*Am. J. Trop. Med. Hyg.*, 85(6), 2011, pp. 994–1001
doi:10.4269/ajtmh.2011.11-0338


# Analysis of Polymorphisms in the *Merozoite Surface Protein*-3α Gene and Two Microsatellite Loci in Sri Lankan *Plasmodium vivax*: Evidence of Population Substructure in Sri Lanka

Mette L. Schousboe,* Rupika S. Rajakaruna, Priyanie H. Amerasinghe, Flemming Konradsen, Rosalynn Ord, Richard Pearce, Ib C. Bygbjerg, Cally Roper, and Michael Alifrangis

*Centre for Medical Parasitology, Department of International Health, Immunology and Microbiology, Copenhagen University Hospital, Denmark; Department of Zoology, University of Peradeniya, Peradeniya, Sri Lanka; International Water Management Institute, Hyderabad, Andhra Pradesh, India; Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom*

*Abstract.* The geographical distribution of genetic variation in *Plasmodium vivax* samples ($N$ = 386) from nine districts across Sri Lanka is described using three markers; the *P. vivax merozoite surface protein*-3α (*Pvmsp*-3α) gene, and the two microsatellites m1501 and m3502. At *Pvmsp*-3α, 11 alleles were found with an expected heterozygosity ($H_e$) of 0.81, whereas at m1501 and m3502, 24 alleles ($H_e$ = 0.85) and 8 alleles ($H_e$ = 0.74) were detected, respectively. Overall, 95 unique three locus genotypes were detected among the 279 samples positive at all three loci ($H_e$ = 0.95). Calculating the pairwise fixation index ($F_{ST}$) revealed statistically significant population structure. The presence of identical 2-loci microsatellite genotypes in a significant proportion of samples revealed local clusters of closely related isolates contributing to strong linkage disequilibrium between marker alleles. The results show evidence of high genetic diversity and possible population substructure of *P. vivax* populations in Sri Lanka.

## INTRODUCTION

Malaria is a serious public health problem in the tropical and sub-tropical regions of the world. Five human malaria parasite species exist, but attention is commonly focused on the most virulent and lethal species infecting humans, *Plasmodium falciparum.* However, in recent years the importance of *P. vivax* as the cause of significant morbidity and relapse of malaria has been increasingly recognized.[1,2] The publication of the complete nuclear genome sequence of the *P. vivax* strain Salvador-1 (Sal-1) opened the way for new studies of this species,[3] for instance studies of genetic diversity, which can be used as an indicator of population stability. New knowledge of the transmission dynamics of local parasite populations and understanding of how they are interconnected or isolated is important in development of suitable and effective control strategies against *Plasmodium* species, which will eventually expand our knowledge of the global distribution and diversity of *P. vivax* populations.

In this study, we have evaluated parasite genetic diversity and possible geographical clustering of *P. vivax* parasites in Sri Lanka. On the island, the two malaria species *P. falciparum* and *P. vivax* are present, with *P. vivax* being the dominating species causing about 90–95% of all malaria infections.[4,5] Throughout history, the malaria situation in Sri Lanka has been fluctuating and unstable with occasional epidemics, but during the last decade prevalence of malaria has decreased dramatically from 210,000 confirmed malaria cases in 2000 to 558 in 2009.[5] The dramatic decrease in reported malaria cases led Sri Lanka to embark on the malaria pre-elimination stage in 2008, with the target of zero incidences of locally acquired malaria infections through deliberate intervention measures to prevent re-establishment of transmission.[6,7] An important factor in this strategy is constant identification and surveillance of transmission hotspots, and the analysis of genetic diversity can assist in doing this.

Genetic diversity in *Plasmodium* parasites has been estimated by studies of allelic variation of polymorphic microsatellite (MS) markers and/or various antigen loci. The MS markers are strings of repetitive DNA (usually non-coding) that possess high polymorphism caused by strand slippage during DNA replication. They are scattered across the *Plasmodium* genomes and have use as selectively neutral markers except when in proximity to drug resistance genes or similarly selected loci.[8] The antigen loci have been used frequently in studies of genetic diversity, recently often in combination with MS loci, but unlike MS loci these are subject to immune selection and thus are not selectively neutral. Several polymorphic genes have been used to examine genetic diversity in natural populations of *P. vivax*, whereof the *P. vivax merozoite surface protein*-3α (*Pvmsp*-3α) gene is one of the most polymorphic genes analyzed to date.[4,9–13] It is a potential vaccine candidate and has been frequently used in population studies since first described in 1999.[14,15] Studies have revealed that the *Pvmsp*-3α is a member of the *msp*-3 family, which includes the three structurally related proteins; *Pvmsp*-3α, β, and γ.[16,17] It encodes a merozoite surface protein weighing between 148 and 150 kD with an alanine-rich central domain responsible for the size polymorphism caused by deletions, small insertions, and single nucleotide polymorphisms.[17–19]

The genetic diversity of the included *P. vivax* samples ($N$ = 386) was studied by analyzing the allelic diversity of three loci; the *Pvmsp*-3α gene and two MS loci, m1501 located on chromosome 1 with a 7-bp tandem repeat and m3502 located on chromosome 3 with an 8-bp tandem repeat.[20] The MS loci have high diversity and long repeat lengths,[20,21] which enable an easy distinction between different alleles/genotypes, compared with, for example, mono-, di-, and trinucleotide repeats. In this work, we studied the geographical clustering of genotypes, with the objective of evaluating the level of polymorphism of these three loci in samples from a restricted area, and to create a better understanding of the epidemiology of vivax malaria in Sri Lanka.

## MATERIALS AND METHODS

**Study area and samples.** In the main study period (September 2004–March 2006), a total of 2,717 *P. vivax* infections were recorded by the Anti Malaria Campaign (AMC), of which 2,149 cases were from the nine districts included in this study,

*Address correspondence to Mette L. Schousboe, Centre for Medical Parasitology, University of Copenhagen, CSS Building 22/23, Øster Farimagsgade 5, PO Box 2099, 1014 Copenhagen K, Denmark. E-mail: mesch@sund.ku.dk

with the majority being collected in the North-central district Anuradhapura (Figure 1).[22] However, the number of samples collected per district is not a standardized indicator of malaria incidence in the individual districts. It was influenced by the efforts to detect cases by AMC and health care services personnel, and the wide seasonally and yearly fluctuations in malaria prevalence across Sri Lanka.[23]

Generally, the country can be divided into three climatic zones; dry, wet, and intermediate, with the highest level of malaria transmission found mainly in the dry zone, which spans a large area from North to South, but malaria transmission also includes the smaller intermediate zone.[24] Another significant factor is political instability in the northern and eastern parts of Sri Lanka, which has resulted in poor health care provision in many parts up to the end of the war 2 years ago.

