GENETIC DIVERSITY OF PLASMODIUM FALCIPARUM FIELD SAMPLES FROM AN ISOLATED COLOMBIAN VILLAGE

DIANA GÓMEZ, JACQUELINE CHAPARRO, CLAUDIA RUBIANO, MARÍA OFREA ROJAS, AND MOISES WASSEMAN

Laboratory of Biochemistry, Instituto Nacional de Salud, Bogota, Colombia; LIBBIQ, Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract: Colombian field isolates of Plasmodium falciparum were analyzed for genetic diversity. Fifty-three samples were collected as thick smears from patients living in Panguí, an isolated area with low migration. While the samples were being collected, Panguí was experiencing an epidemic outbreak of malaria. The samples were typed using nested polymerase chain reaction (PCR) amplification of block 2 of the merozoite surface protein 1 (MSP1) gene and nested PCR with mutation-specific primers for position 108 of the dihydrofolate reductase enzyme gene. The results for the circulating population of parasites in Panguí show low diversity—four allelic forms—using MSP1 as a marker, a fact that contrasts with data reported for certain Asian and African zones. A high percentage of mixed infections was observed, as was high complexity of the infection. No differential distributions were found for any allelic type.

INTRODUCTION

The genetic diversity of Plasmodium falciparum has been regarded as one of the main reasons that only partial immunity to malaria develops and only after multiple exposure to the parasite. This diversity is, therefore, a major problem in the development of vaccines. Various studies in endemic areas have shown natural parasite populations to be highly diverse. Often, more than one genetically different parasite population is found in the same individual.\(^1\)\(^-\)\(^5\)

In the present study, we analyzed the genetic diversity of \(P.\ falciparum\) on the basis of 53 blood samples collected from infected patients during an epidemic outbreak. All the patients live in Panguí, a small area in the Department of Chocó in northwest Colombia. It is an isolated zone where the population has a low rate of migration and the endemicity of malaria is high compared with other parts of Colombia and Latin America, but not so high when compared with some African zones. However, when the samples were collected, the prevalence of malaria in the population was 60%.

The genetic markers used in the study to analyze the variability of the samples were the merozoite surface protein 1 (MSP1) and the enzyme dihydrofolate reductase (DHFR) genes. The antigen MSP-1 has been considered a good candidate for the development of a vaccine against intraerythrocytic stages of the parasite. The gene codifying this protein has been widely used as a genetic marker and has shown a high degree of polymorphism in various studies of diverse geographic origin.\(^6\)\(^,\)\(^7\) Based on sequence analysis, the MSP1 gene is reported to be divided into 17 blocks: ten conserved or semiconserved and seven highly variable.\(^8\) This gene is dimorphic in each block, and the alleles have been designated as K1 and MAD20. The only exception occurs in block 2, where there is a third allele known as RO33. The three alleles establish themselves depending on the presence, type, and number of tripeptide repetitions found in the sequence of this block. Block 2 of MSP1 was, therefore, chosen as the marker. For its part, DHFR has three allelic forms in position 108, the wild type serine (Ser), the mutant type asparagine (Asn), and the mutant type threonine (Thr). DHFR is a target of several widely used antimalarials, and the genotype in position 108 has been associated with \textit{in vitro} resistance to pyrimethamine and cycloguanyl.

In a previous study in Colombia on diversity of the parasite, 31 samples from patients in different and distant regions of the country were analyzed. This work used MSP1 as the marker and showed the predominant allele to be RO33; a low percentage of mixed infections was observed as well.\(^8\) We believe it is important to study the genetic variability of \(P.\ falciparum\) found simultaneously in a specific location where the parasite is highly prevalent. This will provide an idea of the potential diversity of the parasite in Colombia.

