Plasmodium falciparum Genetic Diversity in Western Kenya Highlands

Daibin Zhong,* Yaw Afane, Andrew Githeko, Zhaoqing Yang, Liwang Cui, David M. Menge, Emmanuel A. Temu, and Guiyun Yan

Program in Public Health, College of Health Sciences, University of California, Irvine, California; Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya; Department of Parasitology, Kunming Medical College, Kunming, Yunnan, People’s Republic of China; Department of Entomology, Pennsylvania State University, University Park, Pennsylvania; Center for Infectious Diseases and Microbiology Translational Research, University of Minnesota, Minneapolis, Minnesota; The Institute of Tropical Medicine, Nagasaki University, Nagasaki-shi, Japan

Abstract. The present study examined the genetic diversity and population structure of Plasmodium falciparum in western Kenya by analyzing the polymorphism of 12 microsatellite loci and two antigen loci. Malaria in highland areas is unstable and epidemic whereas malaria in lowland areas is endemic. Transmission intensity and malaria prevalence are substantially lower in the highlands than in the lowlands. Despite that the highland parasite populations exhibited reduced number of alleles, lower expected heterozygosity, and infection complexity in comparison to the surrounding lowland population, genetic diversity of the highland populations remained high in comparison to parasites from other meso-endemic regions. More than 70% of infections from western Kenya highland study sites were mixed genotype infections. Small but statistically significant genetic differentiation between highland and lowland populations was detected. These findings are discussed in the context of human travel and local transmission in the study area.

INTRODUCTION

Epidemic malaria is a major public health problem in the African highlands (highlands are generally referred to areas with elevation more than 1,500 meters above sea level). For instance, since the 1980s a series of Plasmodium falciparum epidemic malaria has occurred in the western Kenyan highlands, a region that was previously free of malaria or exhibited low incidences of the disease.1–4 The epidemic malaria in the western Kenyan highlands is characterized by expanded geographic areas,5 increased frequencies,6 and increased case-fatality rates.7 Previous studies have examined the risk factors associated with clinical malaria in the highlands.8–10 Population dynamics of mosquito vectors,11 effects of topography and land use on the spatial distribution of vectors and malaria infections,12–14 and the potential role of climate factors5,15,16 and drug resistance on malaria epidemics.10,17

Genetic diversity and population structure of malaria parasites have profound impacts on clonal diversity,18 competitive or positive interactions among the clones,19,20 dynamics of drug resistance,21 persistence of asexual infection and gametocyte production,22 infectivity in mosquito vectors,23,24 and malaria vaccine development.25,26 Global-scale studies on P. falciparum genetic diversity using specimens with varying transmission intensities from sites in Africa, South America, and Asia found a general pattern of low genetic diversity, strong linkage disequilibrium, and high genetic differentiation in low transmission areas, whereas in areas of high transmission intensity, parasite populations exhibited a panmictic pattern with high genetic diversity and weak linkage disequilibrium.18 Previous studies established that infection complexity within a host was dependent on exposure (transmission intensity) and age of the host (reflective of level of immunity), and negatively associated with the severity of the disease.27–32

A recent study found that allelic diversity of the merozoite surface protein 1 gene (msp-1) is affected not only by transmission intensity, but also by other factors such as the number of alleles prevalent and infection complexity.33 Moreover, studies of daily parasite population dynamics showed that such complex patterns of multiple infections would undergo rapid changes.34 To date, few studies have examined the genetic diversity and population structure of malaria parasites in highland areas where malaria transmission is unstable and occurrences of epidemic is common.

The objective of this study was to determine genetic diversity of P. falciparum in two highland sites (Kakamega and Kisii districts) in western Kenya. We then compared our results with those found in a surrounding, malaria-endemic lowland site (Kisumu district), also in western Kenya. The transmission intensity and malaria prevalence in the highlands was substantially lower than in the lowlands.11 If transmission intensity is an important determinant of the genetic diversity of malaria parasites, one would expect that epidemic malaria in the highlands would lead to few major genotypes in the population, thus low genetic diversity and low infection complexity. To the contrary, high genetic diversity and infection complexity is expected in the malaria-endemic lowlands because of frequent meiotic recombination events in mosquito vectors resulting from high infection rates. Conversely, travel to malaria-endemic areas was found to be significantly associated with increased malaria risk in highland residents.10 Introduction of new alleles by travelers from malaria-endemic areas to the highlands would increase population genetic diversity. We examined genetic diversity of block 2 of msp-1 and block 3 of msp-2 genes that code for major blood-stage surface antigens. In malaria-endemic areas in Africa, allelic types of block 2 of msp-1 were strongly associated with protection from P. falciparum malaria,35 demonstrating that these genes are under strong selection by human immune systems.36,37 Thus, interpretation of population structure and diversity results based on a small number of strongly selected markers such as msp-1 and msp-2 is problematic because the results are confounded by population history and natural selection.38 Therefore, we also included 12 microsatellite markers that are presumably selective neutral.38,39

