Malaria's eve: Evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum

Permalink
https://escholarship.org/uc/item/13h7q1qw

Journal
Proceedings of the National Academy of Sciences of the United States of America, 95(8)

ISSN
0027-8424

Authors
Rich, SM
Licht, MC
Hudson, RR
et al.

Publication Date
1998-04-14

DOI
10.1073/pnas.95.8.4425

License
CC BY 4.0

Peer reviewed
Malaria’s Eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*  
(genetic polymorphism/demographic sweep/clonality/selective sweep/parasitic protozoa)

**STEPHEN M. RICH**, **MONICA C. LICHT**, **RICHARD R. HUDSON**, AND **FRANCISCO J. AYALA**†  
Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525

*Contributed by Francisco J. Ayala, February 23, 1998*

**ABSTRACT** We have analyzed DNA sequences from world-wide geographic strains of *Plasmodium falciparum* and found a complete absence of synonymous DNA polymorphism at 10 gene loci. We hypothesize that all extant world populations of the parasite have recently derived (within several thousand years) from a single ancestral strain. The upper limit of the 95% confidence interval for the time when this most recent common ancestor lived is between 24,500 and 57,500 years ago (depending on different estimates of the nucleotide substitution rate); the actual time is likely to be much more recent. The recent origin of the *P. falciparum* populations could have resulted from either a demographic sweep (*P. falciparum* has only recently spread throughout the world from a small geographically confined population) or a selective sweep (one strain favored by natural selection has recently replaced all others). The selective sweep hypothesis requires that populations of *P. falciparum* be effectively clonal, despite the obligate sexual stage of the parasite life cycle. A demographic sweep that started several thousand years ago is consistent with worldwide climatic changes ensuing the last glaciation, increased anthropophilia of the mosquito vectors, and the spread of agriculture. *P. falciparum* may have rapidly spread from its African tropical origins to the tropical and subtropical regions of the world only within the last 6,000 years. The recent origin of the world-wide *P. falciparum* populations may account for its virulence, as the most malignant of human malarial parasites.

There is an extensive literature indicating that the agent of malignant malaria, *Plasmodium falciparum*, is highly polymorphic. The studies have focused on antigenic determinants, drug-resistance phenotypes, allozymes (1–4), and chromosome sizes (5). Assessment of DNA sequence variation has been based almost exclusively on examination of several genes coding for antigenic determinants, where amino acid polymorphisms (nonsynonymous nucleotide polymorphisms) are common and likely to be affected by natural selection (6, 7). Our study seeks to ascertain certain population parameters by examining synonymous polymorphisms, which do not change the amino acid sequence of the encoded proteins and are less likely to be under selective pressure.

**MATERIALS AND METHODS**

DNA Sequences. The 10 genes studied and the geographic origin of the strains are listed in Table 1. The genes are from isolates of *P. falciparum* collected from global malaria endemic regions. The *Dhfr* and *Ts* genes are found directly adjacent to one another on the parasite’s fourth chromosome and encode the bifunctional dihydrofolate reductase–thymidylate synthetase (DHFR–TS) domain. Certain mutations in the *Dhfr* gene have been widely associated with *P. falciparum* resistance to antifolate drugs, including pyrimethamine. Two other genes have been implicated with drug-resistant phenotypes of *P. falciparum*: the gene coding for dihydropteroate synthetase (*Dhps*) and the gene for multidrug resistance (*Mdr1*). The circumsporozoite protein (encoded by *Csp1*) is antigenic, and the rhoptry-associated protein (encoded by *Rap1*) may also be immunogenic. The other four genes in Table 1 are not known to be immunogenic or associated with resistance to any antimalarial drug currently in use. They code for calmodulin (*Caln*), glucose-6-phosphate dehydrogenase (*G6pd*), heat-shock protein 86 (*Hsp86*), and triose phosphate isomerase (*Tpi*).

Alignment and Phylogenetic Analysis. We align each set of the gene sequences by means of a progressive multiple-sequence alignment algorithm, using the CLUSTALW computer program (12), with corrections made by eye. Only the gene coding sequence is analyzed; introns and flanking regions are not considered, because of limited availability. The center region of the *Csp1* gene contains several repeating units that cannot readily be aligned (see ref. 13) and, therefore, it has been removed from this analysis. We analyze separately the two regions, 5′ and 3′, that embrace the multiple-repeat middle region of *Csp1*.

