–9} Among them, several major vaccine candidate antigens (including MSP1, MSP4, MSP5, MSP8, and MSP10) are either known or presumed glycosylphosphatidylinositol (GPI)-anchored membrane proteins. *Plasmodium vivax* MSP1 is the largest and most abundant protein on the *P. vivax* merozoite surface.^{5,10} The gene that encodes this protein (*Pvmsp1*) is highly polymorphic and consists of a mosaic of conserved and variable blocks with numerous recombination sites distributed throughout the gene. However, the fragment that encodes the 19-kDa C-terminal epidermal growth factor (EGF)-like

domain is relatively conserved.¹¹ This gene has been used as a polymorphic marker for investigations of the genetic structure of *P. vivax* populations and in molecular epidemiology.¹²

With the completion of *P. vivax* genome sequencing, GPI-anchored proteins of *P. vivax* have been predicted by comparison with validated *P. falciparum* GPI-anchored proteins.^{13,14} *Plasmodium vivax msp1p* (*Pvmsp1p*), a novel paralog of the *Pvmsp1* gene, was found immediately upstream of *Pvmsp1*.¹⁴ This gene is predicted to encode a 1,854-amino-acid protein (predicted molecular mass of 215 kDa) with an N-terminal signal sequence, C-terminal EGF-like domains, and a GPI-attachment motif (Figure 1).¹⁴ The functions of this molecule remain unknown. Thus, we have analyzed available genomic data from PlasmoDB (<http://www.plasmodb.org/>) to search for distinctive pattern of diversity in *msp1p* and *msp1* genes among *Plasmodium* species. We have also assessed the nature and extent of polymorphisms in PvMSP1P from worldwide isolates and laboratory lines of *P. vivax*.

MATERIALS AND METHODS

Gene sequences. The following sequences of malarial parasites were used for the analyses: human *P. falciparum* PfMSP1 (CAA27070), PfMSP8 (PFE0120c), and PfMSP10 (PFF0995c), and *P. vivax* PvMSP1 (PVX_099980), PvMSP8 (PVX_097625), PvMSP10 (PVX_114145), and PvMSP1P (PVX_099975); rodent malaria *P. berghei* PbMSP1 (AAC28871), PbMSP8 (PBANKA_110220), and PbMSP10 (PBANKA_111960), *P. yoelii* PyMSP1 (PY05748), and *P. chabaudi* PchMSP1 (PCAS_083080); primate malaria *P. knowlesi* PkMSP1 (PKH_072850) and PkMSP1P (PKH_072840), *P. reichenowi* PrMSP1 (CAH10285), and *P. cynomolgi* PcyMSP1 (BAI82251); and avian malaria *P. gallinaceum* PgMSP1 (CAH10838). The *P. gallinaceum* sequence database (<http://www.sanger.ac.uk/>) was used to search for homologs of PvMSP1P and PvMSP10.

Blood samples and DNA preparation. Blood samples were collected, after informed consent had been obtained, from 81

*Address correspondence to Eun-Taek Han, Department of Parasitology, Kangwon National University College of Medicine, Hyoja2-dong, Chuncheon, Gangwon-do 200-701, Republic of Korea. E-mail: ethan@kangwon.ac.kr

[†]These authors contributed equally to this article.

FIGURE 1. Schematic diagram of the *Plasmodium vivax* merozoite surface protein 1 paralog. A putative signal peptide (SP), the epidermal growth factor (EGF)-like domains, and the glycosylphosphatidylinositol (GPI)-anchor attachment site are indicated. TR = tandem repeat region of hepta-peptide; PR = polymorphic E/Q rich region. White color indicates region conserved among strains.

symptomatic patients diagnosed by microscopic examination with *P. vivax* infection at Korea University Ansan Hospital, local health centers, and clinics in Gyeonggi and Gangwon provinces, Republic of Korea. Genomic DNA was purified from 200 μ L of whole blood by using a QIAamp DNA Blood Mini Kit (QIAGEN, Valencia, CA), according to the manufacturer's protocol. Genomic DNA of *P. vivax* isolates ($n = 33$) obtained from Thailand ($n = 26$), Indonesia ($n = 3$), India ($n = 1$), Papua New Guinea ($n = 1$), Western Samoa ($n = 1$), and Pakistan ($n = 1$), and nine *P. vivax* laboratory lines (Africa Mauritania, New Guinea, Honduras III, Brazil I, Salvador I, Vietnam IV, Indonesia I, India VII, and Columbia Rio Meta) were used for polymerase chain reaction (PCR).