MATERIALS AND METHODS

Clinical samples. The study was conducted in Panguí, a community of 400 inhabitants in the Department of Chocó, in northwest Colombia. The area has little industrial development, is infested with a primary vector, \textit{Anopheles albimanus}, and is characterized by low altitude, relatively high humidity, constant rain, and an average temperature of 27°C. The analyzed samples were collected during an outbreak in which the prevalence of malaria was 60%, measured as the percentage of people with parasites among those who presented with malaria symptoms at a health service. The Corporación de Investigaciones Biológicas in Medellín collected the samples from July to September 1997 by finger puncture in the form of thick smears on slides. There was no active survey to collect the samples, but during the period under study, almost 200 inhabitants approached the local health authority for an examination. The present study was done on the first 53 positive samples, all taken before treatment was administered. The preferred treatment was chloroquine, with the combination pyrimethamine-sulfadoxine used in few recurrent infections. There is no information on the degree of resistance to pyrimethamine-sulfadoxine in the community at the time of the outbreak. The samples were stained with Giemsa, and the presence of \(P.\ falciparum\) was detected under microscopic observation. The slides were then sent to the Instituto Nacional de Salud in Bogota, and parasitemia was determined and normalized for 100 leukocytes. This information was converted into the number of parasites per microliter, assuming a leukocyte count of 8,000/µL. To prepare the DNA of the clinical samples, the thick smear, moistened previously with 1% saponin, was carefully removed from the slide and resuspended in 1% saponin, then incubated for one hour at 4°C. The sample was centrifuged at 12,000xg for five minutes, and the supernatant discarded. The precipitate was resuspended in 40 µL of Chelex-100 5% and boiled for 10 minutes. Finally,
the sample was centrifuged at 12,000xg for five minutes, and the supernatant was recovered and stored at 4°C. We used 8 µL of the extract to amplify the MSP1 gene through polymerase chain reaction (PCR) and 1 µL for the DHFR gene.

These first amplifications were used as a template for respective nested PCR assays. We took 5 µL to identify the allotypes in block 2 of MSP1 and 0.1 µL to identify the genotype in position 108 of DHFR.

**DNA extraction from reference strains.** For the assays with MSP1, DNA from the Haiti 135, FCB2, and 7G8 strains was used, respectively, as a positive control for the K1, MAD20, and RO33 alleles. DNA from the 3D7, FCDB, and T4 strains was used, in turn, in the DHFR assays as a positive control for the genotypes Ser-108, Thr-108, and Asn-108. The strains were kept in culture via the Trager and Jensen method, using human blood type O Rh+. The blood was suspended at 5% in an RPMI-1640 culture medium supplemented with 0.2 mM hypoxanthine, 25 mM HEPES, 32 mM NaHCO3, 1 mg/L reduced glutathione GSH, 50 µg/mL gentamycin, and 10% inactive human serum. The culture was synchronized through differential lysis with 0.5% sorbitol. To extract DNA, these cultures were treated with 0.15% saponin in HBS at room temperature for 10 minutes. The parasites were recovered through centrifugation at 15000xg for 10 minutes at 4°C and incubated in a buffer with proteinase K 0.2 mg/mL, SDS 0.5%, EDTA pH 8.0 25 mM, and Tris-HCl pH 8.0 10 mM for 18 hours at 50°C. The lysate was subjected to two phenol extractions followed by a phenol:chloroform extraction (1:1), and dialyzed against TE (10 mM Tris-HCl, 1 mM EDTA).

**Allotype detection in block 2 of MSP1.** In the PCR-based amplifications for the MSP1 gene and in the nested PCR to identify the allotypes in block 2 of this gene, we used an amplification profile with an initial denaturation at 94°C/2 min followed by 72°C/2 min, time during which the Taq DNA polymerase was added. Then, 35 cycles at 94°C/30 sec, 55°C/45 sec, and 74°C/45 sec, with a final extension at 72°C/2 min. The following primers were used to amplify the MSP1 gene: OK1 (5’ TAG AAG ATG CAG CAG TAT CAG GAT CAG TTT TA 3’) and OK12 (5’ CAT CAG TAT TGA CAG GTT A 3’) to define the allele present in block 2 of the MSP1 gene, the primers OK1 (5’ CTA ATT CAA GTG GAT CAG TAA ATA A 3’) and OK2 (5’ GAG GCC TTG CAC CAG ATG AAG T 3’) were used for the K1 allelic type. The primers OK3 (5’ GTA TTA AAT GAA GGA ACA AGT GGA ACA 3’) and OK4 (5’ TAT CTG AAG GAT TTG TAC GTC TTG AAT T 3’) were used to amplify the MAD20 allelic type. The primers OK5 (5’ ATT AAA GGA TGG AGC AAA TAC TCA AGT TGT 3’) and OK6 (5’ TCT GAA GTA TTG GCA CCT GGA GA 3’) were used to amplify the RO33 allelic type. Both the primary and the nested PCR were conducted in a final volume of 50 µL, using 200 µM of each dNTP, 1 µM of each primer, 2.5 µL of Taq DNA polymerase (Promega) per reaction, and 1.5 mM of MgCl2 in the enzyme buffer (50 mM KCl, 10 mM Tris-HCl, pH 9.0, 0.1% Triton X-100).

