Geographic Structure of *Plasmodium vivax*: Microsatellite Analysis of Parasite Populations from Sri Lanka, Myanmar, and Ethiopia


Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts; Department of Parasitology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka; Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil; Parasitology Research Division, Department of Medical Research (Lower Myanmar), Myanmar; Centre for Medical Parasitology, Department of International Health, Immunology and Microbiology, University of Copenhagen, Copenhagen, Denmark; Departments of Infectious Diseases and Clinical Microbiology, Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark; Department of Zoology, University of Peradeniya, Peradeniya, Sri Lanka; International Water Management Institute, Delhi, India; Anti-Malaria Campaign, Ministry of Health, Colombo, Sri Lanka; Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts

Abstract. Genetic diversity and population structure of *Plasmodium vivax* parasites can predict the origin and spread of novel variants within a population enabling population specific malaria control measures. We analyzed the genetic diversity and population structure of *P. vivax* isolates from Sri Lanka, Myanmar, and Ethiopia using 12 trinucleotide and tetranucleotide microsatellite markers. All three parasite populations were highly polymorphic with 3–44 alleles per locus. Approximately 65% were multiple-clone infections. Mean genetic diversity (H) was 0.7517 in Ethiopia, 0.8450 in Myanmar, and 0.8610 in Sri Lanka. Significant linkage disequilibrium was maintained. Population structure showed two clusters (Asian and African) according to geography and ancestry. Strong clustering of outbreak isolates from Sri Lanka and Ethiopia was observed. Predictive power of ancestry using two-thirds of the isolates as a model identified 78.2% of isolates accurately as being African or Asian. Microsatellite analysis is a useful tool for mapping short-term outbreaks of malaria and for predicting ancestry.

INTRODUCTION

Malaria caused by *Plasmodium vivax* is often regarded as a benign and self-limiting infection. However, there is increasing evidence that the overall burden, economic impact, and the severity of disease caused by *P. vivax* have been underestimated. The current knowledge on genetic diversity and population structure of *P. vivax* lags behind that of *P. falciparum*, mostly because of lack of appropriate genetic markers for the *P. vivax* genome and its inability to be propagated continuously in the laboratory except in non-human primates. Sequencing of the *P. vivax* genome was completed recently and provides a valuable resource adding stimulus to the much needed study of this neglected species.

Analysis of allelic variation at multiple independent loci provides an effective means to determine population structure. Microsatellites have been estimated to mutate at high rates, and the mechanism of mutation appears to be strand slippage at replication leading to either a lengthening or shortening of the DNA. This variation in length is assumed to give rise to new electrophoretically distinguishable alleles. Thus, genetic diversity and evolutionary history inferred from microsatellite data are based on similarity of electrophoretic mobility, rather than by descent.

The distinct pathologic and epidemiologic nature of *P. vivax* malaria is caused by several biological characteristics of *P. vivax*, which makes it differ from *P. falciparum*. *P. vivax* forms hypnozoite stages in the liver, which may persist for months or years before initiating relapses (secondary infections), which provide a mechanism for the parasite to hibernate during less optimal transmission periods. Differences in *Anopheles* mosquito dynamics, which provides *P. vivax* with a broader temperature tolerance, enables transmission even in temperate climates not tolerated by *P. falciparum*. Relapses and broader temperature tolerance could be expected to increase gene flow between different geographic regions. However, the distinctive phenotypic features found in parasite populations from different countries suggest barriers to gene flow, or alternatively, strong selection maintaining differentiation at loci underlying these traits.

The ability of *P. vivax* malaria to re-emerge in regions where malaria eradication or control efforts in the past have been successful is already evident in countries such as Uzbekistan, Azerbaijan, the Republic of Korea and northern Afghanistan. Because of increased international travel and migration, the risk of re-introduction of *P. vivax* in areas where species of *Anopheles* with greater vectorial capacity are present is high. The analysis of the parasites genetic make-up would be useful in identifying the geographic origins of these parasites and facilitate meaningful control and preventive measures to be implemented.

In this study, we analyzed the genetic diversity and population structure of 425 *P. vivax* patient isolates collected in Sri Lanka (2003–2008), Myanmar (2007), and Ethiopia (2006–2008) using 12 highly polymorphic trinucleotide and tetranucleotide microsatellite markers, which have been validated and tested. We considered the geographic and temporal influences on the genetic diversity and population structure of these *P. vivax* patient isolates. We further evaluated the use of these microsatellite haplotypes as a model for predicting the ancestry of *P. vivax* isolates using the Bayesian algorithm software STRUCTURE.