Amplification and sequencing of target genes. Genomic DNA from 20 isolates from the Republic of Korea and 9 isolates from other locations was used for amplification of the *Pvmsp1p* full-length gene. Primers SeqF1 (5'-TGC ATA TTC ATA CGC GTG TGT-3') and SeqR8 (5'-GGC TGC CCC TAA CTT AGC A-3') were designed based on the *Pvmsp1p* sequence of the *P. vivax* Sal I strain. These were used to amplify DNA fragments from 90 basepairs upstream to 80 basepairs downstream of the *Pvmsp1p* coding sequence by using LA Taq DNA polymerase (TaKaRa, Tokyo, Japan). The PCR amplification was performed on a MyCycler Thermal Cycler (Bio-Rad, Hercules, CA) by using the following temperature profile: 94°C for 2 minutes; 35 cycles at 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 6 minutes; and a final extension at 72°C for 10 minutes.

Both strands of the PCR products were directly sequenced by using a series of sequencing forward primers (SeqF1; SeqF2: 5'-ATC AAC CGG AAG AAC TCC CT-3'; SeqF3: 5'-CAA AGG GAG AAG AAA AAA ATG TAC C-3'; SeqF4: 5'-GGT GAG CTA ATC GAA CGG G-3'; SeqF5: 5'-TGG GGC GCA CAT AAC CT-3'; SeqF6: 5'-CCC GTC TAC TCC AAG GAT GTG ATA AG-3'; SeqF7: 5'-TGA AGT GCA ACA CGT GGA AT-3'; SeqF8: 5'-GTG GACTACTAC GGG CTA AGG A-3' and SeqF9: 5'-ATT CTC TAT GCA GAC AAG GAG GTG-3') and reverse primers (SeqR1: 5'-AAG GCA GGA TTA GAG AGG ACG; SeqR2: 5'-CGC ACG TTT AGG TGG TAG TC-3'; SeqR3: 5'-AGC GTC AAA TCG TGG CAG-3'; SeqR4: 5'-TTC GTG ATG ATC GCG TTG GTT AGC AG-3'; SeqR5: 5'-TCC CCG ATG AAG AAA TAT GC-3'; SeqR6: 5'-ACT GCA GAT GGA TGG TCA TCT-3'; SeqR7: 5'-AAC TGC ATC GCG TCC GTA T-3'; and SeqR8) using an ABI Prism 377 DNA sequencer (Genotech, Seoul, South Korea).

Analysis of the full-length gene sequences showed one highly polymorphic region. This region was amplified by PCR from genomic DNA of 61 samples from the Republic of Korea and 24 samples from other locations, as well as nine laboratory strains, and sequenced. To examine variable tandem repeat regions, as found in the *Pvmsp1p* gene sequence, primers (TR-F: 5'-CCT ACA CGG GAT GGG AGA T-3' and

TR-R: 5'-CGG AGA GCG AGT TCG TGA T-3') were used to amplify a 200-basepair fragment encompassing this region from 33 worldwide isolates and 9 laboratory lines.

Data analysis. Sequence data were submitted to GenBank under accession numbers GU556592–GU556620. Amino acid sequence alignments were constructed using the MUSCLE program, with manual corrections.¹⁵ The number of nonsynonymous substitutions per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) were computed by using the Nei-Gajbordi method¹⁶ with the Jukes-Cantor correction, as implemented in the MEGA 4 program.¹⁷ An unrooted tree was constructed by the neighbor-joining method with the Jones-Taylor-Thornton amino acid substitution model,¹⁸ accompanied by bootstrap analysis with 1,000 replicates for the neighbor-joining method and 100 for the maximum parsimony method implemented in PHYLIP version 3.68 after excluding insertions/deletions (indels) and unreliable amino acid sites.¹⁹