The *P. vivax* positive samples originate from individuals seeking treatment of malaria at government health facilities located in nine different malarious districts across Sri Lanka.[22] Finger-prick blood samples were collected routinely and dried on filter paper by staff at the facilities trained by the AMC, mainly from September 2004 to March 2006, thereby including the malaria peak transmission season in January and the shorter transmission period around July.[25] Additionally, 11 samples from the district Mannar collected in 2002, one from the district Anuradhapura collected in 2003, and a single sample from Kurunegala collected in 2007 were included in the study. The number of samples collected from each of the nine districts varied: Ampara ($N$ = 7), Anuradhapura ($N$ = 183), Batticaloa ($N$ = 14), Kurunegala ($N$ = 48), Mannar ($N$ = 15), Monaragala ($N$ = 3), Polonnaruwa ($N$ = 51), Trincomalee ($N$ = 58), and Vavuniya ($N$ = 7). Four samples from Anuradhapura and one from Vavuniya were included in the study, although only the year, and not specific collection date, was known. The extraction of DNA was carried out by the chelex-100 method.[22]

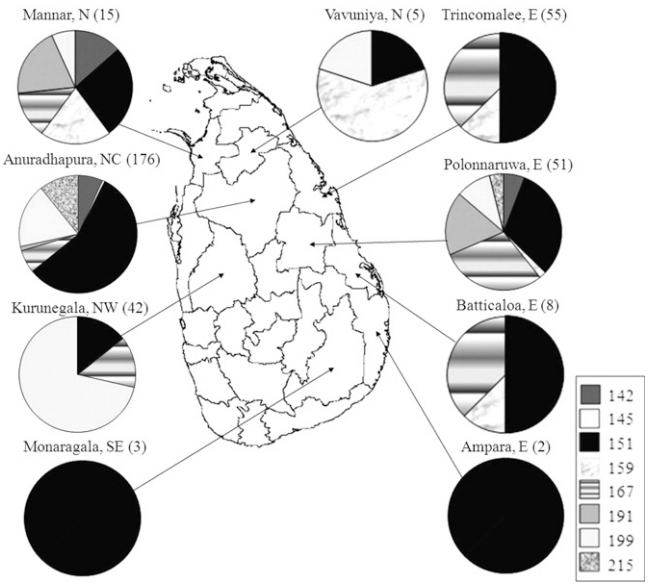


FIGURE 1. Genetic diversity of the microsatellite m3502 in nine districts of Sri Lanka. The number of positive samples from each district is mentioned in brackets. The districts from where the samples were collected are mentioned, together with the geographical location of the districts; N = North, NW = North-west, NC = North-central, E = East, and SE = South-east.

**Amplification of the *Pvmsp*-3α gene.** Primers (0.1 M) for the primary and nested polymerase chain reaction (PCR), used to amplify three major size fragments of the *Pvmsp*-3α gene (referred to as A, B, and C), are described in Bruce and others.[14] Amplification was carried out in a volume of 20 L containing 1 L template, 10 L TEMPase Hot Start Master Mix (Ampliqon, Skovlunde, Denmark), 1 mM $MgCl_2$ (Eppendorf, Hoersholm, Denmark), and $H_2O$. The thermal cycling conditions for the primary PCR were as follows: Initial denaturation at 94°C for 15 min, followed by 35 cycles of 94°C for 30 s, 53°C for 30 s, and 72°C for 2.5 min., and subsequently a 5 min extension step at 72°C. The secondary PCR was the same as the primary PCR conditions, but using 1 L of DNA from the primary reaction and with an annealing temperature at 57°C for only 30 cycles.

The PCR products were visualized under UV illumination after electrophoresis on a 0.8% agarose gel containing ethidium bromide. Positive controls for the PCR were *P. vivax*-confirmed DNA used in previous studies[22] and for the negative control samples *P. falciparum* DNA was used.

**Amplification of the microsatellite loci m1501 and m3502.** Two MS loci, m1501 and m3502, were amplified by semi-nested PCR and analyzed on an ABI 3730 XL genetic analyzer (Applied Biosystems, Foster City, CA), with a FAM fluorescent-labeled inverse primer. The primers are described by Imwong and others (2007).[20] The primary reaction included 1 L template, 0.5 unit Taq polymerase, 1.1 L Thermopol Reaction buffer (New England Biolabs Inc., Glostrup, Denmark), and 0.4 M dNTPs, 0.1 M of forward (F), and reverse primers (R) with cycling conditions as follows: 2 min at 94°C and then 25 repeated cycles of 30 s at 94°C, 30 s at 42°C, 30 s at 40°C, and 40 s at 65°C followed by 2 min at 65°C and a minimum of 10 min at 15°C. In the secondary PCR, the same concentrations of reagents were added, but with 0.15 M of reverse primers (R) and fluorescent-labeled inverse primers (I). The cycling conditions were initiated with 2 min at 94°C followed by 25 repeated cycles of 20 s at 94°C, 20 s at 45°C, 30 s at 65°C, and finished with 2 min at 65°C and 10 min at 15°C.

**Restriction fragment length polymorphism (RFLP) analysis of the amplified products of the *Pvmsp*-3α gene.** Digestion of the *Pvmsp*-3α gene with restriction enzyme *Hha I* (New England Biolabs Inc.) further divided the major-genotypes into different sub-genotypes, using the procedure described by Bruce and others (1999).[14]

A sample was considered unmixed/single if the number and size of bands observed was consistent with the presence of a single allelic sequence, i.e., 1–3 bands of fragment sizes ranging between 150 and 600 bp, and the sum of the bands per sample equalized ≤ 1.9 kb (Figure 2). If more bands than expected were seen per sample and the total size of bands exceeded 1.9 kb, the sample was considered mixed.

**Data analysis of the microsatellite loci.** The length of the MS alleles was measured by reference to the Genescan 500 Liz size standard (Applied Biosystems), using Genemapper version 3.2 (Applied Biosystems).