MATERIALS AND METHODS

Parasite sampling. Blood samples were collected from malaria outpatients attending three clinics, Iguhu (1,500–1,580
For microsatellite loci, the analyses of gene on chromosome 11 and 10 ng of template DNA. An initial denaturation was estimated at 16.6 infectious bites per person per year in the Kakamega site, 0.4 in Kisii, and 31.1 in Kombewa. In 2005, *Plasmodium falciparum* prevalence in asymptomatic children (6–16 years of aged) ranged from 20% to 48% in Kakamega, less than 10% in Kisii, and consistently more than 50% in Kombewa (Afrane Y, unpublished data). *Plasmodium falciparum* constituted more than 98% of malaria infections in these sites. On the basis of elevation, the Kakamega and Kisii sites are classified as highland sites, and the Kombewa site as a lowland site. The catchment area of these village clinics is small, typically less than 10,000 residents within a 15-km² area. A blood sample of approximately 200 μL was obtained from symptomatic volunteers using the finger prick method and then preserved on Whatman (Maidstone, United Kingdom) filter papers. All samples were collected in August 2005. An exclusion criterion was unwillingness to participate in the study. Blood samples from all study participants were tested for *P. falciparum* infection by microscopy, and all positive cases were treated with artemether/lumefantrine (Coartem; Novartis, Basel, Switzerland). The mean age of donors was 16.8 years (range = 1–86 years) in Kakamega (n = 98), 16.7 years (range = 1–80 years) in Kisii (n = 90), and 15.4 years (range = 1–60 years) in Kombewa (n = 100). The study was reviewed and approved by the ethical review committee of the Kenya Medical Research Institute and the institutional review board of the University of California at Irvine.

**Parasite DNA extraction and microsatellite genotyping.** Genomic DNA of the malaria parasite was extracted from 50 microscopically-positive samples per site using the Saponin/ Chelex method. The nested ribosomal DNA–polymerase chain reaction (PCR) method described by Singh and the others was used to confirm *P. falciparum* species identification. Parasites were genotyped using 12 microsatellite markers (Polya, TA42, TA81, TA1, TA87, TA109, ARA2, 2490, TA40, PIPK2, Pfg377, and TA60) that have previously been used for population genetic studies. The single-clone parasite supplied by the Malaria Research and Reference Reagent Resource Center (Manassas, VA) (MR4) and distilled water were used as positive and negative controls, respectively. We followed the protocol of Anderson and others for PCR conditions, primer specificity and reproducibility, with small modifications for automated genotyping in a model 4300 automated DNA analyzer (Li-Cor, Lincoln, NE). For the apparatus to detect PCR products, one primer in every microsatellite marker must be fluorescently labeled (infrared dye, IRD800). To reduce the cost associated with synthesis of fluorescently labeled primers, we used the tailed primer method. In this method, the forward primer for each marker was synthesized with an additional 19-basepair M13 sequence (5'-CACGACGTTGTAAAACGAC-3') added to the 5' end of the oligonucleotide. A primer with the same 19-basepair sequence was directly labeled with the fluorescent dye (IRD800) and was used as the sole type of labeled primer for the detection of microsatellite alleles for all 12 markers. The tailed primer method dramatically reduced the synthesis cost of fluorescently labeled primers.

The PCR was carried out in a total volume of 25 μL for each sample and locus and contained 20 μL of 1.1× PCR Master Mix (Abgene, Rochester, NY), 0.25 μL of 10 μM each primer, 0.5 μL of M13 primer, an additional 1 μL of 25 mM MgCl₂, and 10 ng of template DNA. An initial denaturation period of for 5 minutes at 95°C preceded the 45 amplification cycles (annealing for 30 seconds at 45°C, extension for 45 seconds at 65°C, and denaturation for 30 seconds at 94°C), followed by a final extension of 7 minutes. A Gene ImagIR 4.33 software (Li-Cor) program was used to quantify allele size on the basis of the height pattern of signal peaks. Multiple alleles were scored at a given locus if minor peaks were more than one-third the height of the predominant peak. An infection was considered to contain multiple clones if one or more loci showed more than one allele. Infection complexity provides a surrogate indicator of the level of transmission within populations, as well the opportunity for genetic recombination between different malaria clones.