MATERIALS AND METHODS

Field isolates. *Plasmodium vivax* microsatellite typing was conducted on 425 field isolates collected in Sri Lanka (140), Myanmar (167), and Ethiopia (118). A field isolate is defined in this study as a sample of parasites derived from a
single-infected patient collected on a single occasion. Most patients had fever and were diagnosed as having \textit{P. vivax} malaria by microscopic examination of Giemsa-stained thick and thin blood smears. Transmission levels of malaria differed in the three countries, with the incidence rates in 2006 ranging between 50 and 200, 5 and 49, and 0 and 4 per thousand population in Ethiopia, Myanmar, and Sri Lanka respectively.\textsuperscript{16} Malaria transmission has decreased steadily in Sri Lanka since 2001 from 55,922 cases to 189 cases of \textit{P. vivax} malaria in 2007 (Annual Report Anti-Malaria Campaign, Ministry of Health, Sri Lanka). Malaria is endemic or hypoendemic in Myanmar. Although morbidity and mortality rates for malaria have decreased gradually in Myanmar, malaria still contributes to a large proportion of deaths in the southeast Asia region.\textsuperscript{18} Malaria transmission in Ethiopia is unstable and seasonal with occasional devastating epidemics.\textsuperscript{20}

The 140 field isolates of \textit{P. vivax} from Sri Lanka were collected from patients attending medical health clinics in Trincomalee, Batticaloa, Ampara (Eastern Province), Anuradhapura, Polonnaruwa (Northcentral Province), Kurunegala (Northwestern Province), Vavuniya (Northern Province), and Colombo. Venous blood samples were collected from these patients in 2003 (SL47, \( n = 01 \)), 2004 (SLA1–SLH4, \( n = 38 \)), 2005 (SL45, SLA7–SLP2E1, \( n = 61 \)), 2006 (SLP2A2–SLP2H3, \( n = 18 \)), 2007 (SL24–SL44, \( n = 21 \)), and 2008 (SL46, \( n = 01 \)). The 21 isolates collected during 2007 were from an outbreak of malaria that occurred in Trincomalee.

All 167 field isolates of \textit{P. vivax} from Myanmar were collected between July and November 2007 from patients attending medical health clinics at Kayin State, which is situated along the Myanmar–Thailand border (MYY3–MYY229, \( n = 96 \)), Kashin State (MYKC19–MYKC1478, \( n = 56 \)), and Rakhine State, along the Myanmar-Bangladesh border (MYY148–MYY1264, \( n = 15 \)). Capillary blood from a fingerprick was collected from each patient on a filter paper.

The 118 \textit{P. vivax} field isolates from Assendabo, Ethiopia, were collected from fever patients attending health clinics during 2006 (E1–E61, \( n = 59 \)), 2007 (E84V–E2400V, \( n = 52 \)), and 2008 (E434–E488V, \( n = 07 \)). Capillary blood from a fingerprick was collected from each patient on a filter paper.

Either a venous or capillary blood sample was collected from each infected patient after informed consent was obtained. Filter paper samples from Sri Lanka were de-linked from patient identifiers and provided as discarded samples from the clinic. DNA templates for polymerase chain reaction (PCR) amplification were isolated from venous blood by using the Nucleon genomic DNA extraction kit (Tepnel Life Sciences, Manchester, United Kingdom) and from filter papers by using either the fast methanol-based protocol\textsuperscript{21} or the QIAamp DNA purification Mini kit (Qiagen, Valencia, CA) according to the manufacturer’s instructions. Extracted DNA was dissolved in sterile distilled water and stored at \(-20\)°C. Because of limited availability of genomic DNA in these samples, the extracted DNA was further subjected to whole genome amplification (WGA) using a REPLI-g Mini-kit (Qiagen). The study protocol was reviewed and approved by the Human Subjects Committee of the Harvard School of Public Health (#P10299-111/0209GENE) and the Ethical Review Committee of the Faculty of Medicine, University of Colombo (EC/08/092).