In the cases where multiple (≥ 2) microsatellite alleles were detected in a single sample only one allele was used for further analysis. When the electropherogram peak height of the minor allele was half the size (or less) of the major/predominant allele, the major allele was chosen to represent the sample. If the electropherogram peak height of the minor allele was more than half the height of the major allele, the allele

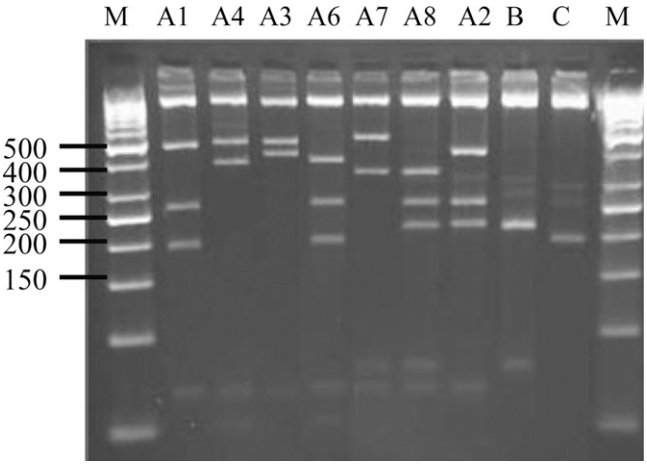


FIGURE 2. Restriction fragment length polymorphism (RFLP) with restriction enzyme *Hha I* of selected *pvmsp*-3α alleles in eight different *Plasmodium vivax* samples from Sri Lanka. With exclusion of bands above 600 bp, and below 150 bp, the estimated allele sizes of the A variants are; A1 (510, 270, 200), A2 (480, 270, 230), A3 (500, 480), A4 (500, 400), A5 (250, 210), A6 (400, 270, 200), A7 (510, 360), A8 (360, 270, 230), and A9 (440, 260, 210). A5 and A9 are not shown. The letters B and C refers to the other teo major-genotypes. DNA size markers are shown in lanes labeled "M" with sizes shown in basepairs (bp).

to represent the major one was chosen by computerized randomization.

**Statistical analysis.** Genetic diversity of the Sri Lankan *P. vivax* population was examined by calculating expected heterozygosity ($H_e$) of each locus, isolation by distance (IBD), and pairwise fixation index ($F_{ST}$). The software used for the calculations were Excel add-in MS Toolkit software, Fstat version 2.9.3 and Arlequin.[26–28] Pairwise $F_{ST}$ estimates, with 10.000 permutations, were calculated with significance based on a permutation process,[26] and used as a measure of genetic differentiation between populations (districts). To test for IBD, pairwise $F_{ST}/(1\text{-}F_{ST})$ were plotted against the natural logarithm of the geographic distance between paired sites and a test for IBD performed by use of a Partial Mantel Test (10.000 permutations) were performed.[27] The IBD calculations were performed using samples from the five districts with the highest number of samples, including both single and mixed infections, and pairwise $F_{ST}$ estimates based on the combined 3-loci genotype. Furthermore, the geographical distances between each main collection site (hospital) for each district was calculated by use of Google Earth.[29] To test for linkage disequilibrium (LD), the MS Toolkit and Arlequin software was used. These detect the presence of significant non-random association between alleles at the two loci, m1501 and m3502, which is measured by the formula $D' = D/D_{max}$.[26] The LD coefficient, $D$, measures derivation of random association between alleles at different loci, while $D'$ equals $D$ standardized by the maximum value ($D_{max}$), given the allele frequencies. The $D'$ coefficient ranges from –1 to +1, with the value 1 referring to a complete non-random association between the alleles. Only unmixed samples were used for the LD analysis.

## RESULTS

**Genetic diversity of the *Pvmsp*-3α gene.** Of the 386 *P. vivax* samples available, 314 (81.3%) *Pvmsp*-3α fragments were successfully PCR amplified. The fragments varied in size from ~1.9kb (A) to 1.4kb (B), and 1.1 kb (C) with 4.8% ($N$ = 15) being mixed. The vast majority of the samples carried the A major-genotype ($N$ = 266, 84.4%), whereas only 13.1% ($N$ = 41) and 2.2% ($N$ = 7) of the samples carried the B and C types, respectively. A refinement of the allelic diversity was obtained by digestion with restriction enzyme *Hha I* which subdivided the A-type alleles into nine allelic variants named A1–A9 (Figure 2, Table 1). The sub-classification of A alleles resulted in an increased detection rate of mixed samples to 8.0% ($N$ = 25/311). The B and C alleles did not digest into further alleles. The distribution of the three major- and 11 sub-types were unevenly distributed across Sri Lanka as shown in Table 1. The A type was found in all districts, but at different frequencies; from complete dominance in the NW and SE districts to only 25% in the North (Mannar). The most common sub-genotype, A1, accounted for 42% of the A-type alleles, and this allele also varied in prevalence across the island. Two alleles were district specific, the A5 allele in the eastern Polonnaruwa district and the A9 allele in the two northern districts Mannar and Trincomalee (Table 1). Other alleles were more widely dispersed, either at districts geographically far from each other or clustered mainly in Northern and Eastern districts. Likewise, an uneven distribution of the B and C type alleles was observed (Table 1). After digestion, significant between-districts differentiation was confirmed for all district comparisons with $F_{ST}$ estimates between 0.12 and 0.44 ($P < 0.0001$) (Table 2).

**Genetic diversity of the microsatellite loci and isolation by distance.** Significant diversity was observed at the two MS loci, m1501 and m3502. At the m1501 locus, 352 samples were amplified and a total of 24 different alleles were found ($H_e$ = 0.85), with 21.0% found to be mixed ($N$ = 74) (Table 3). One common allele, "128," observed in 32.7% ($N$ = 115) of all isolates and with significant variation in frequency across the island (Table 4). Some alleles were district specific and only found once, whereas others were present in 2–6 districts (Table 4). Likewise, the $H_e$ values differed between the districts from 0.68 to 0.88 (Table 3). The highest $H_e$ estimate was found in Mannar, though low sample size and long time span of collection might bias the result.