**Merozoite surface protein genotyping.** We genotyped three *P. falciparum* populations using polymorphic block 2 of *msp-1* gene on chromosome 9 and block 3 of *msp-2* gene on chromosome 2. A nested PCR were performed individually using previously published primer sequences and conditions that distinguish three major allelic families (K1, MAD20, and RO33) for *msp-1* and two allelic families (FC27 and IC1) for *msp-2*. The PCR products were subjected to electrophoresis on a 2.0% agarose gel. Fragment length was determined using the Kodak Digital Science 1D Image Analysis Software (Eastman Kodak Company, New Haven, CT). Each *P. falciparum* infection was characterized on the basis of the allele size for each locus. Sizes of alleles were allocated into bins of 40-basepair size ranges.

**Statistical analysis.** For microsatellite loci, the analyses focused on population genetic diversity, infection complexity,
linkage disequilibrium, and genetic differentiation. Genetic
diversity was measured by the number of alleles per locus and
expected heterozygosity on the basis of allele frequency data
in each population. In the case of the presence of multiple
alleles at a locus within an infection, only the predominant
allele defined by the allele with the highest peak in phero-
grams was used for calculation of allele frequency. This
procedure is appropriate for estimation of population allele
frequency if the composition of PCR products is representa-
tive of the composition of templates. The unbiased expected
heterozygosity, $H_e$, was calculated as $H_e = \frac{n(n-1)}{2} - \Sigma$, where
$n$ is the number of infections sampled and $p_i$ is the
frequency of the $i$th allele. Observed infection complexity is
defined as the largest number of alleles at any $msp$ locus
detected in the sample; it is often referred to as minimum
number of clones. For example, an infection with two
$msp$-1 alleles and one $msp$-2 allele is scored as two clones.
This measure is conservative and it likely underestimates the
number of clones likely to be present. Therefore, we used the
maximum likelihood methods described by Hill and Babiker to estimate the mean number of clones, the support
limit and $msp$ allele frequencies, assuming a Poisson distribu-
tion of number of clones per individual. Because detection of
multiple alleles per locus depends critically on the polymor-
phism of the markers and sensitivity of the methods used, the
data provides relative assessment rather than absolute mea-
sures of infection complexity. The non-parametric Wilcoxon
test was used to determine the statistical differences in genetic
diversity among populations.

Linkage disequilibrium (LD), which characterizes statisti-
cal independence of alleles at pairs of loci, was determined for
all pairs of microsatellite loci, using the LIAN 3.0 software
developed for multilocus haplotype data. This software tests
the null hypothesis of no linkage by a Monte Carlo simulation
(10,000 permutations) on the variance of genetic distances
between isolates ($V_{p}$). The sample variance is then compared
with the variance expected under linkage equilibrium $V_{E}$. A
distribution of $V_{E}$ is generated by Monte Carlo simulations,
and its percentiles provide 95% confidence intervals. The output
file gives $V_{E}$ and $V_{p}$ values, as well as a standardized index of association ($I_{AS} = \frac{V_{p} - V_{E}}{V_{E} - 1} / (1 - \rho)$, where $r$ is
the number of loci, a measure of haplotype-wide linkage and the
95% confidence limits determined by Monte Carlo simula-
tions ($I_{MC}$). Because $msp$ markers are not selectively neu-
tral, population genetic structure was determined using
microsatellite markers alone. The analysis of molecular vari-
ance (AMOVA) was conducted using the Arlequin genetic
analysis software. The AMOVA partitions the molecular
variability (microsatellite allele size) into three categories: among
groups (highland and lowland), among populations within
groups, and among individuals within populations. A
Permutation test ($n = 1,000$) was used to test the null hy-
pothesis that the variance component is zero. Wright’s $F$-
statistics between pairs of populations is also computed using
the FSTAT computer software.

Structure 2.0 computer program was used to assign indi-
viduals from all populations to a predetermined number of
clusters (K) based on multilocus microsatellite data. For
each run, a burn-in period of 50,000 steps was followed by
1,000,000 iterations under the admixture model and the as-
sumption of correlated allele frequencies. Various values of $K$
(2 to 4) were tested. Individuals were assigned to the cluster
only if the proportion ancestry was $\geq 0.8$. The individual
was unassigned when the proportion ancestry from any single
cluster was < 0.8.