RESULTS

An MSP1P homolog can be found in the *P. vivax* and *P. knowlesi* genome databases, but not in the *P. falciparum*, *P. yoelii*, or *P. berghei* genome databases. To investigate the evolutionary relationship of MSP1P with other MSPs, we searched for *Pvmsp1p* homologs in the available *Plasmodium* genome database and analyzed their relationship with MSP1 by using the distantly related MSP8 and MSP10 sequences as outgroups. A TBLASTN search of the *P. gallinaceum* sequence database (<http://www.sanger.ac.uk>) was conducted by using the PvMSP1P amino acid sequence as a query. This search identified a contig (28a.d000006175.Contig1) that contained a putative partial sequence of the *Pgmsp1p* gene (encoding the C-terminal end included EGF-like domains) (Figure 2). Based on the BLOSUM matrix, the amino acid sequence identity/similarity of the EGF-like domains to those of PvMSP1P and PkMSP1P were 56/71% and 58/75%, respectively (Figure 3A). The identity/similarity of the N-terminal region of this gene

FIGURE 2. Alignment of the partial amino acid sequences of *Plasmodium vivax* merozoite surface protein 1 paralog (PvMSP1P), *P. knowlesi* merozoite surface protein 1 paralog (PkMSP1P), and *P. gallinaceum* merozoite surface protein 1 paralog (PgMSP1P). Dashes indicate deletions. Cys residues with light areas indicate Cys residues conserved among all sequences and those with dark areas and the arrowhead indicate the additional two Cys residues conserved among MSP8, MSP10, MSP1P, and Pf/Pr/PgMSP1 (Figure 3). Asterisks, dots, and colons under the alignment indicate identical, conserved, and semi-conserved substitutions, respectively, based on BLOSUM. The glycosylphosphatidylinositol (GPI) attachment site is indicated with an arrowhead.

FIGURE 3. Relationship between *Plasmodium vivax* merozoite surface protein 1 paralog (MSP1P) and other *Plasmodium* merozoite surface proteins possessing epidermal growth factor (EGF)-like domains. **A**, Amino acid sequence alignment of the EGF-like domains of *Plasmodium* MSP1, MSP8, MSP10, and MSP1P. Dashes indicate a deletion. Cys residues with light areas indicate Cys residues conserved among all sequences and those with dark areas masks and the arrowhead indicate the additional two Cys residues conserved among MSP8, MSP10, MSP1P, and Pf/Pr/PgMSP1. Asterisks, dots, and colons under the alignment indicate identical, conserved, and semi-conserved substitutions, respectively, based on BLOSUM. **B**, Unrooted dendrograms of the EGF-like region of MSP1, MSP8, MSP10, and MSP1P amino acid sequences. Trees were constructed by the neighbor-joining and maximum parsimony methods using amino acid positions 1759, 1782, 1783, 1784, 1785, 1786, 1787, 1812, 1813, 1814, 1815, 1822, 1823, and 1824 (after *P. vivax* MSP1P amino acid sequence) after excluding indel and unreliable sites. Numbers on branches indicate bootstrap values. **C**, Schematic diagram of the proposed evolutionary history of the *msp1p* gene locus in *Plasmodium* spp. The *msp1p* gene locus was generated by duplication of the *msp1* gene locus in the common ancestor of known *Plasmodium* species. This locus was then deleted in *P. yoelii*, *P. berghei*, and *P. falciparum*. Sequences of *P. falciparum* PfMSP1 (CAA27070), PfMSP8 (PFE0120c) and PfMSP10 (PFF0995c); *P. vivax* PvMSP1 (PVX_099980), PvMSP1P (PVX_099975), PvMSP8 (PVX_097625) and PvMSP10 (PVX_114145); *P. knowlesi* PkMSP1 (PKH_072850) and PkMSP1P (PKH_072840); *P. berghei* PbMSP1 (AAC28871), PbMSP8 (PBANKA_110220) and PbMSP10 (PBANKA_111960); *P. reichenowi* PrMSP1 (CAH10285); *P. gallinaceum* PgMSP1 (CAH10838), PgMSP1P (encoded in 28a.d000006175.Contig1), PgMSP10 (encoded in 28a.d000005716.Contig1); *P. cynomolgi* PcyMSP1 (BAI82251); *P. yoelii* PyMSP1 (PY05748); and *P. chabaudi* PchMSP1 (PCAS_083080) were used.