**Microsatellite typing of \textit{P. vivax}**. Twelve validated highly polymorphic microsatellite markers\textsuperscript{22} (MS1, MS2, MS3, MS4, MS5, MS7, MS8, MS10, MS12, MS15, MS16, and MS20) were used for typing all isolates. Apart from MS2, which has a tetrancleotide repeat array, all other markers consist of trinucleotide tandem repeat motifs. These 12 markers are situated on different chromosomes (Table 1).\textsuperscript{23} The 12 markers were PCR-amplified using oligonucleotide primers with the forward primer labeled with fluorescent dyes (6-FAM, VIC, NED, and PET) as shown in Table 1. The PCRs with all primer pairs (6.4 pmol working solution of each) were performed on a PTC-200 thermo cycler (MJ Research, Waltham, MA) with 1.5 µL of genomic DNA template, 1 unit of Platinum Taq polymerase, 0.7 µL of 50 mM MgCl\textsubscript{2}, 2.0 µL of 10× reaction buffer, and 1 mM of each dNTP (all supplied by Invitrogen, Carlsbad, CA) in a final reaction volume of 20 µL. Cycling parameters were the same for all primer pairs: 1 cycle at 94°C for 2 minutes; 40 cycles at 94°C for 30 seconds, 58°C for 40 seconds, and 72°C for 30 seconds; and a final cycle at 72°C for 5 minutes. Length variation of labeled PCR products was measured on an ABI PRISM 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA) by using ABI GS500LIZ internal size standards and the GENESCAN, GENOTYPER, and GENEMAPPER software (Applied Biosystems).

**Data analysis.** The single or predominant allele at each locus was considered for computing allele frequencies.\textsuperscript{23} All 12 markers are single-copy loci. Because blood-stage malaria parasites are haploid, the presence of one or more additional alleles at a particular locus was interpreted as a co-infection with two or more genetically distinct clones (multiple-clone infections) in the same isolate.\textsuperscript{5,24} An additional allele was scored if the peak was at least one-third the height of the predominant allele (highest peak) as shown by the electropherogram traces, a method that has been validated.\textsuperscript{50} Only the predominant allele was used to define haplotypes in multiple-clone infections.\textsuperscript{5,6,24}

The genetic diversity of the parasite populations from Sri Lanka, Myanmar, and Ethiopia was determined by calculating the virtual heterozygosity (\(H_v\)) at each locus in each separate population. Virtual heterozygosity is the average probability that a pair of alleles randomly obtained from the population is different and was defined as \(H_v = [n(n - 1)]/[1 - \Sigma p_i^2]\), with \(n\) being the number of isolates analyzed and \(p_i\) the frequency of the \(i\)th allele in the population. Virtual heterozygosity ranges between 0 and 1, with values close to 1 reflecting high genetic diversity levels in a population. The eBURST version 3 software\textsuperscript{52,54} was used to search for nearly identical multilocus haplotypes (those differing by a single locus).\textsuperscript{27}

A standardized index of association (\(F_{is}\)) was used to test for evidence of overall multilocus linkage disequilibrium in each parasite population from Sri Lanka, Myanmar, and Ethiopia. This test compares the variance (\(V_{hi}\)) of the number of alleles shared between all pairs of haplotypes observed in the population (\(D\)) with the variance expected under random association of alleles (\(V_{hi}\)) and is defined as \(F_{is} = (V_{hi} - 1)(r - 1)\), with \(r\) being the number of loci analyzed.\textsuperscript{29} \(V_{hi}\) is derived from 10,000 simulated data sets in which alleles were randomly reshuffled among haplotypes. Significant linkage disequilibrium is detected if \(V_{hi}\) is greater than 95% of the values derived from the reshuffled data sets. Data was analyzed using LIAN software version 3.5.\textsuperscript{29,30}

The Bayesian clustering model to assign isolates to \(K\) populations according to allele frequencies at each locus was applied using STRUCTURE 2.2 software\textsuperscript{31} to test whether microsatellite haplotypes clustered according to the geographic
### Table 1