**RESULTS**

**Microsatellite polymorphism.** High levels of polymorphism
were shown among parasites because all 12 microsatellite
markers used were polymorphic for the three populations
studied. The number of microsatellite alleles per locus ranged
from 7 (TA42) to 16 (Polya). The lowland population showed
the highest level of diversity, and the highland samples
showed a significantly lower level of diversity (Table 1). For
example, an average of 8.6 alleles was detected in the lowland
Kombewa population (range = 3–14), whereas the two high-
land populations, Kakamega and Kisii, showed an average of
6.7 alleles (range = 4–9) and 5.8 alleles (range = 2–9), respec-
tively ($\chi^2 = 7.44, df = 2, P = 0.02$). The levels of unbi-
ased expected heterozygosity of the lowland Kombewa popu-
lation were marginally higher than those observed in the high-
land population ($\chi^2 = 4.86, degrees of freedom [df] = 2,
P = 0.09$). The estimated expected heterozygosity ($H_e$) was
0.75 (range = 0.55–0.89) in Kombewa, 0.69 (range = 0.53–
0.78) in Kakamega, and 0.66 (range = 0.38–0.81) in Kisii
(Table 1).

The overall infection complexity (minimum number of
clones in an infected detection by all 12 microsatellite
markers) was marginally higher in the lowland Kombewa popu-
lation than in the two highland populations ($\chi^2 = 5.54, df = 2,
P = 0.06$; Table 2). The proportion of infections with more
than one clone was 87.0%, 82.4%, and 80.9% in Kombewa,
Kakamega and Kisii, respectively ($\chi^2 = 1.44, df = 2, P = 0.49$).

**Genetic diversity of $msp$-1 and $msp$-2 genes.** The $P. falciparum$
parasite populations exhibited a high degree of diver-
sity on the basis of length polymorphism of the two antigen
loci. Using a conservative bin size range of 40 basepairs for
$msp$-1 and $msp$-2, a total of 13 alleles were identified for
$msp$-1 and 15 alleles for $msp$-2 for the three locations (Table 3).
At the $msp$-1 locus, a total of six alleles of the K1 family
(size range = 140–370 basepairs), five MAD20 alleles
(range = 120–320 basepairs) and two alleles of the RO33
family (range = 180–260 basepairs) were detected in the three

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>Lowland</th>
<th>Highland</th>
<th>Kisii</th>
<th>Allele size (basepairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polya</td>
<td>4</td>
<td>14</td>
<td>0.89</td>
<td>0.79</td>
<td>6.73</td>
</tr>
<tr>
<td>TA42</td>
<td>5</td>
<td>5.66</td>
<td>0.53</td>
<td>0.55</td>
<td>204–277</td>
</tr>
<tr>
<td>TA81</td>
<td>5</td>
<td>0.80</td>
<td>0.72</td>
<td>0.63</td>
<td>129–160</td>
</tr>
<tr>
<td>TA1</td>
<td>6</td>
<td>0.58</td>
<td>0.55</td>
<td>0.53</td>
<td>152–221</td>
</tr>
<tr>
<td>TA87</td>
<td>6</td>
<td>0.86</td>
<td>0.73</td>
<td>0.78</td>
<td>109–144</td>
</tr>
<tr>
<td>TA109</td>
<td>6</td>
<td>0.82</td>
<td>0.73</td>
<td>0.69</td>
<td>170–237</td>
</tr>
<tr>
<td>ARA2</td>
<td>11</td>
<td>0.81</td>
<td>0.72</td>
<td>0.82</td>
<td>79–108</td>
</tr>
<tr>
<td>2490</td>
<td>10</td>
<td>0.64</td>
<td>0.41</td>
<td>0.38</td>
<td>90–115</td>
</tr>
<tr>
<td>TA40</td>
<td>10</td>
<td>0.75</td>
<td>0.69</td>
<td>0.61</td>
<td>195–252</td>
</tr>
<tr>
<td>PIPK2</td>
<td>12</td>
<td>0.85</td>
<td>0.76</td>
<td>0.77</td>
<td>167–217</td>
</tr>
<tr>
<td>PIP377</td>
<td>12</td>
<td>0.55</td>
<td>0.59</td>
<td>0.67</td>
<td>110–130</td>
</tr>
<tr>
<td>TA60</td>
<td>13</td>
<td>0.84</td>
<td>0.78</td>
<td>0.81</td>
<td>188–237</td>
</tr>
</tbody>
</table>