product to the corresponding region of PvMSP1P (amino acid positions 1,402–1,675) and PkMSP1P (1,425–1,699) were 37/60% and 39/61%, respectively (Figure 2), where the similarity with PgMSP1 was less than 30%. This gene product formed a single clade with Pv/PkMSP1P with high bootstrap values (99–100%; Figure 3B), confirming that this gene was *Pgmsp1p*. Thus, *P. gallinaceum*, *P. vivax*, and *P. knowlesi* have both *msp1* and *msp1p* in their genome and PvMSP1P has weak homology with PvMSP1. This finding, in turn, suggests that the *msp1* and *msp1p* gene loci were generated by a gene duplication event prior to diversification of these parasite species. Because rodent malaria parasite species form a single clade with *P. vivax* and *P. knowlesi*, the lack of a *msp1p* homolog in the *P. yoelii* and *P. berghei* genomes is likely caused by deletion of the *msp1p* gene locus during their evolution. Deletion of this gene locus may also have occurred in *P. falciparum* (Figure 2C).

In a previous study, Carlton and others reported that the PvMSP1P EGF-like domains contained extra two Cys residues absent in PvMSP1 and rodent malaria parasite MSP1, but present in PfMSP1 (Figure 3A).¹⁴ This finding suggested that PvMSP1P might be evolutionarily closer to PfMSP1 than to PvMSP1. However, this appears not to be so for three reasons. First, the dendrogram using EGF-like domains indicated that PvMSP1P formed one clade with PkMSP1P and PgMSP1P and was separated from the MSP1 clade (Figure 3B). Second, beside the EGF-like domains, N-terminal side of the

PgMSP1P showed greater similarity to Pv/PkMSP1P (> 60%) than PgMSP1 (< 30%). Third, EGF-like domains of the distantly related MSP8 and MSP10 proteins contain two extra Cys residues, similar to MSP1P and Pf/Pr/PgMSP1. This finding indicates that the common ancestral protein of these MSPs possessed 12 Cys residues (Figure 3A). Collectively, these data suggest that the two Cys sites of the first MSP1 EGF-like domain in *P. vivax*, *P. knowlesi*, *P. yoelii*, and *P. berghei* were substituted with other amino acids during their evolution.

To assess the global genetic diversity of the *Pvmsp1p* gene, we determined the *Pvmsp1p* full-length sequence of 20 isolates of *P. vivax* from the Republic of Korea and 9 strains of *P. vivax* from other locations worldwide. One highly polymorphic glutamate (Glu, E)/glutamine (Gln, Q)-rich region and a polymorphic hepta-peptide motif (SAYSYSVS) with number variation (single to triple) were detected. There was no amino acid polymorphism in the EGF-like domains (Figure 1). No diversifying selection was detected, and there was no significant excess of *dN* over *dS*.

To determine the repeat variation of the hepta-peptide motif, additional genomic DNA from *P. vivax* worldwide, isolates and laboratory strains was amplified by PCR. All field isolates and laboratory strains had double repeats of this hepta-peptide motif, except for the isolates from Western Samoa (triple repeats) and Pakistan (single repeat). Outside of one E/Q-rich region and the hepta-peptide motif, the sequences were highly conserved with relatively few amino acid substitutions (Table 1).

TABLE 1

Amino acid polymorphisms in the full-length *Plasmodium* merozoite surface protein 1 paralog sequence of *P. vivax* isolates from various locations*

Positions of amino acids																	
2	3	5	7	//	1	1	1	1	1	1	1	1	1	1	1		
8	9	0	1	5	//	2	3	3	4	5	5	7	7				
9	9	8	7	5	//	3	2	8	0	4	5	0	6				
					//	2	0	6	4	6	3	3	7				
																No. (%)	Isolate
																	Sources
A R E R S // S R E Q Q E P D																	Salvador I (PVX_099975)
. //																9 (31.0)	South Korea, Thailand, Pakistan, India
. // T .																8 (27.6)	South Korea, Papua New Guinea
. I //																3 (10.3)	South Korea
. I // T .																4 (13.8)	South Korea
. I // . . K . . K . .																1 (3.4)	Western Samoa
. // . . K E																1 (3.4)	Indonesia
V . . K . //																1 (3.4)	Thailand
. . K . //																1 (3.4)	Thailand
. Q . . I // N L . K H . . .																1 (3.4)	Thailand
Total																29 (100)	

*// = region included hepta-repeat and E/Q-rich sequences.