Characterization of the 12 polymorphic *Plasmodium vivax* microsatellite loci (n = 425)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>Core repeat sequence in the Salvador-1 strain</th>
<th>Primers (5'-3')</th>
<th>Size range, basepairs</th>
<th>No. alleles</th>
<th>$H_{E}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS1</td>
<td>3 (CM000444)</td>
<td>(GAA)$_{11}$</td>
<td>F:6-FAMTCACACTGTGTTGGAAGGGCAAT R:ctgtcttTGCTGCTGTTTTTGTGCTG</td>
<td>201–249</td>
<td>16 17 13</td>
<td>0.8997 0.9022 0.7692</td>
</tr>
<tr>
<td>MS2</td>
<td>6 (CM000447)</td>
<td>(TAAA)$_2$TATA(TAAA)$_6$TATA (TAAA)$_2$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>182–286</td>
<td>16 21 13</td>
<td>0.8514 0.8929 0.8014</td>
</tr>
<tr>
<td>MS3</td>
<td>4 (CM000445)</td>
<td>(GAA)$_{11}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>152–200</td>
<td>15 16 14</td>
<td>0.8835 0.8445 0.8617</td>
</tr>
<tr>
<td>MS4</td>
<td>6 (CM000447)</td>
<td>(AGT)$_{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>171–228</td>
<td>11 18 14</td>
<td>0.7848 0.8116 0.7850</td>
</tr>
<tr>
<td>MS5</td>
<td>5 (CM000446)</td>
<td>(CAG)$<em>{10}$(CAA)$</em>{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>139–232</td>
<td>17 16 6</td>
<td>0.8068 0.7933 0.867</td>
</tr>
<tr>
<td>MS7</td>
<td>12 (CM000453)</td>
<td>(GAA)$_{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>124–178</td>
<td>17 17 9</td>
<td>0.8526 0.9004 0.6523</td>
</tr>
<tr>
<td>MS8</td>
<td>12 (CM000453)</td>
<td>(CAG)$<em>{10}$(CAA)$</em>{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>165–345</td>
<td>42 44 37</td>
<td>0.9701 0.9666 0.9534</td>
</tr>
<tr>
<td>MS10</td>
<td>13 (CM000454)</td>
<td>GAA(GGA)$_3$AGA(GGA)$_3$AGA(GGA)$_3$AGA(GGA)$_3$AGA(GGA)$_3$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>174–249</td>
<td>19 23 18</td>
<td>0.8973 0.9095 0.9035</td>
</tr>
<tr>
<td>MS12</td>
<td>5 (CM000446)</td>
<td>(TTC)$<em>{10}$(TGC)$</em>{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>191–263</td>
<td>20 24 19</td>
<td>0.9305 0.9397 0.8983</td>
</tr>
<tr>
<td>MS15</td>
<td>5 (CM000446)</td>
<td>(TCT)$_{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>236–263</td>
<td>10 9 8</td>
<td>0.8502 0.8570 0.7536</td>
</tr>
<tr>
<td>MS16</td>
<td>9 (CM000450)</td>
<td>(ACA)$_3$GCA(ACA)$_3$GCA(ACA)$_3$GCA(ACA)$_3$GCA(ACA)$_3$GCA(ACA)$_3$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>132–330</td>
<td>21 10 3</td>
<td>0.6893 0.3682 0.1995</td>
</tr>
<tr>
<td>MS20</td>
<td>10 (CM000451)</td>
<td>(GAA)$<em>{10}$GAG(GAA)$</em>{10}$</td>
<td>F:VICGAGCTAGCCAAGAGTTACA R:ctgtcttTGAGAGAGAGAGAGAGCTA</td>
<td>150–246</td>
<td>31 30 20</td>
<td>0.9164 0.9540 0.7843</td>
</tr>
</tbody>
</table>
origins of the isolates. The program was run 10 times at each of 8 different K values (K1–6) with a burn-in period of 50,000 iterations followed by 107 iterations. The admixture model was used in all analyses. The number of populations was inferred by plotting the log probability of the data [Ln P(D)] for each K value. The strongest statistical support was associated with K = 2 (P > 0.999). For predicting the power of ancestry, two-thirds of isolates (n = 283) were selected randomly as a training model. These isolates with known geographic origin were used by STRUCTURE to distinguish between Asian and African samples. The remaining isolates (N = 142) were analyzed in a blinded manner. The program was run using different migration rates (MIGRPRIOR = 0.001, 0.01, 0.05, and 0.1); best results were obtained with MIGRPRIOR = 0.1. Isolates with predominant ancestry (>70%) were considered as members of that particular population.

RESULTS

**Microsatellite diversity in natural P. vivax isolates from Sri Lanka.** All parasite isolates from Sri Lanka were highly polymorphic and showed unique haplotypes, with the number of alleles per locus ranging from 10 to 42 (mean = 19.6; Table 1 and Supplemental Tables 1 and 2, available at www.ajtmh.org). No single haplotype was shared between isolates. Adjacent haplotypes (differing at a single locus) were seen between two pairs of isolates from Sri Lanka (SL32-SL40 and SL28-34), which were from one location in Sri Lanka (Trincomalee) collected during an outbreak in 2007. The average genetic distance (1 − proportion of microsatellite alleles shared by pairs of haplotypes) of isolates from Sri Lanka collected during the 2007 outbreak in Trincomalee (outbreak isolates = 0.7845) was significantly lower than that calculated for pairwise comparisons involving outbreak isolates and the rest of the isolates collected from Sri Lanka (non-outbreak isolates = 0.854; P < 0.0001, by two-sample randomization test performed with Poptools 2.5 software).32 No identical haplotypes occurred in either outbreak or non-outbreak isolates. 3.4% (N = 57) alleles were private (present only in the Sri Lankan population) and not shared between populations. 55.0% (N = 77) of isolates were mixed-clone infections.