### TABLE 1

Genetic diversity of *Plasmodium falciparum* from western Kenya at 12 microsatellite loci

* $Na = number of alleles for microsatellite loci, $He = expected heterozygosity.*
populations. Alleles of the RO33 family in the msp-1 gene were not found in the Kisii population. At the msp-2 locus, seven IC1 alleles (range = 440–760 basepairs) and eight FC27 alleles (range = 300–720 basepairs) were observed. Overall, the Kombewa population showed the highest number of alleles, and the Kisii population showed the lowest number of alleles (Table 3). The K1 and IC1 allele families were dominant alleles in western Kenya study sites. The minimum number of clones was 1.98, 1.92, and 1.58 in Kombewa, Kakamega, and Kisii, respectively (P > 0.05). The number of clones estimated by the maximum likelihood method was substantially higher than the observed number of clones (Table 3), but the trend on the difference between highland and lowland sites was similar to those detected by the minimum number of clones. The proportion of mixed infections based on msp-1 and msp-2 genes was 81.7%, 76.7%, and 73.3% in Kombewa, Kakamega, and Kisii, respectively (χ² = 1.21, df = 2, P = 0.55). This result indicated that multiple infections in western Kenya highlands were as frequent as in the lowland populations.

### Table 2
Mean infection complexity and percentage of mixed infections detected by microsatellite markers in three populations of *Plasmodium falciparum* from highland and lowland areas in western Kenya

<table>
<thead>
<tr>
<th>Locus</th>
<th>Kombewa</th>
<th>Kakamega</th>
<th>Highland</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean complexity ± SE</td>
<td>% Multi-infection</td>
<td>Mean complexity ± SE</td>
</tr>
<tr>
<td>Polya</td>
<td>2.38 ± 0.22</td>
<td>33.9</td>
<td>2.00 ± 0.24</td>
</tr>
<tr>
<td>TA42</td>
<td>1.14 ± 0.10</td>
<td>20.3</td>
<td>1.10 ± 0.11</td>
</tr>
<tr>
<td>TA81</td>
<td>3.33 ± 0.25</td>
<td>72.9</td>
<td>3.00 ± 0.25</td>
</tr>
<tr>
<td>TA1</td>
<td>2.43 ± 0.21</td>
<td>59.3</td>
<td>1.90 ± 0.19</td>
</tr>
<tr>
<td>TA87</td>
<td>2.65 ± 0.23</td>
<td>49.2</td>
<td>2.30 ± 0.26</td>
</tr>
<tr>
<td>TA109</td>
<td>1.83 ± 0.16</td>
<td>28.8</td>
<td>1.56 ± 0.15</td>
</tr>
<tr>
<td>ARA2</td>
<td>2.82 ± 0.24</td>
<td>55.9</td>
<td>2.10 ± 0.20</td>
</tr>
<tr>
<td>2490</td>
<td>1.92 ± 0.16</td>
<td>66.1</td>
<td>1.70 ± 0.17</td>
</tr>
<tr>
<td>TA40</td>
<td>1.71 ± 0.13</td>
<td>44.1</td>
<td>1.20 ± 0.12</td>
</tr>
<tr>
<td>PIPK2</td>
<td>2.06 ± 0.16</td>
<td>50.9</td>
<td>1.29 ± 0.16</td>
</tr>
<tr>
<td>Pfg377</td>
<td>2.43 ± 0.16</td>
<td>71.2</td>
<td>2.20 ± 0.18</td>
</tr>
<tr>
<td>TA60</td>
<td>2.31 ± 0.20</td>
<td>44.1</td>
<td>2.10 ± 0.24</td>
</tr>
<tr>
<td>Overall††</td>
<td>4.76 ± 0.22</td>
<td>87.0</td>
<td>4.02 ± 0.26</td>
</tr>
</tbody>
</table>

* Complexity is the minimum number of parasite genotypes per infected person.
† Percentage of infections that exhibited more than one genotype.
‡ The overall mean infection complexity is the average of the highest number of clones detected by any microsatellite markers tested. The overall proportion of multi-infection refers to the percentage of infections with more than one clone detected by any 12 microsatellite markers.