At the m3502 locus, eight alleles were found among the 357 amplified samples ($H_e$ = 0.74, Table 3) and only 6.2% ($N$ = 22) of samples were found to be mixed. The geographical distribution of alleles is shown in Figure 1. No obvious geographical isolation was observed, although there was a tendency of allele "199" to be dominating in Kurunegala, and allele "159" to be more common in the Northern districts (Figure 1). Allele "151" was the most commonly observed allele with an overall frequency of 44.5% ($N$ = 159), and was present in all districts with frequencies ranging from 14.3% to 100% across the country. The maximum number of different alleles observed per district was seven (Anuradhapura and Polonnaruwa). As with the m1501 loci, the district-wise $H_e$ estimates differed, ranging from 0.46–0.87, with samples from Kurunegala possessing the lowest $H_e$ estimates, while samples from Mannar possessed the highest $H_e$ estimates (Table 3).

Isolation by distance was calculated to determine any relationship between genetic diversity and geographical distance between districts, however no significant association was found using the Partiel Mantel Test ($P$ = 0.73).

All $H_e$ and $F_{ST}$ calculations were performed both by including and excluding mixed samples. Because either no, or minor,

TABLE 1

Genetic diversity of the *Pvmsp*-3α gene in *Plasmodium vivax* samples from nine districts of Sri Lanka analyzed with restriction fragment length polymorphism (RFLP)*

| *Pvmsp*-3α | Kurunegala (NW) | Anuradhapura (NC) | Mannar (N) | Vavuniya (N) | Trincomalee (E) | Polonnaruwa (E) | Batticaloa (E) | Ampara (E) | Monaragala (SE) | All districts |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 1 | 88 | | 1 | 15 | 6 | 1 | | | 111 |
| A2 | 5 | 18 | | 45 | 1 | 15 | 2 | 2 | 2 | 45 |
| A3 | | 6 | 2 | 16 | 3 | 4 | | 1 | | 16 |
| A4 | | 11 | | 16 | | 4 | 1 | | | 16 |
| A5 | | | | | | 1 | | | | 1 |
| A6 | 5 | 5 | | | 20 | | | | | 30 |
| A7 | | 1 | | 4 | 3 | 4 | | | | 12 |
| A8 | 23 | 2 | | | 2 | | 3 | | | 30 |
| A9 | | | 1 | | 1 | | | | | 2 |
| B | | 19 | 7 | | 1 | 11 | 1 | 2 | | 41 |
| C | | | 2 | 1 | 3 | 1 | 1 | | | 7 |
| *N*/alleles | 34/4 | 149/8 | 12/4 | 6/3 | 49/9 | 45/7 | 9/6 | 5/3 | 2/1 | 311/11 |

*The districts from where the samples were collected are mentioned, together with the geographical location of the districts; N = North, NW = North-west, NC = North-central, E = East, and SE = South-east. The total number of positive samples and alleles are mentioned in the last row of the figure.

effect of excluding the mixed infections was found, the results described in the text, tables, and figures are based on the total sample set, unless mentioned otherwise.

**Temporal differentiation.** To test if years of sample collection had an effect on the observed genetic differentiation across Sri Lanka, two tests were performed. First, the within-district diversity per locus for each of the 3 years (2004–2006) with the highest number of analyzed samples was analyzed by calculation of $H_e$ estimates (Table 3). The difference in genetic diversity between the 3 years was minor, although differences in samples sizes might bias the calculations. Second, the samples from the following districts were divided into years to test for any genotypic differences between the years, measured by $F_{ST}$: Anuradhapura (2004, 2005, and 2006), Polonnaruwa (2004 and 2005), Kurunegala (2005 and 2006), and Trincomalee (2005 and 2006) (data not shown). The remaining districts were not analyzed because of sample size limitations. Solely the 2-loci MS genotypes were used for the $F_{ST}$ calculations, because 3-loci genotypes narrowed down sample size too much. The only temporal differentiation was found in Kurunegala, although with low $F_{ST}$ estimates ($F_{ST} = 0.15$, $P = 0.007$) and Anuradhapura ($F_{ST} = 0.01$, $P = 0.049$).

Since the majority of samples from Mannar were collected in 2002 ($N = 11$), thus beyond the main sample collection period from 2004 to 2006, temporal changes in this particular district has been investigated separately. Results from PCR and RFLP of the *Pvmsp*-3α gene revealed a possible switch from the B type in 2002 to the A and C type in 2004–2005. The two MS markers did not reveal any radical change, although the m165 allele dominated at the m1501 locus in 2002, while absent in 2004–2005. However, the low sample sizes hindered any statistical calculations.

**Linkage disequilibrium.** When variation at the two unlinked MS markers was combined and the resulting 2-loci genotype was analyzed, substructuring in the total sample set was apparent (Table 5). In particular, certain 2-loci genotypes dominated in specific districts, e.g., the "128-151" genotype in Anuradhapura, the "241-199" genotype in Kurunegala, and the "179-159" genotype in Trincomalee (Table 5). Therefore, LD was calculated for the total Sri Lankan *P. vivax* population, and by district and year. For all LD calculations, the unmixed sample set was used, whereas only the alleles occurring more than two times were included in the LD calculation. Despite the different chromosomal location, highly significant LD was found between the "206-142" and "241-199" genotypes ($D' = 1, P < 0.0001$) when pooling all samples from Sri Lanka (Table 5). When calculating LD district wise, a significant LD could be detected for certain 2-loci genotypes in Anuradhapura, Polonnaruwa, Trincomalee, and Kurunegala ($D' = 1, P < 0.0001$).

## DISCUSSION

The last decade has seen significant advancements in *P. vivax* research, but much is still unknown about the epidemiology and diversity of *P. vivax* parasites in natural populations. This study investigated the level of allelic diversity at the *Pvmsp*-3α gene and two MS markers in 386 *P. vivax* field samples from nine districts in Sri Lanka.