The virtual heterozygosity (H_{EV}) values for the entire parasite population from Sri Lanka ranged from 0.6893 to 0.9701 (mean ± SE = 0.8610 ± 0.0216; Table 1). High diversity values were present even when considered according to the year of collection of the isolates (Table 2).

Significant multilocus linkage disequilibrium was found in the isolates (n = 140) from Sri Lanka (P_{A} = 0.0236, P < 0.001; Table 2). P_{A} was also calculated separately for the isolates considering their time of collection, and significant linkage disequilibrium was seen for subpopulations from 2004, 2005, and 2007: P_{A} = 0.0120, 0.0275, and 0.2709, respectively; P < 0.05. The association was extremely high in the 2007 subpopulation (mean ± SE = 0.8450 ± 0.0236 < 0.001 (Table 2). Significant H_{EV} was encountered in collections from Kayin State (2007, n = 96; P_{A} = 0.0130, P < 0.001) and Kachin State (2007, n = 56, P_{A} = 0.0313, P < 0.001), but not from Rakhine State (2007, n = 15, P_{A} = 0.0164, P = 0.139). Significant multilocus linkage disequilibrium was observed in collections from Kayin State (2007, n = 96; P_{A} = 0.0130, P < 0.001) and Kachin State (2007, n = 56, P_{A} = 0.0313, P < 0.001), but not from Rakhine State (2007, n = 15, P_{A} = 0.0164, P = 0.139).

**Microsatellite diversity in natural P. vivax isolates from Myanmar.** The parasite isolates from Myanmar were also highly polymorphic, with the number of alleles per locus ranging from 9 to 44 (mean = 20.4; Table 1 and Supplemental Tables 1 and 2). All isolates showed unique haplotypes with 5.6% (n = 112) of the alleles being private to Myanmar. A total of 67.1% (n = 112) were mixed-clone infections. The virtual heterozygosity (H_{EV}) values for the isolate population from Myanmar (n = 167) ranged from 0.3682 to 0.9666 (mean ± SE = 0.8450 ± 0.0460; Table 1) and showed significant multilocus linkage disequilibrium with P_{A} = 0.0149, P < 0.001 (Table 2). Significant H_{EV} was encountered in collections from Kayin State (2007, n = 96, P_{A} = 0.0130, P < 0.001) and Kachin State (2007, n = 56, P_{A} = 0.0313, P < 0.001), but not from Rakhine State (2007, n = 15, P_{A} = 0.0164, P = 0.139).

**Microsatellite diversity in natural P. vivax isolates from Ethiopia.** Parasite populations from Ethiopia were also highly polymorphic, having between 3 and 37 alleles per locus (mean = 14.5; Table 1 and Supplemental Tables 1 and 2). All isolates had unique haplotypes with only 0.07% (n = 1) of the alleles being private to Ethiopia. A total of 73.70% (n = 87) were mixed-clone infections. The virtual heterozygosity (H_{EV}) values for the isolate population from Ethiopia ranged from 0.1995 to 0.9534 (mean ± SE = 0.7517 ± 0.0568; Table 1). Significant levels of genetic diversity were found for isolates collected in 2006 (n = 59; H_{EV} = 0.7432 ± 0.0551), 2007 (n = 52; H_{EV} = 0.7314 ± 0.0695), and 2008 (n = 7; H_{EV} = 0.6190 ± 0.0839). Significant multilocus linkage disequilibrium was also seen among the isolates from Ethiopia (n = 118), with P_{A} = 0.0137,