The PCR products for the two markers were subjected to electrophoresis in a 2.0% agarose gel stained with an aqueous solution of 0.5 µg/mL ethidium bromide and analyzed with the One-Dscan one-dimensional electrophoresis program (Scanalytics, Fairfax, VA).

**Allelic distribution and complexity of infection.** The prevalence of each allelic type analyzed was determined as the percentage of PCR fragments for the type in the total number of amplified bands for the corresponding locus. The complexity of infection, which is the average number of PCR bands per infected individual, was determined as described earlier. The percentages for the type and complexity of infection were calculated independently for each genetic marker. A x² test compared the distribution of the DHFR genotypes, and a Yates correction was conducted when required (Sigma Stat 2.0, Jandel, Ringoes, NJ).

**RESULTS**

Fifty-three samples were analyzed: 57% (N = 30) pertained to children between birth and 17 years of age, 40% (N = 21) to adults between 16 and 70 years of age, and 3% (N = 2) to patients of unknown age. Of the people analyzed, 57% (N = 30) were females. Most of the children were students. The adults were farm or domestic workers.

Genetic diversity was analyzed through PCR amplification of polymorphic regions in the MSP1 and DHFR genes. The three previously reported allelic types were found in block 2 of MSP1, and in addition, two subtypes were found for MAD20. (We called them MAD20-1 and MAD20-2.) These subtypes were established on the basis of differences in size observed through electrophoresis and showed respective sizes of 218 pb and 238 pb. The detected fragments of type K1 showed only one size (210 pb), as was the case with those of type RO33 (190 pb). Accordingly, they were regarded as simple alleles. The allele most often present in the parasite population of the samples studied was MAD20, with a frequency of 63% (N = 66). Type K1 registered a frequency of 11% (N = 12) and RO33 26% (N = 27). Using DHFR as a marker, the frequency of the Ser wild genotype was 10% (N = 3), and the frequency of the Thr mutant genotype and the Asn mutant genotype each was 45% (N = 13).

In analyzing single-copy genes such as MSP1 and DHFR, only one fragment of PCR is expected to be generated by the
haploid parasite. This study showed a high frequency of mixed infections, with the presence of more than one type of genetically different parasite population in the same individual. As illustrated in Figure 1A, using MSP1 as the marker, 30% (N = 16) of the isolates were catalogued as simple infections and 70% (N = 37) as mixed. With respect to the simple infections, the frequency of MAD20-1 was 26% (N = 14) and that of RO33 was 4% (N = 2). None was found with the K1 allotype or with MAD20-2. Seven combinations were found among the mixed infections, with the following distribution: 9% (N = 5) MAD20-1 + MAD20-2; 9% (N = 5) K1 + MAD20-1; 27% (N = 14) RO33 + MAD20-1; 4% (N = 2) MAD20-1 + MAD20-2 + K1; 11% (N = 6) MAD20-1 + MAD20-2 + RO33; 6% (N = 3) MAD20-1 + K1 + RO33; and 4% (N = 2) MAD20-1 + MAD20-2 + K1 + RO33. Up to four MSP1 genotypes were detected in a single infected person. According to the method of analysis used in the study, this is the maximum number possible.