We perform the phylogenetic analysis of the *Dhfr* gene by using the branch-and-bound search method with maximum parsimony optimization criterion of PAUP (14).

**RESULTS**

Table 2 summarizes the polymorphisms observed in the 10 genes studied. Amino acid replacement polymorphisms occur in the three drug-resistance genes *Dhfr*, *Dhps* and *Mdr1*, as well as in *Csp1*, the presumed immunogenic *Rap1* gene, and *G6pd*. No silent polymorphisms are observed in these or any other genes. Silent polymorphisms are polymorphisms that do not affect the amino acid sequence coded. Silent polymorphisms occur in degenerate codons. We will distinguish 4-fold and 2-fold degenerate codons. The 6-fold degenerate codons for amino acids serine, leucine, and arginine are grouped according to their 4-fold or 2-fold codons. The 3-fold degenerate codon for isoleucine is included with the group of 4-fold degenerate codons for the present analysis.

We do not include β-tubulin (*β-tub*), of which three sequences are known (GenBank accession nos. M28398, M31205, and X16075), which exhibit three silent polymorphisms and 24 amino acid polymorphisms. The polypeptides α-
and β-tubulin form the dimeric protein tubulin, which is a major structural component of microtubules. In most protozoans, including species of *Trypanosoma* and *Leishmania*, these genes occur in tandem, as repeated linear arrays along the chromosome. In *P. falciparum*, the α- and β-tub genes are located on different chromosomes, and each may be represented by at least two copies, as evidenced by multiple mRNA transcripts differing in their untranslated regions (15), although it has been suggested that these transcripts may result from posttranscriptional modification of a single mRNA (16).

The large number of nonsilent polymorphisms in the three β-tub sequences suggests that they are not orthologous, i.e., alleles of the same locus. This is corroborated by noticing that the average genetic distance (incidence of amino acid replacements) between the three *falciparum* sequences is $D = 0.040 \pm 0.025$, whereas between *P. falciparum* and *Plasmodium berghei*, $D = 0.023 \pm 0.010$, although these two species diverged 55–129 million years (My) ago (17).

We first consider the hypothesis that the absence of silent polymorphism is a consequence of a recent population bottleneck (or demographic sweep), so that the extant world populations of *P. falciparum* can be traced to a single recent ancestor. If the population grew to a large size after the bottleneck, it is reasonable to assume that the genealogy of a sample of multiple strains collected from widely distributed localities would be a star-like phylogeny (Fig. 1) with the most recent common ancestor at the vertex of the star (18). Under this assumption, and ignoring the possibility of multiple hits at