TABLE 2

Genetic differentiation between *Plasmodium vivax* parasites sampled in five districts in Sri Lanka estimated by pairwise fixation index, $F_{ST}$*

| Locations compared | *Pvmsp*-3α $F_{ST}$ | m1501 $F_{ST}$ | m3502 $F_{ST}$ | m1501–3502 $F_{ST}$ | 3 loci $F_{ST}$ |
|---|---|---|---|---|---|
| Anuradhapura/Kurunegala | 0.39 *** | 0.24 *** | 0.24 *** | 0.21 *** | 0.17 *** |
| Anuradhapura/Mannar | 0.31 *** | 0.14 ** | 0.14 ** | 0.11 ** | 0.09 *** |
| Anuradhapura/Polonnaruwa | 0.16 *** | 0.06 ** | 0.06 ** | 0.09 *** | 0.08 *** |
| Anuradhapura/Trincomalee | 0.15 *** | 0.19 *** | 0.19 *** | 0.15 *** | 0.13 *** |
| Kurunegala/Mannar | 0.44 *** | 0.22 ** | 0.22 *** | 0.19 ** | 0.13 |
| Kurunegala/Polonnaruwa | 0.30 *** | 0.25 *** | 0.25 *** | 0.20 *** | 0.13 *** |
| Kurunegala/Trincomalee | 0.29 *** | 0.27 *** | 0.27 *** | 0.24 *** | 0.19 *** |
| Mannar/Polonnaruwa | 0.12 *** | 0.08 | 0.08 | 0.05 | 0.03 |
| Mannar/Trincomalee | 0.26*** | 0.12* | 0.12* | 0.11 ** | 0.10 |
| Polonnaruwa/Trincomalee | 0.17 *** | 0.14 *** | 0.14 *** | 0.12 *** | 0.10 *** |

*The 3 loci mentioned are the gene *Pvmsp*-3α after restriction fragment length polymorphism (RFLP), and the two microsatellites m1501 and m3502. *P* values ≤ 0.0033 are considered significant after Bonferroni multiple test correction, and are marked with asterisks; * < 0.003–0.001, ** < 0.001–0.0001, and *** < 0.0001. The non-significant *P* values, after Bonferroni correction, range from 0.180 to 0.004.

TABLE 3

Genetic diversity, measured by expected heterozygosity ($H_e$), of *Plasmodium vivax* parasites collected in districts in Sri Lanka, by each of the three *P. vivax* loci; the gene *Pvmsp*-3α (after restriction fragment length polymorphism [RFLP]), and the two microsatellites m1501 and m3502, and the combined 2-MS and 3-loci genotypes*

| | *Pvmsp*-3α | m1501 | m3502 | m1501–3502 | *Pvmsp*-3α-m1501–m3502 |
|---|---|---|---|---|---|
| Kurunegala (NW) | 0.51 (34/4) | 0.68 (41/7) | 0.46 (42/3) | 0.69 (30/9) | 0.78 (29/10) |
| Anuradhapura (NC) | 0.62 (149/8) | 0.74 (176/16) | 0.63 (176/7) | 0.82 (172/29) | 0.86 (142/43) |
| Mannar (N) | 0.65 (12/4) | 0.88 (15/8) | 0.87 (15/6) | 0.97 (15/12) | 0.98 (12/11) |
| Trincomalee (E) | 0.74 (49/9) | 0.75 (51/8) | 0.66 (55/5) | 0.82 (50/15) | 0.83 (44/17) |
| Polonnaruwa (E) | 0.81 (45/7) | 0.77 (48/14) | 0.78 (51/7) | 0.90 (48/19) | 0.95 (42/24) |
| $H_e$, 9 districts | 0.81 (311/11) | 0.85 (352/24) | 0.74 (357/8) | 0.92 (336/63) | 0.95 (279/95) |
| 2004 | 0.67 (109/8) | 0.73 (116/17) | 0.67 (119/7) | 0.83 (115/28) | 0.88 (106/40) |
| 2005 | 0.83 (154/10) | 0.87 (180/20) | 0.71 (176/7) | 0.93 (167/41) | 0.96 (134/59) |
| 2006 | 0.65 (36/6) | 0.77 (40/9) | 0.60 (46/5) | 0.78 (38/12) | 0.84 (28/11) |
| 2004–2006 | 0.81 (299/11) | 0.85 (336/24) | 0.73 (341/8) | 0.91 (320/60) | 0.94 (268/88) |

*The first 6 rows mention the genetic diversity per each of the five districts with sample sizes above 10 samples, and for all districts, whereas the last 4 rows of the table mention genetic diversity per 2004, 2005, and 2006, and for 2004–2006. In brackets are mentioned number of positive samples and alleles by locus and district. The geographical locations are abbreviated as; N = North, NW = North-west, NC = North-central, and E = East.

The *Pvmsp*-3α gene was found to be a highly polymorphic marker as shown by other *P. vivax* population studies from India, Sri Lanka, Thailand, Iran, and Papua New Guinea.[4,9–13,30] Using only PCR, three allelic types (A, B, and C) were identified and these revealed low and significant differentiation within and between populations in Sri Lanka. Although the A and B types were scattered across the country, the C type was more commonly observed in the Northern and some Eastern districts of Sri Lanka (Table 1). Because a high number of C alleles have been found in India,[12] it could be hypothesized that this allelic variant is transferred by migration of people between Sri Lanka and India, mainly the Indian state Tamil Nadu where many refugees moved to, although the malaria transmission in this area was low at the time of the armed conflict (Konradsen F, personal communication). Conversely, the limited similarity in *P. vivax* populations between Mannar and the more Southern populations might signal isolation caused by the long-lasting conflict in the Northern, North-eastern, and some Eastern parts of Sri Lanka. With the exception of the constant movement of armed services personnel who might have transported *P. vivax* infections around, the conflict has severely hindered the movement of people, and thereby parasites, back and forth between the affected areas. However, a recent study of the *Pvmsp*-3α gene examined 13 samples collected in 2000 in Kataragama in the Southern Sri Lanka, and it detected A ($N$ = 11) and C ($N$ = 2) variants.[4] Although the sample size was small, the presence of the C variant and absence of B variant may suggest that the regional trend we observe may not be consistent from year to year and varies at more local levels. In this study, the low number of C types detected was unusual and differs from observations from other groups.[4,10,12,18,31,32]

Digestion of the nested *Pvmsp*-3α PCR products with *Hha I* revealed further genetic variation within the A-type allele.