Using position 108 of the DHFR gene as a marker, 55% (N = 29) of the samples were catalogued as simple infections (Figure 1B). Most of them correspond to the mutant genotypes Thr and Asn, each 25% (N = 13). The Ser wild genotype was found in only 5% (N = 3) of the samples. With this marker, the remaining 45% (N = 24) of the samples were catalogued as mixed infections, with four combinations being found: 19% (N = 10) Thr + Asn; 15% (N = 8) Ser + Asn; 5.5% (N = 3) Ser + Thr; and 5.5% (N = 3) Ser + Thr + Asn. It should be noted that some patients were carrying the three DHFR genotypes studied.

Figure 2 compares the distribution of genotypes in position 108 of the DHFR gene, corresponding to the infant population of the sample, with the parasitemia (Figure 2A) and the age of the children (Figure 2B). When comparing with the parasitemia, the samples were grouped into: a) fewer than 10,000 parasites/μL; b) 10,000 to 20,000; and c) more than 20,000. Half of the samples in the first group were mixed infections. The single-infection distribution was three Thr, two Asn mutants, and no Ser wild type. In the second group, mixed infections also were the most frequent (five), and the single infections were distributed in the three possible cases as shown in the figure. A similar result was observed in the third group. No statistically significant association was found between parasite density and genotype distribution.

A similar result was obtained when the distribution of genotypes was compared with the age of the infected children (Figure 2B). Three groups were conformed as follows: a) samples from children younger than 4 years; b) from children 4 to 8 years; and c) from children over 8 years. The distribution of the genotypes of the parasites do not show any statistically significant association with the age of the infected children.

When the results for MSP1 and DHFR were analyzed as a whole, the frequency of simple infections declined to 17% (N = 9). We found 31 types of multilocus MSP1 + DHFR associations and up to six genotypes in a single patient (not shown). Table 1 shows the complexity of infection, which was 1.98 for MSP1 and 1.51 for DHFR. Combining the data from the two loci yielded 2.15 fragments per infected person.

DISCUSSION

The structure of natural P. falciparum populations plays a highly important role in the natural acquisition of immunity in malarial infections. Knowledge of this structure is necessary to develop strategies to control the disease, beginning with the design of effective vaccines against P. falciparum and including policy on the use of antimalarial medicines. We ana-

FIGURE 1. Distribution of the type of infection and allelic combinations. Plasmodium falciparum samples were collected at Panguí, Colombia, from July to September 1997. In (A), block 2 of the MSP1 gene was used as a marker. The conventions are K for K1, M1 for MAD20-1, M2 for MAD20-2, and R for RO33. In (B), position 108 of DHFR was used as a marker. The conventions are S for serine, T for threonine, and A for asparagine.
lyzed the genetic diversity of *P. falciparum* isolates collected during an epidemic outbreak in an area of Colombia endemic for malaria but one where there is no permanent transmission of the parasite, as in certain African zones. However, prevalence was extremely high when the samples were collected. The community is closed, with little migration, and is geographically isolated. The only way in or out of the area is by river or bridle path.

From July to September 1997, four MSP1 genotypes were detected in the *P. falciparum* population in the Colombian endemic zone under study. Although two subtypes for MAD20 were found, no other variants were encountered for types K1 and RO33. Nonetheless, given the analytic method used, we cannot rule out the possibility that subtle differences in the size of the amplified fragments or specific changes in the sequence went unobserved. This limited polymorphism is in keeping with what was reported for areas of low endemicity, such as French Guiana,16 Honduras,17 and Vanuatu,18 and contrasts with what was found in Senegal4 and Thailand.3 The particular features of the study zone—namely, geographic isolation and a closed population—may cause the low introduction of new allelic types via migration, which could be a probable cause of the limited polymorphism observed.