### Table 1. Genes examined in the present survey of genetic variation among global populations of *P. falciparum*

<table>
<thead>
<tr>
<th>Location</th>
<th>Strain</th>
<th>Dhfr</th>
<th>Ts</th>
<th>Dhs</th>
<th>Mdr1</th>
<th>Rap1</th>
<th>Calm</th>
<th>G6pd</th>
<th>Hsp86</th>
<th>Tpi</th>
<th>Csp1</th>
</tr>
</thead>
<tbody>
<tr>
<td>South America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honduras</td>
<td>HB3</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>7G8</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>It.G2.F6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>It.D12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El Salvador</td>
<td>INTM22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambia</td>
<td>FCR3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambia</td>
<td>FC27</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>SL/D6</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>NF7</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>M24</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Palo Alto</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>1/CDC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papua New Guinea</td>
<td>MAD20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>K1</td>
<td>2</td>
<td>2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>K39</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>Cxi-2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>GH2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>T4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>T9</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>Tak9</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>TM335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>806-946</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>V1</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>V1/S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>Ind C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>W2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>Ind-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>Ind-J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>Dd2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>CVD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>3D7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>NF54*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The symbol x indicates a single DNA sequence per strain, whereas numbers indicate multiple isolates from the same strain; for Csp1, the Thailand sequences labeled 806-946 are from 13 strains. The GenBank accession numbers (or reference), successively from top to bottom, are as follows: Dhfr, J03772, J04643, J03772, J03028, J03772, M22159, J03772, J03772, J03772, and J03772; Tst, J03772, J04643, J03772, J03028, J03772, M22159, J03772, J03772, J03772, and J03772; Ts, J03772, J04643, J03772, J03028, J03772, M22159, J03772, J03772, J03772, and J03772; Calm, X56950 (ref. 9), M99442 (ref. 9), and M59770 (ref. 9); G6pd, M80655 and X74988 (the authors of ref. 10 submitted the G6pd sequence for isolates K1 and 3D7 under a single accession number); Mdr1, X56855, S53996, and L02513; Rap1, J02985, L10322, L10323, M32653, U41077, U41073, U41075, U41074, and U41076; Calm, X56950 (ref. 9), M99442 (ref. 9), and M59770 (ref. 9); Tpi, L01654 and L01655; Csp1, K02194, U20969, M15505, M83172, M83174, M19752, M57499, M57498, M83173, M83149, M83150, M83155, M83156, M83161, M83163, M83164, M83165, M83166, M83167, M83168, M83169, M83170, M83886, X15363, and M22982. *NF54, CVD1, and 3D7 are laboratory strains isolated from a patient in the Netherlands who did not traveled in malaria endemic regions.

1Ref. 11 and J. Reeder (personal communication).
individual sites, the number of neutral polymorphisms that we observe in a sample of multiple strains will have a Poisson distribution with a mean that depends on the neutral mutation rate, the time elapsed since the bottleneck; and \( l_i \) and \( m_i \) are the numbers of 4-fold and 2-fold synonymous sites examined at the \( i \)th locus. This expression suggests an estimator of the time of the bottleneck, obtained by solving for \( t \) and replacing \( \lambda \) by \( S \), the observed number of polymorphisms:

\[
i = \frac{S}{\mu_4 \Sigma n_i l_i + \mu_2 \Sigma n_i m_i}.
\]

In our sample \( S = 0 \), so \( i = 0 \). Because \( S \) is assumed to be Poisson-distributed, we can estimate an upper 95% confidence limit, \( t_{95} \), for the time of the bottleneck by finding the value of \( t \) for which the probability of no polymorphism \((e^{-\lambda})\) equals 0.05. Because \( e^{-2.996} = 0.05 \), we calculate the \( t_{95} \) by writing 2.996 in the numerator of Eq. 1.

To use Eq. 1, we need an estimate of the neutral mutation rate at third-position degenerate codons. We estimate this neutral mutation rate for \( Dhfr \) by comparing the \( P. falciparum \) gene sequences with those available for other species, the rodent parasites \( P. berghei, Plasmodium chabaudi, Plasmodium vinckei \), and \( Plasmodium yoelii \). Fig. 2 shows the phylogenetic relationships among these taxa, based on the \( Dhfr \) gene sequences. These relationships are in agreement with phylogenies based on other genes (6, 19). There are 146 amino acid differences between \( P. berghei \) and \( P. falciparum \). Among the unchanged amino acid sites, there are 69 and 51 observed synonymous nucleotide differences among the 4-fold (\( n = 142 \)) and 2-fold (\( n = 264 \)) degenerate codons, respectively.

The observed number of nucleotide differences is probably less than the actual number of substitutions that have occurred since divergence, because multiple substitutions at individual

![Fig. 1. Schematic representation of a star phylogeny; \( t \) represents the time elapsed between the population bottleneck (cen ancestor) and the present.](image)