### Table 3
Polymorphism of merozoite surface protein (msp) genes and infection complexity in *Plasmodium falciparum* from highland and lowland areas in western Kenya

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Kombewa</th>
<th>Frequency</th>
<th>Kakamega</th>
<th>Frequency</th>
<th>Highland</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>msp-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1</td>
<td>6</td>
<td>0.602</td>
<td>4</td>
<td>0.622</td>
<td>5</td>
<td>0.815</td>
</tr>
<tr>
<td>MAD20</td>
<td>5</td>
<td>0.186</td>
<td>5</td>
<td>0.180</td>
<td>2</td>
<td>0.185</td>
</tr>
<tr>
<td>RO33</td>
<td>2</td>
<td>0.212</td>
<td>2</td>
<td>0.197</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td></td>
<td>11</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>msp-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC1</td>
<td>7</td>
<td>0.615</td>
<td>6</td>
<td>0.644</td>
<td>3</td>
<td>0.554</td>
</tr>
<tr>
<td>FC27</td>
<td>8</td>
<td>0.385</td>
<td>7</td>
<td>0.356</td>
<td>3</td>
<td>0.446</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td></td>
<td>13</td>
<td></td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Average minimum no. of clones (95% confidence interval) 1.98 (1.77–2.19) 1.92 (1.69–2.15) 1.58 (1.34–1.82)

No. of clones estimated by the maximum likelihood method (range of support limit) 3.86 (3.02–4.53) 2.83 (2.15–3.79) 2.19 (1.53–2.87)

Percentage of infections with more than one clone 81.7 76.7 70.3

### Linkage disequilibrium.
For microsatellite analysis, 9 and 17 of the 66 possible tests for LD showed significant results at P ≤ 0.05 in the lowland and highland populations, respectively. None of the significant pairwise associations involved loci located on the same chromosome. The LD estimated by the LIAN program was significant in the lowland population (observed mismatch variance V_D = 3.66, expected mismatch variance V_E = 1.93, standardized index of association I_AS = 0.082, simulated 5% critical value L_MC = 2.09, P = 10^-4) and in the highland populations (V_D = 3.92, V_E = 1.94, I_AS = 0.093, L_MC = 2.16, P = 10^-4).

### Population structure.
The AMOVA found that most (92.2%) of the variance in allele frequencies was among individuals within populations; only 6.9% of the variation was explained by differences between highland and lowland populations (P < 0.001) (Table 4). Wright’s F-statistics indicated that 9 of 12 microsatellite markers detected significant population structure (Table 5). The overall genetic differentiation index among the three populations (F_st) was 0.027 (P <
0.0001), whereas differences between highland and lowland populations were pronounced ($F_{ST} = 0.036, P < 0.0001$). Seven of 12 microsatellite markers detected significant structure between Kombewa and Kisii, which had the lowest transmission intensity ($F_{ST} = 0.032, P < 0.0001$). Between the two highland populations, four of 12 markers detected significant structure, giving an average differentiation index of 0.017 ($P < 0.05$).

**Assignment test.** Two clusters, each representing parasites with similar microsatellite genotypes, were identified by the structure analysis. Assignment test found that 91.1–94.0% individuals could be assigned to particular locations (Table 6). From the location perspective, most (76%) of the malaria parasites from the lowland site (Kombewa) were assigned to one cluster, whereas parasites from the two highland sites represented a mixture of both clusters in similar frequencies (Table 6). These results indicate a high rate of parasite infiltration from lowland to highland areas.

**DISCUSSION**

In this study, microsatellite markers detected a significantly lower number of alleles, marginally lower expected heterozygosity, and reduced infection complexity in the highland populations in comparison to the lowland population. Similar trends in genetic diversity measures were found in *msp-1* and *msp-2*; however, the difference between highland and lowland populations was not statistically significant. Malaria in the lowland site used in this study is hyperendemic according to its transmissions intensity (annual EIR = 31.1) and prevalence in asymptomatic children (< 50%). The lack of statistical significance in population genetic diversity by *msp* genes is likely due to the fact that *msp* genes are under strong Selection pressure, the limited sample size analyzed in this study, and limited difference in malaria transmission intensities among the study sites.

The expected heterozygosity of microsatellite markers in our highland samples ($He$ range = 0.66–0.69) was substantially higher than that observed in South America (Colombia, Bolivia, and Brazil; $He$ range = 0.30–0.40) and in Thailand ($He = 0.51$), but was comparable to the samples from the Republic of the Congo. In comparison to other malaria-endemic African regions, genetic diversity of malaria parasites in western Kenya highlands was slightly lower. Studies with microsatellite markers showed expected heterozygosity of 0.76–0.80 for *P. falciparum* from malaria-endemic sites in Uganda, Congo, and Zimbabwe, and 0.79 for those from the Republic of the Congo. These estimates were comparable with our lowland Kombewa sample.