TABLE 4

Distribution of 24 alleles at the m1501 microsatellite locus among 9 districts in Sri Lanka*

| m1501 | Kurunegala (NW) | Anuradhapura (NC) | Mannar (N) | Vavuniya (N) | Trincomalee (E) | Polonnaruwa (E) | Batticaloa (E) | Ampara (E) | Monaragala (SE) | All districts |
|---|---|---|---|---|---|---|---|---|---|---|
| 79 | | 1 | | | | | | | | 1 |
| 86 | | | | 1 | | | | | | 1 |
| 93 | | | | | 1 | | | | | 1 |
| 100 | 2 | 7 | 2 | | 2 | | | | | 13 |
| 107 | | 11 | 2 | 1 | 11 | 16 | | 1 | | 42 |
| 114 | | 2 | 1 | 1 | | 1 | | | 2 | 7 |
| 128 | 4 | 83 | 2 | | 6 | 17 | 3 | | | 115 |
| 139 | | 3 | | | | 1 | | | | 4 |
| 142 | | | 1 | | | | | | | 1 |
| 150 | | 18 | | | 4 | 2 | 1 | 1 | | 26 |
| 158 | | | 1 | 3 | | 1 | | | | 5 |
| 165 | | 1 | 5 | | | 2 | | 2 | | 10 |
| 172 | | 3 | | | | | | 1 | | 4 |
| 179 | | | 1 | 1 | 22 | | 1 | | | 25 |
| 185 | 7 | 4 | | | | 1 | | | | 12 |
| 192 | 1 | | | | | 2 | | | | 3 |
| 199 | | | | | 4 | 1 | | | | 5 |
| 206 | | 6 | | | | 1 | 1 | | | 8 |
| 213 | | | | | | 1 | 1 | | | 2 |
| 220 | | 4 | | | | 1 | | | | 5 |
| 227 | 1 | 28 | | | 1 | 1 | | | | 31 |
| 241 | 22 | 3 | | | | | | | | 25 |
| 247 | 4 | 1 | | | | | | | | 5 |
| 283 | | 1 | | | | | | | | 1 |
| $N$/alleles | 41/7 | 176/16 | 15/8 | 7/5 | 51/8 | 48/14 | 7/5 | 5/4 | 2/1 | 352/24 |

*Sample size and number of different alleles by district and for all 9 districts are included in the table. In brackets are abbreviations of regional location of the districts; N = North, NW = North-west, NC = North-central, E = East, and SE = South-east.

TABLE 5
Genetic diversity for the combined genotype of the two microsatellite loci, m1501 and m3502 for all districts analyzed in Sri Lanka*

| m1501–m3502 | Kurunegala (NW) | Anuradhapura (NC) | Mannar (N) | Vavuniya (N) | Trincomalee (E) | Polonnaruwa (E) | Batticaloa (E) | Monaragala (SE) | All districts |
|---|---|---|---|---|---|---|---|---|---|
| 128–151 | 3 | 48 | | | 2 | 5 | | | 58 |
| 241–199† | 16 | 3 | | | | | | | 19 |
| 179–159 | | | | | 17 | | | | 17 |
| 107–167 | | 4 | 1 | | | 10 | | | 15 |
| 107–151 | | 3 | | | 8 | 3 | | | 14 |
| 227–199 | 1 | 13 | | | | | | | 14 |
| 227–151 | | 12 | | | 1 | | | | 13 |
| 150–215 | | 11 | | | | | | | 11 |
| 128–199 | | 7 | | | | 3 | | | 10 |
| 100–151 | 1 | 6 | 1 | | 1 | | | | 9 |
| 185–167 | 5 | 3 | | | | 1 | | | 9 |
| 128–191 | | 1 | | | | 5 | | | 6 |
| 206–142† | | 6 | | | | | | | 6 |
| 150–167 | | | | | 3 | | 1 | | 4 |
| 220–142 | | 4 | | | | | | | 4 |
| 114–151 | | 1 | | | | | | 2 | 3 |
| 158–159 | | | | 2 | | 1 | | | 3 |
| 165–151 | | | 1 | | | 2 | | | 3 |
| 165–191 | | 1 | 2 | | | | | | 3 |
| 179–199 | | | 1 | | 2 | | | | 3 |
| n/alleles | 30/9 | 136/24 | 12/11 | 2/1 | 38/11 | 37/14 | 2/2 | 2/1 | 259/49 |

* Excluded in the table are genotypes with an overall count below 2, all mixed samples, and samples not positive at both loci. The district Ampara is excluded because the district only included 2-loci genotypes with an overall count of 1. Sample size and number of different alleles by district for all 9 districts are included in the table. In brackets are abbreviations of the geographical location of the districts, N = North, NW = North-West, NC = North-Central, E = East, and SE = South-East.
† Highly significant linkage disequilibrium between the two microsatellite alleles are found when pooling all samples into one population.

None of the digestion patterns correspond to that reported for the reference strains Belem and Sal-1.[14] The trend of substructuring was further validated with this technique, with a few alleles being district specific while others were more commonly distributed with diverging frequencies. Differentiation between districts revealed significant differences with $F_{ST}$ estimates between 0.15 and 0.44 ($P < 0.0001$). Focusing solely on the A alleles, this study detected nine different digestion patterns, with allele A1 as the predominant allele. The dominance of a few alleles is also seen in another study from Sri Lanka,[30] and limits the usefulness of the *Pvmsp*-3α gene as a solitary marker for studies aiming at distinguishing between individual parasites, e.g., when evaluating parasites' drug response where highly polymorphic loci are needed.

The polymorphism at two MS loci, m1501 and m3502, was analyzed to test the validity of conclusions based on the *Pvmsp*-3α gene using markers that are highly polymorphic yet selectively neutral.[20,21] The two markers showed different levels of polymorphism, the m1501 being three times more polymorphic than m3502. The trend of m1501 being the most polymorphic marker corresponds with a recent study from Papua New Guinea, which found m1501 to be more polymorphic than m3502, though with a difference in polymorphism of only 1.5 times (m1501 $N = 19$, m3502 $N = 13$ allelic variants).[21] Another study, including samples from South America and Asia, reports the m3502 to be 1.4 times more polymorphic than m1501 (m3502 $N = 23$, m1501 $N = 17$ allelic variants).[20] Furthermore, the latter study observed a much higher diversity among the Asian samples compared with the South American samples, with 10 m1501 and 16 m3502 allelic variants specific for the Asian samples. Thus, the diversity of these MS markers seems to be determined by origin and parasite population rather than an intrinsic property of the locus itself.