Of these substitutions, only the S755I or P1686T (or both) mutations were found in 15 of the isolates from the Republic of Korea and Papua New Guinea, whereas more mutants were found in the worldwide isolates (Table 1).

A short and highly diverse region, composed of Glu and Gln as 3–5 Glu residues, followed by one or several basic E/Qn (n = 1–6) units, was found in the *Pvmsp1p* gene. Twenty-one distinguishable allelic types (A1–A21) were identified in 127 isolates (clones), based on a comparison with corresponding regions in the *P. vivax* Sal I strain (Table 2). Type A1 had an identical sequence to that of the Sal I strain, which was found in only two laboratory strains, from Central and South America. Type A2 predominated (33.1%, 42 of 127) in all *P. vivax* samples, and in the Korean (35.8%, 29 of 81), Thai (26.7%, 8 of 30), Pacific (67%, 2 of 3), and African isolates (100%, 1 of 1), which share 96.7% amino acid identity with type A1. The 81 Korean isolates appeared to have limited diversity because only four genotypes (allelic types A2, A7, A20, A21) were found, whereas isolates from other locations worldwide, and laboratory strains, showed 20 allelic types (the exception being A21). Interestingly, type A21 was detected only in Korean isolates (18.5%, 15 of 81).

Polymerase chain reaction amplification resulted in two or three target bands in each of three Thai isolates, which suggested multi-clone infection. To confirm this finding, PCR products of the polymorphic region amplified from these samples were cloned and sequenced. Three types (A2, A16, A20) were detected from the Thai T21 isolate, two (A7, A15) from isolate T25, and two (A19, A20) from isolate T29.

DISCUSSION

We have assessed the evolutionary relationship of the *msh1p* gene with other *msh* genes and propose that a duplication event (*msh1* and *msh1p*) occurred before the diversification of the clades in *P. vivax* and *P. gallinaceum*. This account requires two independent deletions of *msh1p*, one in the rodent lineage (after its divergence from the primate lineage to *P. knowlesi* and *P. vivax*) and another deletion in the lineage to *P. falciparum* (after its divergence from the primate lineage to *P. knowlesi* and *P. vivax*). We also propose that the common ancestor of *P. vivax*, *P. knowlesi*, *P. yoelii*, and *P. berghei* possessed MSP1 that had 12 Cys residues in the first EGF-like domain, and that two Cys sites were substituted to other amino

acids during their evolution. We further investigated the genetic diversity of the *PvmSP1p* gene in isolates from locations worldwide, including the Republic of Korea. We found an E/Q-rich polymorphic region, a hepta repeat region, and several polymorphic sites. However, no diversifying selection was apparent by comparing *dN* and *dS*. Although the molecular data (e.g., size, molecular mass, number, location of Cys residue) were similar to those of PvMSP1, PvMSP1P is not polymorphic and appears to not be under noticeable host immune pressure. However, the repeat-length polymorphism of the E/G-rich region may prove useful as a genetic marker for epidemiologic studies.

High conservation of the double EGF-like domains was also detected in other merozoite surface proteins, such as MSP1 and MSP4. These are involved in putative ligand-receptor interactions during erythrocyte invasion by merozoites.^{20,21} Thus, the lack of variation in the C-terminus sequence of PvMSP1P, especially the high conservation of the double EGF-like domains, suggests that these regions play an important role in this process.