![Fig. 2. Phylogeny of five \( Plasmodium \) species based on 592 coding nucleotides of the \( Dhfr \) gene. The tree is a majority-rule consensus, with midpoint rooting of 1,000 bootstrap replicates derived by the branch-and-bound method using the maximum parsimony optimality criterion of the PAUP 3.1.1 program (14). Bootstrap values are given above each branch. GenBank accession numbers of the sequences used are given in parentheses after the species names. Thick branches indicate the divergence between \( P. falciparum \) and \( P. berghei \), which occurred no earlier than 129 My ago.](image)
sites would be observed as single differences or no difference. To correct for overlapping substitutions between sequences we use the Jukes–Cantor correction, which is conservative (i.e., it undercorrects) because it assumes that all changes occur with equal frequency, an unrealistic assumption, particularly for genomes such as that of *Plasmodium* spp. for which A+T richness is around 74% or greater (17). With the Jukes–Cantor method, we estimate that 111.2 4-fold and 64.5 2-fold synonymous substitutions have occurred between *P. falciparum* and *P. berghei*. The number of substitutions at 4-fold degenerate sites is estimated by 142(−3/4)ln[1 − (−4/3) × (69/142)] = 111.2. At 2-fold degenerate sites a two-state model is assumed, and the number of substitutions is estimated by 264(−1/2)ln[1 − (−2 × (51/264))] = 64.5. The conservative bias of the Jukes–Cantor model yields lower estimates of μ and correspondingly higher estimates of t₀ than would be obtained with other methods. The radiation of the genus *Plasmodium* has been estimated to have occurred between 55 and 129 My ago (19). If we assume, conservatively, that the divergence has been estimated to have occurred between 55 and 129 My ago, this would effectively amount to a population bottleneck of many hundreds of individuals for very many generations until not long ago, this would effectively amount to a population bottleneck.

There seems to be no reason to suspect that spontaneous mutation rates are exceptional in *P. falciparum* (hypothesis ii), and there are two arguments against it. One is the high incidence of polymorphisms at antigenic and drug-sensitivity sites, both in worldwide samples (3, 21–24) and in laboratory selection experiments with mice (25). The other argument is that there is divergence, in synonymous and nonsynonymous sites, between *P. falciparum* and other *Plasmodium* species (7, 26, 27).

Similarly, there seems to be no reason to suspect that there are strong selective constraints against silent variation in *P. falciparum* (hypothesis iii). *P. falciparum* has a 76.4% incidence of A+T, which is typical of other *Plasmodium* species; e.g., we have observed 75.8% in *P. berghei*. In addition, silent polymorphisms appear in the comparisons between *P. falciparum* and other *Plasmodium* species (Fig. 2 for Dhfr; see refs. 7 and 26 for other genes) at the typical rates observed for other organisms. Moreover, two clinical isolates of a related species, the Apicomplexan *Cryptosporidium parvum*, exhibit 29 silent polymorphisms among 1,530 bp of the Dhfr gene (sequences from ref. 28; GenBank accession nos. U41365 and U41366).

Natural selection (hypothesis iv) may account for the rapid spread of a favored genotype throughout populations, particularly when the population is large and/or the selection is strong. The repeated appearance throughout global malaria endemic regions of drug-resistant phenotypes, determined by nonsynonymous substitutions at the Dhfr, Dhps, and other loci, is most likely due to natural selection. Selection sweeps are known in other organisms, such as *Drosophila melanogaster*, where a single nucleotide sequence at the Sod locus is present in about 50% of all haplotypes throughout the world, without any silent substitutions along the 1,500-bp sequence, although there is one widespread nonsynonymous substitution that accounts for a fast–slow allozyme polymorphism (29, 30).

Natural selection can account for the absence of synonymous variation at any of the 10 loci we have analyzed, if the particular gene sequence itself (or a gene with which it is linked) has been subject to a recent worldwide selective sweep, without sufficient time for the accumulation of new synonymous mutations. However, the 10 genes are located on, at least, six different chromosomes, and so six selective sweeps would need to have occurred more or less concurrently, which seems unlikely. A selective sweep simultaneously affecting all chromosomes could happen if the population structure of *P. falciparum* were predominantly clonal rather than sexual.