In this study, the diversity by locus differed remarkably between populations. It was generally high in all districts apart from Kurunegala, which possessed the lowest allelic diversity and the greatest allelic distinctiveness from other populations. The temporal distribution of alleles on the two MS loci within Kurunegala was examined and clearly showed that a specific 2-loci genotype increased in frequency from 2005 to 2006 (data not shown). Together with calculation of a highly significant LD between these alleles in Kurunegala, it might indicate an epidemic expansion of this specific *P. vivax* genotype. Interestingly, an actual outbreak of vivax malaria in Kurunegala during November 2005 to February 2006 has been reported, which might explain our results. That MS loci are useful for mapping substructuring and site-specific outbreaks have also been reported by a recent study from Sri Lanka, Myanmar, and Ethiopia.[33]

When combining all three loci, the mean diversity in five of the districts was high, although a low but distinct population differentiation existed between most districts. Analyzing the between-districts differentiation confirmed statistically the presence of substructuring between all districts, though any isolation caused by geographical distance could not be confirmed by regression analysis. When evaluating the within-district diversity this further revealed clear district specificity of the alleles. This trend was especially apparent in the district of Kurunegala, which stands out from the others as relatively undiverse, with the lowest $H_e$ of all districts and a strong LD between specific MS alleles. All of these observations emphasize the presence of high genetic diversity and spatial and temporal substructuring in Sri Lanka, despite the low malaria transmission intensities. This corresponds with the finding of up to 22 different *Pvmsp*-3α genotypes among 196 samples collected in three districts in Sri Lanka,[30] and another study from Sri Lanka detecting high complexity of the *Pvmsp*-1 gene ($N = 95$).[34] On the other hand, it contradicts studies of genetic diversity in *P. falciparum*, wherein a positive relationship between genetic complexity and transmission intensities is found.[35,36] However, the genetic complexity of local *P. vivax* parasites might be less associated with transmission intensities, possibly because

of special biological features such as early gametocytogenesis and the ability to relapse. Regarding relapses, a study has found that relapsing parasites often are of a different genotype than those that dominated the initial infection,[37] and together with the presence of continuous relapse infections this will increase genetic diversity of vivax populations. However, relapses on Sri Lanka are expected to be low because of the general recommendations of treating vivax infections with a combination of chloroquine and the hypnozoitocidal drug primaquine.

In conclusion, our study found that the three loci together are efficient markers for estimating genetic diversity of *P. vivax* populations in Sri Lanka. The patterns of diversity indicate local transmission and may be used to describe occasional epidemic expansions of *P. vivax* infections, which is relevant for current efforts of malaria elimination.

Received May 27, 2011. Accepted for publication September 12, 2011.

Acknowledgments: The Regional Medical Officers and the technical personnel in Sri Lanka are gratefully acknowledged for supplying the filter paper blood samples. Sharmini Gunawardena from the Department of Immunology and Infectious Diseases, Harvard School of Public Health, USA, is thanked for assistance in statistical calculations, technician Ulla Abildtrup for her excellent support in the laboratory, Kevin Tetteh and Lindsay Stewart (London School of Hygiene and Tropical Medicine), Sylvia Mathiasen and Line Vej Ugelvig (University of Copenhagen) are thanked for their support with MS analysis and statistical guidance.

Financial support: We thank "The Danish Agency for Science, Technology and Innovation" for financial support to fulfill this study.

Disclaimer: No conflicts of interest to declare.

Ethical statement: Ethical clearance for this project was granted by the Committee on Research and Ethical Review at the Faculty of Medicine, Peradeniya, Kandy and verbal consent was obtained from participants, parents, and/or guardians.

Authors' addresses: Mette L. Schousboe, Flemming Konradsen, Ib C. Bygbjerg, and Michael Alifrangis, Centre for Medical Parasitology, University of Copenhagen, Copenhagen K, Denmark, E-mails: mesch@sund.ku.dk, flko@sund.ku.dk, iby@sund.ku.dk, and micali@sund.ku.dk. Rupika S. Rajakaruna, Department of Zoology, Faculty of Science, University of Peradeniya, Peradeniya (20400), E-mail: rupika.r@gmail.com. Priyanie H. Amerasinghe, International Water Management Institute, Hyderabad, Andhra Pradesh, India, E-mail: p.amerasinghe@cgiar.org. Rosalynn Ord, Richard Pearce, and Cally Roper, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom, E-mails: ordrosalynn@yahoo.co.uk, richard.j.pearce@gmail.com, and Cally.Roper@lshtm.ac.uk.