The overall nucleotide diversity of *Pvmsp1p* is much lower than that of other *P. vivax* antigens, such as MSP1, MSP3 β , and apical membrane antigen 1.^{20,22,23} In the PvMSP1P sequences, the E/Q-rich region was shown to be highly polymorphic (21 allelic types in 127 clones/isolates). In the cases of PvMSP1 and Pfs230 (AF269242), the E/Q-rich region was also highly polymorphic and represented the principal source of genetic diversity.^{24,25} In a low-complexity region analysis of *Plasmodium*,²⁶ Gln appeared with a somewhat higher frequency in the repetitive than in the non-repetitive motifs. The E/Q-rich regions and repeat motif of PvMSP1P and Pfs230 were located in a low-complexity region.²⁷ These low-complexity regions harbor tandem repeats identified in *Plasmodium* and correspond to species-specific and rapidly diverging regions.²⁶

This variation in E/Q-rich regions and the number of repeats could be generated by slipped-strand mispairing mechanisms. These result in duplication, deletion, or mutation of certain repeat units.^{28,29} The tandem repeat regions of PvMSP1P may result from rapid diversification, which enables the parasite to evade the immune response of the host by antigenic polymorphism.²⁶

Finally, the highly polymorphic E/Q-rich region sequence of PvMSP1P might be useful as a genetic marker for studies on the population structure and dynamics of *P. vivax* in malaria-endemic areas.

TABLE 2
Allelic types of the E/Q-rich region of *Plasmodium* merozoite surface protein 1 paralog and geographic prevalence of each *P. vivax* isolate from various locations*

Allelic types	Amino acid sequence variation of E/Q rich region sequence	No. (%) following sources						GenBank accession no.
		South Korea	Thailand	Asia†	Pacific region	Central and South America	Africa	
Sal I	1157 EE—EQQQ—EQEQEQEQEQKK 1174							PVX_099975
A1	..E—..EQ—..EQEQEQEQEQKK	29 (35.8)	8 (26.6)	2 (25.0)	2 (66.6)	2 (50.0)	1 (100)	2 (1.6)
A2	..E—..EQ—..EQEQEQEQEQKK		2 (6.7)					42 (33.1)
A3	..EE—..EQ—..EQEQEQEQEQKK		1 (3.3)					2 (1.6)
A4	..E—..EQ—..EQEQEQEQEQKK		1 (3.3)					1 (0.8)
A5	..E—..EQEQ—..EQEQEQEQEQKK		1 (3.3)					1 (0.8)
A6	..E—..EQEQ—..EQEQEQEQEQKK		2 (6.7)					2 (1.6)
A7	..E—..EQEQ—..EQEQEQEQEQKK	18 (22.2)	1 (3.3)					19 (15.0)
A8	..E—..EQ—..EQEQ—..EQEQEQEQKK		5 (16.7)	1 (12.5)	1 (33.3)	1 (25.0)		6 (4.7)
A9	..E—..EQ—..EQEQ—..EQEQEQEQKK			1 (12.5)				2 (1.6)
A10	..E—..EQ—..EQEQ—..EQEQEQEQKK			1 (12.5)		1 (25.0)		1 (0.8)
A11	..E—..EQ—..EQEQ—..EQEQEQEQKK			1 (12.5)				1 (0.8)
A12	..E—..EQEQ—..EQEQ—..EQEQEQEQKK			1 (12.5)				1 (0.8)
A13	..E—..EQ—..EQ—..EQEQEQEQKK			1 (12.5)				1 (0.8)
A14	..E—..EQEQ—..EQEQ—..EQEQEQEQKK			1 (12.5)				1 (0.8)
A15	..E—..EQ—..EQEQEQEQEQ—..EQEQEQEQKK		1 (3.3)					1 (0.8)
A16	..E—..EQ—..EQEQEQEQEQ—..EQEQEQEQKK		1 (3.3)					1 (0.8)
A17	..E—..EQ—..EQEQEQEQEQ—..EQEQEQEQKK		2 (6.7)					2 (1.6)
A18	..E—..EQEQEQEQEQEQ—..EQEQEQEQKK		2 (6.7)	1 (12.5)				3 (2.4)
A19	..E—..EQ—..EQEQEQEQEQ—..EQEQEQEQKK		2 (6.7)					2 (1.6)
A20	..E—..EQ—..EQ—..EQEQEQEQKK	19 (23.5)	2 (6.7)					21 (16.5)
A21	..E—..EQ—..EQ—..EQEQEQEQKK	15 (18.5)	2 (6.7)					15 (11.8)
Total		81 (100)	30‡ (100)	8 (100)	3 (100)	4 (100)	1 (100)	127 (100)

* Dots and dashes represent identical residues and deletions, respectively.