### Table 2

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Location</th>
<th>No.</th>
<th>H_{EV}</th>
<th>P_{A}</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sri Lanka</td>
<td>2003</td>
<td>Colombo</td>
<td>1</td>
<td>0.8491</td>
<td>0.0113</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>Anuradhapura</td>
<td>20</td>
<td>0.8611</td>
<td>0.0143</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Polonnaruwa</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mannar</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ampara</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>38</td>
<td>0.8602</td>
<td>0.0120</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>Trincomalee</td>
<td>20</td>
<td>0.8215</td>
<td>0.0127</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Batticaloa</td>
<td>17</td>
<td>0.7953</td>
<td>0.0669</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ampara</td>
<td>4</td>
<td>0.7778</td>
<td>0.2104</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Polonnaruwa</td>
<td>4</td>
<td>0.8750</td>
<td>0.0350</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vavuniya</td>
<td>7</td>
<td>0.8492</td>
<td>0.0548</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurunegala</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mannar</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monaragala</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Colombo</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>61</td>
<td>0.8542</td>
<td>0.0275</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>Kurunegala</td>
<td>17</td>
<td>0.6906</td>
<td>0.0116</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monaragala</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>18</td>
<td>0.6912</td>
<td>0.0083</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>Trincomalee</td>
<td>21</td>
<td>0.7845</td>
<td>0.2709</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Colombo</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>140</td>
<td>0.8610</td>
<td>0.0236</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Myanmar</td>
<td>2007</td>
<td>Kayin State</td>
<td>96</td>
<td>0.8536</td>
<td>0.0130</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kachin State</td>
<td>56</td>
<td>0.8180</td>
<td>0.0313</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rakhine State</td>
<td>15</td>
<td>0.8317</td>
<td>0.0164</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>167</td>
<td>0.8450</td>
<td>0.0149</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2006</td>
<td>Assendabo</td>
<td>59</td>
<td>0.7432</td>
<td>0.0199</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>Assendabo</td>
<td>52</td>
<td>0.7314</td>
<td>0.0180</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>Assendabo</td>
<td>7</td>
<td>0.6190</td>
<td>0.0097</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>118</td>
<td>0.7517</td>
<td>0.0137</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
$P < 0.001$ (Table 2) remaining significant even when considered according to the time of collection of the isolates: $F_1^A = 0.0199$, $P < 0.001$ for 2006, $F_2^A = 0.0180$, $P = 0.006$ for 2007, and $F_3^A = 0.0697$, $P = 0.044$ for 2008 (n = 7).

**Geographic structure of *P. vivax***. Analysis of 425 isolates from Sri Lanka, Myanmar (Asia) and Ethiopia (Africa) identified haplotypes that were all unique. Private alleles for Asia and Africa were 13.4% and 0.07% respectively. Multiple-clone infections were 64.9% (n = 276).

The clustering patterns obtained with $K = 2$ shown in Figure 1 showed clustering of *P. vivax* isolates into Asian and African origins with much admixture in both regions. Adding

---

on a third population ($K = 3, P < 0.001$) did not help to further differentiate between populations from Sri Lanka and Myanmar (Figure 2). However, the first 21 (Figure 2) isolates from Sri Lanka collected from one site (Trincomalee) during a *P. vivax* malaria outbreak in 2007 and isolates from Assendabo, Ethiopia (308–425; Figure 2) collected during seasonal outbreaks in 2006, 2007, and 2008 are shown clearly in different clusters. The population structure of *P. vivax* according to ancestry at $K = 2$ is shown in Figure 3. The red cluster shows 62.7% ($n = 74$) of isolates with predominant (> 70%) ancestry from Africa, and the green cluster shows only a 48.8% ($n = 150$) membership fraction from Asia.

**Predictive power of ancestry for *P. vivax* isolates.** We further assessed the predictive power of ancestry for *P. vivax* isolates using two-thirds of the isolates ($n = 283$) as a model. Of the 142 isolates that were tested in this model, 47 were from Sri Lanka, 56 from Myanmar, and 39 from Ethiopia. The percentages of test isolates that were correctly identified to have a predominant ancestry (> 70%) from either Asian or African origin were 72.3% ($n = 34$) for Sri Lanka (Asian), 78.6% ($n = 44$) for Myanmar (Asian), and 84.6% ($n = 33$) for Ethiopia (African), giving an overall predictive power of 78.2% ($n = 111$).

**DISCUSSION**

Microsatellite polymorphisms are derived mainly from variability in length rather than in the primary sequence. Genetic variation at many microsatellite loci is characterized by high heterozygosity and the presence of multiple alleles, which is well demonstrated among the 425 isolates from Sri Lanka, Myanmar, and Ethiopia. Most of this microsatellite diversity is probably caused by strand slippage during mitotic replication of the parasites, which causes transient dissociation of the replicating DNA strands followed by misaligned reassociation, resulting in either the addition or deletion of repeat units.