## REFERENCES

1. Mendis K, Sina BJ, Marchesini P, Carter R, 2001. The neglected burden of *Plasmodium vivax* malaria. *Am J Trop Med Hyg 64:* 97–106.
2. 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.
3. 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, Korsinczky 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.
4. Aresha M, Sanath M, Deepika F, Renu W, Anura B, Chanditha H, Wimaladharma A, Rajitha W, 2008. Genotyping of *Plasmodium vivax* infections in Sri Lanka using *Pvmsp*-3alpha and Pvcs genes as markers: a preliminary report. *Trop Biomed 25:* 100–106.
5. World Health Organization, 2010. *World Malaria Report 2010*. Geneva: WHO.
6. Dowdle WR, 1998. The principles of disease elimination and eradication. *Bull World Health Organ 76 (Suppl 2):* 22–25.
7. Rajakaruna RS, Alifrangis M, Amerasinghe PH, Konradsen F, 2010. Pre-elimination stage of malaria in Sri Lanka: assessing the level of hidden parasites in the population. *Malar J 9:* 1–6.
8. Russell B, Suwanarusk R, Lek-Uthai U, 2006. *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends Parasitol 22:* 399–401.
9. Cole-Tobian JL, Biasor M, King CL, 2005. High complexity of *Plasmodium vivax* infections in Papua New Guinean children. *Am J Trop Med Hyg 73:* 626–633.
10. Cui L, Mascorro CN, Fan Q, Rzomp KA, Khuntirat B, Zhou G, Chen H, Yan G, Sattabongkot J, 2003. Genetic diversity and multiple infections of *Plasmodium vivax* malaria in Western Thailand. *Am J Trop Med Hyg 68:* 613–619.
11. Joshi H, 2003. Markers for population genetic analysis of human plasmodia species, *P. falciparum* and *P. vivax*. *J Vector Borne Dis 40:* 78–83.
12. Kim JR, Imwong M, Nandy A, Chotivanich K, Nontprasert A, Tonomsing N, Maji A, Addy M, Day NP, White NJ, Pukrittayakamee S, 2006. Genetic diversity of *Plasmodium vivax* in Kolkata, India. *Malar J 5:* 1–10.
13. Zakeri S, Barjesteh H, Djadid ND, 2006. Merozoite surface protein-3alpha is a reliable marker for population genetic analysis of *Plasmodium vivax*. *Malar J 5:* 1–6.
14. Bruce MC, Galinski MR, Barnwell JW, Snounou G, Day KP, 1999. Polymorphism at the merozoite surface protein-3alpha locus of *Plasmodium vivax*: global and local diversity. *Am J Trop Med Hyg 61:* 518–525.
15. Galinski MR, Corredor-Medina C, Povoa M, Crosby J, Ingravallo P, Barnwell JW, 1999. *Plasmodium vivax* merozoite surface protein-3 contains coiled-coil motifs in an alanine-rich central domain. *Mol Biochem Parasitol 101:* 131–147.
16. Galinski MR, Ingravallo P, Corredor-Medina C, Al-Khedery B, Povoa M, Barnwell JW, 2001. *Plasmodium vivax* merozoite surface proteins-3beta and-3gamma share structural similarities with *P. vivax* merozoite surface protein-3alpha and define a new gene family. *Mol Biochem Parasitol 115:* 41–53.
17. Rayner JC, Corredor V, Feldman D, Ingravallo P, Iderabdullah F, Galinski MR, Barnwell JW, 2002. Extensive polymorphism in the *Plasmodium vivax* merozoite surface coat protein MSP-3alpha is limited to specific domains. *Parasitology 125:* 393–405.
18. Ord R, Polley S, Tami A, Sutherland CJ, 2005. High sequence diversity and evidence of balancing selection in the *Pvmsp*3alpha gene of *Plasmodium vivax* in the Venezuelan Amazon. *Mol Biochem Parasitol 144:* 86–93.
19. Veron V, Legrand E, Yrinesi J, Volney B, Simon S, Carme B, 2009. Genetic diversity of msp3alpha and msp1_b5 markers of *Plasmodium vivax* in French Guiana. *Malar J 8:* 1–8.
20. Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, Mayxay M, Newton PN, Kim JR, Nandy A, Osorio L, Carlton JM, White NJ, Day NP, Anderson TJ, 2007. Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. *Int J Parasitol 37:* 1013–1022.
21. Koepfli C, Mueller I, Marfurt J, Goroti M, Sie A, Oa O, Genton B, Beck HP, Felger I, 2009. Evaluation of *Plasmodium vivax* genotyping markers for molecular monitoring in clinical trials. *J Infect Dis 199:* 1074–1080.
22. Schousboe ML, Rajakaruna RS, Salanti A, Hapuarachchi HC, Galappaththy GN, Bygbjerg IC, Amerasinghe PH, Konradsen F, Alifrangis M, 2007. Island-wide diversity in single nucleotide polymorphisms of the *Plasmodium vivax* dihydrofolate reductase and dihydropteroate synthetase genes in Sri Lanka. *Malar J 6:* 1–6.
23. Zubair L, Connor SJ, Ward N, Galapaththy GN, Mahanama S, Faizal HM, Yahiya Z, Sirwardhana M, Tennakoon U, Chandimala J, Razick S, Nawarathna B, Shaman J, Ndiaye O, Yang H, 2007. Analysis of impacts of climate variability on malaria transmission and the development of an early warning system. Project

Report to NOAA, NSF, EPA and EPRI Joint Program on Climate Variability and Human Health, December 2007.
24. Konradsen F, Amerasinghe PH, Perera D, van der Hoek W, Amerasinghe FP, 2000. A village treatment center for malaria: community response in Sri Lanka. *Soc Sci Med 50:* 879–889.
25. Briet OJ, Gunawardena DM, van der Hoek W, Amerasinghe FP, 2003. Sri Lanka malaria maps. *Malar J 2:* 1–10.
26. Excoffier L, Laval G, Schneider S, 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online 1:* 47–50.
27. Goudet J, 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at: http://www unil ch/izea/softwares/fstat html.
28. Park SD, 2001. Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis, University of Dublin, Dublin, Ireland.
29. Google, 2011. Google Earth.
30. Wickramarachchi T, Premaratne PH, Dias S, Handunnetti SM, Udagama-Randeniya PV, 2010. Genetic complexity of *Plasmodium vivax* infections in Sri Lanka, as reflected at the merozoite-surface-protein-3alpha locus. *Ann Trop Med Parasitol 104:* 95–108.
31. Bruce MC, Galinski MR, Barnwell JW, Donnelly CA, Walmsley M, Alpers MP, Walliker D, Day KP, 2000. Genetic diversity and dynamics of *Plasmodium falciparum* and *P. vivax* populations in multiply infected children with asymptomatic malaria infections in Papua New Guinea. *Parasitology 121:* 257–272.
32. Mueller I, Kaiok J, Reeder JC, Cortes A, 2002. The population structure of *Plasmodium falciparum* and *Plasmodium vivax* during an epidemic of malaria in the Eastern Highlands of Papua New Guinea. *Am J Trop Med Hyg 67:* 459–464.
33. Gunawardena S, Karunaweera ND, Ferreira MU, Phone-Kyaw M, Pollack RJ, Alifrangis M, Rajakaruna RS, Konradsen F, Amerasinghe PH, Schousboe ML, Galappaththy GNL, Abeyasinghe RR, Hartl DH, Wirth DF, 2010. Geographic structure of *Plasmodium vivax*: microsatellite analysis of parasite populations from Sri Lanka, Myanmar and Ethiopia. *Am J Trop Med Hyg 82:* 235–242.
34. Dias S, Longacre S, Escalante AA, Udagama-Randeniya PV, 2011. Genetic diversity and recombination at the C-terminal fragment of the merozoite surface protein-1 of *Plasmodium vivax* (*PvMSP*-1) in Sri Lanka. *Infect Genet Evol 11:* 145–156.
35. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP, 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol 17:* 1467–1482.
36. Babiker HA, Lines J, Hill WG, Walliker D, 1997. Population structure of *Plasmodium falciparum* in villages with different malaria endemicity in east Africa. *Am J Trop Med Hyg 56:* 141–147.
37. Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, Guthmann JP, Nosten F, Carlton J, Looareesuwan S, Nair S, Sudimack D, Day NP, Anderson TJ, White NJ, 2007. Relapses of *Plasmodium vivax* infection usually result from activation of heterologous hypnozoites. *J Infect Dis 195:* 927–933.