### Table 3.

<table>
<thead>
<tr>
<th>Mutation rate (×10⁻⁹)</th>
<th>Assumption</th>
<th>4-fold</th>
<th>2-fold</th>
<th>t₀</th>
<th>t₉₀</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>55 My</td>
<td>7.12</td>
<td>2.22</td>
<td>24,511</td>
<td>5,670</td>
</tr>
<tr>
<td></td>
<td>129 My</td>
<td>3.03</td>
<td>0.95</td>
<td>57,481</td>
<td>13,296</td>
</tr>
<tr>
<td></td>
<td><em>P. falciparum–reichenowi</em></td>
<td>5 My</td>
<td>3.78</td>
<td>1.86</td>
<td>38,136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 My</td>
<td>2.70</td>
<td>1.33</td>
<td>53,363</td>
</tr>
</tbody>
</table>

* t₀ gives the upper boundary of the 95% confidence interval of the estimate; t₉₀ represents the estimated time above which the probability is greater than 50% of observing more variation than actually observed.
Some authors have argued that there is no evidence that *Plasmodium falciparum* propagates clonally in nature but rather predominantly propagates by crossing between dissimilar parasites (31). This inference is based on three observations: (i) high incidence of mixed infections in individual patients, (ii) presence of two unlike alleles in diploid oocysts isolated from mosquito midguts, and (iii) evidence of intragenic recombination. However, the question whether *P. falciparum* has an effectively clonal population structure is not incompatible with the occurrence of recombination in this obligatory sexual parasite. The issues are whether recombination is meiotic (rather than mitotic) and how often it occurs. Our analysis of polymorphisms in antigenic peptide repeats of the *Csp1* gene indicates that these polymorphisms can be generated by somatic (mitotic), rather than sexual (meiotic), recombination. Moreover, the rate of intragenic recombination and the strength of linkage disequilibrium between nucleotide sites are independent of the distance between nucleotides, which is inconsistent with meiotic recombination but is consistent with a clonal population structure for *P. falciparum* (13). The evidence used for rejecting the hypothesis of clonality is based upon genetic polymorphisms at loci that code for antigenic determinants (32, 33). These loci may be under strong selection for avoiding the host’s immune responses, and it is, therefore, difficult to determine how frequently recombination takes place, because even rare recombinants could rapidly reach high frequency. For example, *Escherichia coli* is regarded as being clonal (34) and hence recombination is relatively rare; nonetheless the identification of recombinant segments within genes is not uncommon (35, 36). In light of this, the evidence currently available is not sufficient to reject the hypothesis that *P. falciparum* has, like *Trypanosoma* and *Leishmania* (37–43), a clonal population structure.

It may also be pointed out that a clonal population structure is not inconsistent with high levels of polymorphism, as observed in *P. falciparum* (for review, see ref. 31). Thus, the observation of 60 different phenotypes in 60 isolates provides no grounds for rejecting clonality out of hand, as done in ref. 2. One well known example (among many available ones) of extensive genetic variation in a clonal organism is *E. coli*, in which it is rare to find multiple independent isolates exhibiting the same multiple locus enzyme electrophoresis (MLEE) type; rather, a sample of 60 isolates is likely to contain 60 MLEE types (34, 44, 45). Genetic polymorphism is very extensive in *Trypanosoma cruzi*. On the basis of allozyme variation at 13 loci in a sample of 121 stocks, 43 different genotypes were found, 27 of which were present only once in the sample (37, 41, 43). Yet the clonal population structure of *T. cruzi* is well established (37–43). For *P. falciparum*, the data in Table 2 show that amino acid polymorphisms are common; thus, multiple multilocus genotypes will exist by accumulation of different amino acid polymorphisms independently arisen at different loci by natural selection. Arnot (46, 47) attributed the paucity of synonymous substitutions in *P. falciparum* to a codon-use bias. Codon-use bias is pervasive throughout the *Plasmodium* genus; however, it does not constrain silent polymorphisms between *Plasmodium* species or within species other than *falciparum*.

How could we account for a recent demographic sweep in *P. falciparum*, if it were not due to natural selection? One possible hypothesis is that *P. falciparum* has become a human parasite in recent times, by lateral transfer from some other relative of *P. falciparum*. We thank Mario Coluzzi, Allan Dickerman, Victor DeFilippis, Walter Fitch, Anthony James, Benjamin Rosenthal, Steve Schrödi, Andrew Spielman, and Andrey Tatarenkov for helpful comments on the manuscript. We thank John Reeder for sharing unpublished *Dhfr* gene sequence data. This work was supported by National Institutes of Health Grant GM42397 to F.J.A.