Genetic diversity measured by virtual heterozygosity was high in all three populations, which is expected with microsatellite data. This high diversity extended to all individual loci except for 1 of the 12 analyzed. Interestingly, the MS16 marker showed diversity levels varying from $H_s = 0.1995$ in the parasite population from Ethiopia to $H_s = 0.6908$ in the population from Sri Lanka. This microsatellite marker is located on chromosome 9 of the *P. vivax* genome in a gene coding for a hypothetical protein (PVX_092630, Seq. ID CM000450-PlasmoDB 5.5). Whether the hypothetical protein coded for by this gene is under selection requires further investigation because fluctuating selection pressures within populations may contribute to reduced variability in regions of reduced recombination.

Little is known of the global genetic diversity and population structure of *P. vivax*. This lack of information is due mostly to the lack of appropriate genetic markers for the *P. vivax* genome, which has hampered in-depth analysis of the population structure and evolutionary history of the parasite and prevented efforts to map determinants contributing to important parasite phenotypes. In this study, with the use of 12 highly polymorphic microsatellite markers distributed genome wide, we were able to identify the population structure of *P. vivax* according to their geographic location (Asian versus African, $K = 2$). Although the addition of a third population ($K = 3$) did not improve differentiation between isolates from Sri Lanka and Myanmar (Figure 2), strong clustering of the outbreak isolates from Sri Lanka (first portion of cluster 1 in Figure 2 shown predominantly in blue) and Ethiopia (cluster 3, predominantly in red) was observed, which is consistent with epidemic transmission of clones. The significantly lower average genetic distance between the outbreak and non-outbreak isolates from Sri Lanka (0.7845 versus 0.8584; $P < 0.0001$) also substantiates this result. Unfortunately, comparisons between outbreak and non-outbreak isolates from Ethiopia could not be made because non-outbreak isolates were not available.

Figure 3 shows the *P. vivax* populations sorted according to Asian or African ancestry, with much admixture in both fractions, especially the Asian fraction. Almost 63% of isolates from Ethiopia had a predominant ancestry (> 70%) from Africa, and only 45% from Myanmar and 54% from Sri Lanka were shown to be from Asia. Many alleles were shown to be private for Asia (13.4%; Supplemental Table 2) than Africa.

---

**Figure 2.** Population structure of *Plasmodium vivax* plotted in a single line according to geographic location inferred from microsatellite typing of 425 isolates at $K = 3$. Isolates 1–140 were collected in Sri Lanka; 141–307 in Myanmar; and 308–425 in Ethiopia. This figure appears in color at www.ajtmh.org.

**Figure 3.** Population structure of *Plasmodium vivax* according to ancestry inferred from microsatellite typing of 425 isolates at $K = 2$ (single and multiple line plots). Predominant ancestry (> 70%) for Asian and African isolates in each cluster was as follows: red = 62.7% ($n = 74$) African and 12.7% ($n = 39$) Asian; green = 48.8% ($n = 150$) Asian and 5.9% ($n = 7$) African. This figure appears in color at www.ajtmh.org.
(0.07%), which corroborates the historical origins of \textit{P. vivax} in Asia.\textsuperscript{2,37,38}

Microsatellites are known to be a valuable resource for the study of genetic diversity because they are generally hypervariable, codominant, locus-specific, and not directly subject to host immunity.\textsuperscript{19} However, they have high mutation rates and significant mutation-rate heterogeneity among loci, which means that it might be difficult to translate estimates of genetic distance into absolute timescales.\textsuperscript{35} More importantly, alleles generated by microsatellite data represent similarities in electrophoretic mobility rather than by descent. Therefore, use of microsatellites in population genetics and evolution has its limitations. However, the population structure generated using these microsatellites show promise for microsatellite data being a good tool to identify parasite differentiation occurring over a short period, such as for mapping outbreaks.

The predictive power of ancestry using two-thirds of isolates as a model gave good results for isolates from all three populations. Predominant ancestry (> 70%) for African or Asian origins were 85% African and 3% Asian for isolates from Ethiopia, 79% Asian and 5% African isolates from Myanmar, and 72% Asian and 11% African for isolates from Sri Lanka. Thus, microsatellite data as a model for predicting the origin of \textit{P. vivax} parasites holds promise and would provide useful information regarding genetic structure of parasites, in context to the resurgence of \textit{P. vivax} malaria in many regions worldwide.\textsuperscript{16} This approach would be a useful complement to conventional epidemiologic methods with important implications for control and prevention of this parasite.