† Asia except South Korea and Thailand.

‡ Including four clones found in mixed infection from Thailand isolates.

Received July 30, 2010. Accepted for publication October 7, 2010.

Acknowledgment: We thank A. Escalante for providing the *P. vivax* DNA samples of laboratory strain used in this study.

Financial support: This study was supported by National Research Foundation of Korea grant (2009-075103) from the Korean government.

Disclaimer: The views of the authors do not purport to reflect the position of the U.S. Department of the Army or Department of Defense.

Authors' addresses: Yue Wang and Jun-Hu Chen, Department of Parasitology, Kangwon National University College of Medicine, Chuncheon, Gangwon-do 200-701, Republic of Korea and Institute of Parasitic Diseases, Zhejiang Academy of Medical Sciences, Hangzhou 310013, People's Republic of China, E-mails: wangyuerr@yahoo.com.cn and hzjunhuchen@yahoo.com.cn. Osamu Kaneko, Department of Protozoology, Institute of Tropical Medicine, Nagasaki University, 1-12-4 Sakamoto, Nagasaki, Japan, E-mail: okaneko@nagasaki-u.ac.jp. Jetsumon Sattabongkot, Department of Entomology, Armed Forces Research Institute of Medical Science, Bangkok 10400, Thailand, E-mail: JetsumonP@afirms.org. Feng Lu, Department of Parasitology, Kangwon National University College of Medicine, Chuncheon, Gangwon-do 200-701, Republic of Korea and Jiangsu Institute of Parasitic Diseases, Wuxi 214064, People's Republic of China, E-mail: lufeng981@hotmail.com. Jong-Yil Chai, Department of Parasitology and Tropical Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea, E-mail: cji@snu.ac.kr. Satoru Takeo and Takafumi Tsuboi, Cell-Free Science and Technology Research Center, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan, E-mails: tsuboi@ccr.ehime-u.ac.jp and takeo@ccr.ehime-u.ac.jp. Francisco J. Ayala, Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, E-mail: fjayala@uci.edu. Yong Chen, Zhejiang Medical College, Hangzhou 310053, and Zhejiang Academy of Medical Sciences, Hangzhou 310013, People's Republic of China, E-mail: cyong93@yahoo.com.cn. Chae Seung Lim, Department of Laboratory Medicine, College of Medicine, Korea University, Seoul 152-703, Republic of Korea, E-mail: malarim@korea.ac.kr. Eun-Taek Han, Department of Parasitology, Kangwon National University College of Medicine, Chuncheon, Gangwon-do 200-701, Republic of Korea, E-mail: ethan@kangwon.ac.kr.