Several similar studies in different geographic areas, using block 2 of MSP1 as a marker, report important variations in the frequency of the genotypes. For example, we found MAD20 to be the predominant allele in the parasite population and K1 to be the least frequent. This is contrary to the situation observed in French Guiana,16 where MAD20 is the less frequent type and K1 the most frequent. On the other hand, in a study in Thailand,3 it was found that two variants of the MAD20 allelic family were disassociated and their simultaneous presence rarely observed. However, in our study, we found subtypes MAD20-1 and MAD20-2 to be strongly associated, so much so that no infection was found with genotype MAD20-2 independent of MAD20-1.

When various genotypes for a specific parasite locus are found in a host, the possibility of cross-fertilization and meiotic recombination in the mosquito vector increases. To some extent, genetic diversity is expected to depend on the proportion of mixed infections and the number of clones per individual.6 The presence of mixed infections is one of the prerequisites if cross-fertilization is to generate new genotypes and add to the diversity of the parasite population. Nevertheless, our study showed a high percentage of mixed infections (83%) in infected individuals and a very complex infection, in spite of the low polymorphism observed. In Guiana, (an area of low endemicity similar to that of the present study) using the MSP1, MSP2, and GLURP markers, the percentage of mixed infections was 10%, complexity of infection was 1.1,

**TABLE 1**

<table>
<thead>
<tr>
<th>Percentage of mixed infections and complexity of infection, using MSP1 and DHFR as markers in panguı´, Colombia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSP1</strong></td>
</tr>
<tr>
<td>Percentage of mixed infections</td>
</tr>
<tr>
<td>Complexity of infection (fragments/individual)</td>
</tr>
</tbody>
</table>

Fifty-three samples were analyzed. Both the percentage of mixed infections (individuals with more than one genetically different population of parasites) and the complexity of the infection (number of amplification fragments found per sample) were determined independently for each marker. These parameters were also determined for the combination of the two loci, for which the larger number of mixed samples and the larger number of bands per individual were used, regardless of the locus. The third column, mixed infections with MSP1 + DHFR, represents the percentage of patients who had more than one allele at either locus or both loci.

**FIGURE 2.** Genotype distribution in position 108 of DHFR, according to parasite density and to age of the infant population. Three groups were established based on the parasitemia. (A) shows isolates with fewer than 10,000 parasites/μL, densities of 10,000 to 20,000 parasites/μL, and more than 20,000 parasites/μL. The patients were classified into three age groups (B): under 4 years, 4–8 years, and over 8 years. In both (A) and (B), the conventions are S for serine 108 wild type, T for threonine 108 mutant, A for asparagine 108 mutant, and M for all mixed infections.
and four MSP1 alleles were reported. In our case, four MSP1 variants were detected as well, but the proportion of mixed infections was 83% and complexity of infection was 2. This contrast is even more striking when considering that we used fewer markers (MSP1 and DHFR). The data reported herein on the complexity and percentage of mixed infections are similar to those from mesoendemic zones such as Ndip, Senegal. In the Senegalese study, using three markers (MSP1, MSP2, and GLURP), the proportion of mixed infections was 50–52%, the complexity of infection was 1.65–1.75 fragments/individual for the total population, and 17 MSP1 alleles were reported. In a Thai study, using the same three markers, the proportion of mixed infections was 76%, the complexity of infection was 2.69–3.27, and 10 MSP1 allelic variants were found. A study in Iquitos, Peru, used DHFR and DHPS as markers to look for mutations related to in vitro resistance to pyrimethamine-sulfadoxine; the samples were taken during an epidemic outbreak, and no mixed infections were detected. The authors hypothesized that the simplicity was a result of the epidemic’s arising from very few genetically restricted parasites. This contrasts with our results: a high percentage of mixed infections found in an apparently similar epidemiologic setting. However, very little information is available on the recent history of malaria in Panguí, and there are no clues on the epidemiologic circumstances that led to this outbreak.

On the other hand, inasmuch as the Thr-108 genotype in DHFR was not found when analyzing a large number of samples from different geographic areas, certain authors have suggested that this mutation might occur only in laboratory strains. Nevertheless, the Thr-108 mutation was observed in a high frequency in our study (55%), and it also was found in two Colombian studies in which isolates of different geographic origin were analyzed.