High levels of genetic diversity in \textit{P. vivax} have been demonstrated in Papua New Guinea, Thailand, India, Colombia, and Laos by use of dinucleotide microsatellites.\textsuperscript{5,34,40} High diversity levels have also been observed in India\textsuperscript{18} and Myanmar\textsuperscript{19} with the use of genetic markers. Previous studies with \textit{P. vivax} parasites from Sri Lanka,\textsuperscript{17} Brazil,\textsuperscript{4} and Myanmar\textsuperscript{19} demonstrated high genetic diversity and multiple clone infections despite low levels of transmission in these areas. The co-occurrence of significant linkage disequilibrium in the presence of high diversity and multiple clone infections was also demonstrated in parasites from both Sri Lanka\textsuperscript{17} and Brazil.\textsuperscript{4} Findings of the present study involving three geographic sites add further strength to the above observations. Linkage disequilibrium in parasite populations may arise as an artifact of the admixture of two or more subpopulations with different allele frequencies. Therefore, isolates from each population were analyzed individually according to their times and sites of collection. Apart from some sites in Sri Lanka (Trincomalee, 2005, n = 20), Anuradhapura (2004, n = 20), Polonnaruwa (2004 and 2005, n = 16 and 4, respectively), Vavuniya (2005, n = 7), and Kurunegala (2006, n = 16) and a single site in Myanmar (Rakhine State, 2007, n = 12), significant linkage disequilibrium was consistently present among the remaining sites (Table 2).

The predominance of inbreeding in \textit{P. vivax} with high proportions of multiple-clone infections and genetic diversity is unusual. High rates of selling may occur in the presence of multiple clone infections if only a small proportion of the co-infecting genetically distinct clones produce viable gametocytes that are taken up by mosquitoes or when strand slippage events occur so frequently during mitotic replication of parasites that new haplotypes would be generated without affecting the overall patterns of association between loci.\textsuperscript{4,6}

In conclusion, microsatellite-based analysis of \textit{P. vivax} parasites from Sri Lanka, Myanmar, and Ethiopia showed extensive genetic diversity co-occurring with significant multilocus linkage disequilibrium. These parasite populations were seen to cluster according to their geographic locations and ancestry. The predictive power of ancestry with this microsatellite data as a model is promising and could be useful in identifying the origin of \textit{P. vivax} malaria cases and enabling meaningful control and preventive measures to be implemented. Although microsatellite data may have some limitations in its use in population genetics and evolution of malaria parasites, it appears to serve as an excellent tool for mapping short-term outbreaks of malaria.

Received September 29, 2009. Accepted for publication October 6, 2009.


Acknowledgments: We thank the director of the Anti-Malaria Campaign, Sri Lanka, for authorizing collection of samples; G.M.G. Kapilandananda (Department of Parasitology, Faculty of Medicine, University of Colombo) and Nira Mahesh (Department of Immunology and Infectious Diseases, School of Public Health, Harvard University) for technical assistance; Yihe new Tesfaye (Boston University/ Harvard School of Public Health project), Abayneh Dega, Enaku Keno Doti, Teklu Aklilu (Assendabo Health Center), and Myat Hlut Nyunt (Myanmar) for assistance in providing samples.

Financial support: This study was supported by National Institutes of Health grant 5R03 TW007966-02.

Authors’ addresses: Sharmini Gunawardena, Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA and Department of Parasitology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka. Nadira D. Karunaweera, Department of Parasitology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka. Marcelo U. Ferreira, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil. Myatt Phone-Kyaw, Parasitology Research Division, Department of Medical Research (Lower Myanmar), Myanmar. Richard J. Pollack and Dyann F. Wirth, Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA. Michael Alifrangis, Fleming Konradsen, and Mette L. Schousboe, Centre for Medical Parasitology, Department of International Health, Immunology and Microbiology, University of Copenhagen, and Department of Infectious Diseases and Department of Clinical Microbiology, Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark. Rupika S. Rajakaruna, Department of Zoology, University of Peradeniya, Peradeniya, Sri Lanka. Priyanie H. Amerasinghe, International Water Management Institute, Delhi, India. Gwirie N. L. Galappaththy and Rabindra R. Abyasinghe, Anti-Malaria Campaign, Ministry of Health, Colombo, Sri Lanka. Daniel L. Hartl, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA.

REFERENCES