REFERENCES

- Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM, 2007. Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 77: 79–87.
- Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, del Portillo HA, 2009. Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infect Dis* 9: 555–566.
- Herrera S, Corradin G, Arevalo-Herrera M, 2007. An update on the search for a *Plasmodium vivax* vaccine. *Trends Parasitol* 23: 122–128.
- Udomsangpet R, Kaneko O, Chotivanich K, Sattabongkot J, 2008. Cultivation of *Plasmodium vivax*. *Trends Parasitol* 24: 85–88.
- Galinski MR, Barnwell JW, 2008. *Plasmodium vivax*: who cares? *Malar J* 7 (Suppl 1): S9.
- Black CG, Barnwell JW, Huber CS, Galinski MR, Coppel RL, 2002. The *Plasmodium vivax* homologues of merozoite surface proteins 4 and 5 from *Plasmodium falciparum* are expressed at different locations in the merozoite. *Mol Biochem Parasitol* 120: 215–224.
- Perez-Leal O, Sierra AY, Barrero CA, Moncada C, Martinez P, Cortes J, Lopez Y, Torres E, Salazar LM, Patarroyo MA, 2004. *Plasmodium vivax* merozoite surface protein 8 cloning, expression, and characterisation. *Biochem Biophys Res Commun* 324: 1393–1399.
- Perez-Leal O, Sierra AY, Barrero CA, Moncada C, Martinez P, Cortes J, Lopez Y, Salazar LM, Hoebeke J, Patarroyo MA, 2005. Identifying and characterising the *Plasmodium falciparum* merozoite surface protein 10 *Plasmodium vivax* homologue. *Biochem Biophys Res Commun* 331: 1178–1184.
- de Oliveira CI, Wunderlich G, Levitus G, Soares IS, Rodrigues MM, Tsuji M, del Portillo HA, 1999. Antigenic properties of the merozoite surface protein 1 gene of *Plasmodium vivax*. *Vaccine* 17: 2959–2968.
- Udagama PV, Gamage-Mendis AC, David PH, Peiris JS, Perera KL, Mendis KN, Carter R, 1990. Genetic complexity of *Plasmodium vivax* parasites in individual human infections analyzed with monoclonal antibodies against variant epitopes on a single parasite protein. *Am J Trop Med Hyg* 42: 104–110.
- Putapornpip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho WG, Kaneko A, Kanbara H, Hattori T, Tanabe K, 2002. Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. *Proc Natl Acad Sci U S A* 99: 16348–16353.
- Cui L, Escalante AA, Imwong M, Snounou G, 2003. The genetic diversity of *Plasmodium vivax* populations. *Trends Parasitol* 19: 220–226.
- Gilson PR, Nebl T, Vukcevic D, Moritz RL, Sargeant T, Speed TP, Schofield L, Crabb BS, 2006. Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 5: 1286–1299.
- Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, Del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczyk M, Meyer EV, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, Salzberg SL, Stoeckert CJ, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM, 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455: 757–763.
- Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
- Nei M, Gojobori T, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.
- Tamura K, Dudley J, Nei M, Kumar S, 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Jones DT, Taylor WR, Thornton JM, 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275–282.
- Felsenstein J, 1993. *PHYLIP (Phylogeny Inference Package) Version 3.5c*. Seattle, WA: Department of Genetics, University of Washington.
- Thakur A, Alam MT, Sharma YD, 2008. Genetic diversity in the C-terminal 42 kDa region of merozoite surface protein-1 of *Plasmodium vivax* (PvMSP-1(42)) among Indian isolates. *Acta Trop* 108: 58–63.
- Putapornpip C, Jongwutiwes S, Ferreira MU, Kanbara H, Udomsangpet R, Cui L, 2009. Limited global diversity of the *Plasmodium vivax* merozoite surface protein 4 gene. *Infect Genet Evol* 9: 821–826.
- Thakur A, Alam MT, Bora H, Kaur P, Sharma YD, 2008. *Plasmodium vivax*: sequence polymorphism and effect of natural selection at apical membrane antigen 1 (PvAMA1) among Indian population. *Gene* 419: 35–42.
- Rayner JC, Huber CS, Feldman D, Ingravalle P, Galinski MR, Barnwell JW, 2004. *Plasmodium vivax* merozoite surface protein PvMSP-3 beta is radically polymorphic through mutation and large insertions and deletions. *Infect Genet Evol* 4: 309–319.
- Leclerc MC, Menegon M, Cligny A, Noyer JL, Mammadov S, Aliyev N, Gasimov E, Majori G, Severini C, 2004. Genetic diversity of *Plasmodium vivax* isolates from Azerbaijan. *Malar J* 3: 40.
- Niederwieser I, Felger I, Beck HP, 2001. Limited polymorphism in *Plasmodium falciparum* sexual-stage antigens. *Am J Trop Med Hyg* 64: 9–11.
- Brocchieri L, 2001. Low-complexity regions in *Plasmodium* proteins: in search of a function. *Genome Res* 11: 195–197.
- Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ Jr, Treatman C, Wang H, 2009. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 37: D539–D543.
- Pizzi E, Frontali C, 2001. Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res* 11: 218–229.
- Tanabe K, Sakihama N, Kaneko A, 2004. Stable SNPs in malaria antigen genes in isolated populations. *Science* 303: 493.