Interestingly, although highland parasite populations exhibited reduced infection complexity in comparison with the lowland population, the proportion of mixed infections was similar between the highland and lowland populations. Both microsatellite markers and *msp* gene detected a high proportion of mixed infections in the highland populations. The high proportion of mixed infections paralleled high genetic diversity in the highland. This observation raises a question of how high genetic diversity and high prevalence of mixed infections are generated and maintained in the highlands where malaria transmission intensity is generally low (e.g., EIR was estimated as merely 0.4 infectious bites per person in Kisii). One possible mechanism is the introduction of parasite clones from malaria-endemic lowland areas through human travel. Human travel is an important risk factor for malaria in the highlands. During the epidemic season, malaria transmission intensity is increased substantially over that in the non-epidemic season. Diversity in microsatellite alleles or *msp* haplotypes may be generated through meiotic recombination during the epidemic, a period with enhanced parasite infection rates in vector populations. If low-transmission during non-epidemic seasons did not lead to a substantial genetic

### Table 4

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Expected mean squares</th>
<th>Variance component</th>
<th>Percentage of variation</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups</td>
<td>1</td>
<td>32.4</td>
<td>32.4</td>
<td>0.65</td>
<td>6.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Among populations within groups</td>
<td>1</td>
<td>11.6</td>
<td>11.6</td>
<td>0.08</td>
<td>0.9</td>
<td>0.113</td>
</tr>
<tr>
<td>Among individuals within populations</td>
<td>147</td>
<td>1,231.6</td>
<td>8.4</td>
<td>8.63</td>
<td>92.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>149</td>
<td>1,275.6</td>
<td>52.4</td>
<td>9.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Marker</th>
<th>Among the three populations</th>
<th>Highland versus lowland</th>
<th>Within highland (Kakamega versus Kisii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA42</td>
<td>0.002</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td>Pf377</td>
<td>0.047$^{*}$</td>
<td>0.059$^{*}$</td>
<td>0.042$^{*}$</td>
</tr>
<tr>
<td>ARA2</td>
<td>0.054$^{†}$</td>
<td>0.056$^{†}$</td>
<td>0.004</td>
</tr>
<tr>
<td>TA87</td>
<td>0.005</td>
<td>0.028</td>
<td>0.012</td>
</tr>
<tr>
<td>2490</td>
<td>0.027$^{‡}$</td>
<td>0.041$^{*}$</td>
<td>0.035$^{‡}$</td>
</tr>
<tr>
<td>TA40</td>
<td>0.029$^{‡}$</td>
<td>0.038$^{‡}$</td>
<td>0.032$^{‡}$</td>
</tr>
<tr>
<td>PIPK2</td>
<td>0.02</td>
<td>0.015</td>
<td>0.006</td>
</tr>
<tr>
<td>TA109</td>
<td>0.038$^{‡}$</td>
<td>0.036$^{‡}$</td>
<td>0.015</td>
</tr>
<tr>
<td>PolyA</td>
<td>0.023$^{‡}$</td>
<td>0.043$^{‡}$</td>
<td>0.029</td>
</tr>
<tr>
<td>TA60</td>
<td>0.036$^{†}$</td>
<td>0.048$^{‡}$</td>
<td>0.007</td>
</tr>
<tr>
<td>TA81</td>
<td>0.023$^{‡}$</td>
<td>0.030$^{‡}$</td>
<td>0.011</td>
</tr>
<tr>
<td>TA1</td>
<td>0.022$^{‡}$</td>
<td>0.031$^{‡}$</td>
<td>0.013</td>
</tr>
<tr>
<td>Mean</td>
<td>0.027$^{*}$</td>
<td>0.036$^{*}$</td>
<td>0.017$^{‡}$</td>
</tr>
</tbody>
</table>

* $P < 0.001$.
† $P < 0.001$.
‡ $P < 0.1$.