Data also have been reported in which parasitemia is related to the genotype in the 108 position of DHFR. For example, a study in Tanzania on a population of 91 children showed that distribution of the genotype in position 108 varies according to parasite density. The authors explain this by suggesting that pyrimethamine-resistant parasites may be more virulent and produce disease with fewer quantities of parasites, or that the resistant populations grow more slowly and are found in a lesser proportion in natural populations. However, when analyzing the Colombian isolates, we found no statistically significant relationship between parasite density and the presence of one genotype or another in position 108 of DHFR ($\chi^2 = 2.12$). In infections with more than 20,000 parasites per microliter, we detected the Ser-108 genotype in 13% of the cases, while all the samples from Tanzania with this parasitemia exhibited the Ser-108 genotype. Along the same line, all the isolates from Tanzania with very few parasites exhibited the Asn-108 genotype; we detected this genotype in only 20% of the Colombian samples. The previously cited study in Iquitos, Peru, showed that the presence of the DHFR genotypes Asn-108, Ile-51, and Leu-164 combined with DHPS Gly-437, Glu-540, and Gly-581 was highly correlated ($P < 0.01$) with the level of in vivo resistance to pyrimethamine-sulfadoxine.

In the case of infant populations, it has also been suggested that the genotype in position 108 of DHFR is related to the patient’s age. Based on these results, the child’s age has been suggested as a possible factor contributing to confusion in epidemiologic data on DHFR genotypes. For example, a study on the infant population in Nigeria ($N = 131$) found that frequency of the Asn-108 genotype declined with an increase in age. Asn-108 frequency was 60% in the 0–2 age group and 18% among children age 6 and over. However, our analysis of Colombian samples—in contrast to the Nigerian study—did not find a decline in the frequency of Asn-108 with an increase in the age of the group studied. Nor was it possible to establish a statistically significant relationship between children’s ages and presence of the Ser-108 genotype or the presence of mixed infections ($\chi^2 = 8.91$). This indicates that some results reported in the literature cannot be regarded as events of a general nature but appear to be particular to the area and the time of the study.

It is important to bear in mind that our study was conducted with Giemsa-stained thick smears. This overcomes the difficulties inherent in obtaining clinical samples for field studies. The main advantage of using thick smears as a source of DNA is that they are obtained routinely for diagnosing the disease. Furthermore, no special equipment, gathering, or storage conditions are required to extract the DNA. This facilitates the collection of samples in remote areas and allows studies to be developed easily to detect variations in the parasite population over time.

The considerable differences between the data on parasite diversity obtained through our study and those reported for several natural populations of P. falciparum clearly indicate that much remains to be known about the dynamics involved in transmission of the parasite. These may be particular for each population and endemic area, where the combination of different factors such as the genetics and immune response of the host, differential adaptation of the parasite to the vector, preferential distribution of certain vectors, and the genetic diversity of the parasite itself, among other aspects, plays a decisive role in the development of malaria. Therefore, generation of local knowledge on the structure of the parasite population and the genetic variability it contains is of fundamental importance, since it is involved in circumstances such as the pathology of the disease, the acquisition of immunity, and resistance to drugs—and thus the potential for transmission.

Acknowledgments: This work was supported by the Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología ‘Francisco José de Caldas’ Colciencias. Project BID-Colciencias 2104-04170-95, by the Instituto Nacional de Salud and by the Universidad Nacional de Colombia. The authors are indebted to Dr. William Rojas (Corporacion de Investigaciones Biologicas—Medellin) for providing the samples analyzed in this study.

Authors’ addresses: Diana Gómez, Jacqueline Chaparro, Claudia Rubianol, María Orfa Rojas, and Moises Wasserman, Laboratory of Biochemistry, Av. El dorado Cra. 50, Instituto Nacional de Salud, Bogotá, Colombia. Moises Wasserman, LBBIQ, Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia.

Reprint requests: Moises Wasserman, Laboratory of Biochemistry, Instituto Nacional de Salud, Av. Eldorado Cra. 50, Bogotá, Colombia, E-mail: mwasser@ciencias.unal.edu.co

REFERENCES