### Table 6

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample size</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Not assigned</th>
<th>Percent assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kombewa</td>
<td>50</td>
<td>38</td>
<td>8</td>
<td>4</td>
<td>92.0</td>
</tr>
<tr>
<td>Kakamega</td>
<td>50</td>
<td>26</td>
<td>21</td>
<td>3</td>
<td>94.0</td>
</tr>
<tr>
<td>Kisii</td>
<td>45</td>
<td>21</td>
<td>20</td>
<td>4</td>
<td>91.1</td>
</tr>
<tr>
<td>Total</td>
<td>145</td>
<td>85</td>
<td>49</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

* Numbered clusters represent distinct groups identified by Bayesian cluster analysis using Structure computer software. 
† Not assigned refers to the number of individuals that could not be assigned to any cluster statistically.
bottleneck effect, population diversity and mixed infection could be maintained in the population. The question of whether low transmission intensity in the long non-epidemic season leads to parasite selfing and thus dramatic reduction in genetic diversity has not been determined. Evidence from the lack of significant seasonal fluctuations in the allelic frequencies of msp-1, msp-2, and the gene coding for glutamate-rich protein in Sudan, after a six-month dry season during which malaria transmission was not even detectable, suggests that genetic bottleneck in P. falciparum during low transmission season should not be severe. It would be interesting to determine the changes in genetic diversity and infection complexity throughout an epidemic season in the highland areas.

We observed small, but statistically significant population structure between the highland and lowland malaria populations (FST = 0.036, P < 0.0001). The lack of strong genetic differentiation between highland and lowland parasite populations is consistent with the notion of parasite introduction by human travel from the malaria-endemic lowland to the epidemic highland region. Shanks and others found that travel to malaria-endemic areas was significantly associated with increased malaria risk in highland residents. In addition to human travel, genetic structure of malaria parasite populations may be affected by parasite life history, site-specific epidemiology, and malaria control measures such as insecticide-treated bed nets and indoor residual spray. Selection by antimalarial drugs would also result in unique genetic imprints on parasite populations.

We found significant linkage disequilibrium in both highland and lowland populations in the present study. None of the significant linkage disequilibrium involved loci located on the same chromosome, which suggested that these loci are only statistically linked and not physically linked. Linkage disequilibrium in our highland populations was consistent with the expectation of low transmission. However, linkage disequilibrium in our lowland population present with a substantially stronger index of association (Iasso = 0.082) than those reported in other African sites (Iasso < 0.02) with comparable malaria prevalence. Sakihama and others also reported strong linkage disequilibrium in the msp-1 gene in P. falciparum populations from the Solomon Islands where the EIR exceeds several hundred infectious bites per person per year. These investigators suggested that transmission intensity is not the sole determinant of the strength of linkage disequilibrium. Other variables such as the number of prevalent alleles, the prevalence of mixed infections, and infection complexity may also be involved. In our study, more than 70% of infection cases were mixed genotype infections in the highlands and lowlands on the basis of microsatellite or msp markers. Multiple repeated multilocus genotypes may have contributed to the observed linkage disequilibrium. No significant linkage disequilibrium was found in P. falciparum populations from hypendemic areas in Africa when only genotypes of the mono-infected isolates were used.

In summary, the present study examined the genetic diversity of P. falciparum from symptomatic infections in western Kenya highland and lowland areas with contrasting transmission intensities. Despite reduced genetic diversity in the highland populations in comparison to populations in the surrounding endemic lowland, population genetic diversity in the highland area remained high. More than 70% of the infections were mixed genotype infections, even in areas of with an EIR < 1 infectious bite per person per year. Small geographic differentiation was detected between highland and lowland populations. Together with the results of risk factor analysis and malaria infection dynamics in the highlands, these results suggest that both human travel and local transmission may be important in shaping the genetic structure of malaria parasites in the highlands.

Received May 2, 2007. Accepted for publication August 23, 2007.

Acknowledgments: This study is published with the permission of the Director of the Kenya Medical Research Institute. We thank Claire Garro, Hong Chen, and two anonymous reviewers for helpful suggestions on the manuscript.

Financial support: This work was supported by grants R01 A150243, D43 TW01505, and R03 TW007360 from the National Institutes of Health.

Author’s addresses: Daibin Zhong and Guiyun Yan, Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697. Yaw Afrane and Andrew Githeko, Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya. Zhaqing Yang, Department of Parasitology, Kunming Medical College, Kunming 650031, Yunnan, People’s Republic of China. Liwang Cui, Department of Entomology, Pennsylvania State University, University Park, PA 16802. David M. Menge, Center for Infectious Diseases and Microbiology Translational Research, University of Minnesota, Minneapolis, MN 55455. Emmanuelle A. Temu, The Institute of Tropical Medicine, Nagasaki University, Nagasaki-shi 852-8523, Japan.

Reprint requests: Daibin Zhong, Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697, E-mail: dzhong@uci.edu.

REFERENCES
Genetic diversity


