Malaria in Humans: *Plasmodium falciparum* Blood Infection Levels Are Linked to Chromosome 5q31-q33

Pascal Rihet, Yves Traoré, Laurent Abel, Christophe Aucan, Thérèse Traoré-Leroux, and Francis Fumoux

1Institut National de la Santé et de la Recherche Médicale (INSERM) U 399, Immunology and Genetics of Parasitic Diseases, Marseille; 2Centre Muraz, Organisation de Coordination et de Coopération contre les Grandes Endémies, Bobo-Dioulasso, Burkina Faso; and 3INSERM U 436, Mathematical and Statistical Modeling in Biology and Medicine, Hospital Pitié-Salpêtrière, Paris

Summary

*Plasmodium falciparum* malaria remains a major cause of morbidity and mortality in many tropical countries, especially those in sub-Saharan Africa. Human genetic control of malaria infection is poorly understood; in particular, genes controlling *P. falciparum* blood infection levels remain to be identified. We recently evidenced the existence of complex genetic factors controlling blood infection levels in an urban population living in Burkina Faso. We performed, on 153 sibs from 34 families, sib-pair linkage analyses between blood infection levels and chromosome 5q31-q33, which contains numerous candidate genes encoding immunological molecules. Our results, obtained by means of the two-point Haseman-Elston (HE) method and a nonparametric (NP) approach, show linkage of parasitemia to D5S393 (P = .002) and D5S658 (P = .0004). Multipoint analyses confirmed linkage, with a peak close to D5S658 (P = .0013 and P = .0007 with the HE and NP methods, respectively). The heritability of the locus was .48, according to the two-point results, and .43, according to the multipoint results; this indicates that its variation accounted for ~45% of the variance of blood infection levels and that the locus plays a central role in the control of parasitemia. The identification of the gene is, therefore, of major interest in understanding the mechanisms controlling *P. falciparum* parasitemia.

Introduction

*Plasmodium falciparum* malaria (MIM 109270, 110750, 305900, 111300, 141900, 141800, 142830, 146631, 248310, and 601798) is a major cause of morbidity and mortality in many developing countries, especially in sub-Saharan Africa, where childhood mortality is ~2,000,000/year; worldwide, one-third of humanity is at risk (Butler et al. 1997). Both host and parasite genetic factors are thought to determine the outcome of malaria infection. On one hand, parasite strains differ in virulence (Mendis and Carter 1995), and, on the other hand, the influence of host genetic factors has been demonstrated in experimental models (Stevenson et al. 1982; Fortin et al. 1997). There is accumulating evidence for genetic control of malaria infection in humans. Genetic factors have been shown to control malaria disease (Hill 1996) and blood infection levels (Abel et al. 1992; Garcia et al. 1998; Rihet et al. 1998) and to regulate antimalarial immune responses (Stjøberg et al. 1992; Jepson et al. 1997a).

Case-control studies have detected associations between severe malaria and major histocompatibility complex genes and genes encoding red blood cell proteins (Hill et al. 1991; McGuire et al. 1994; Ruwende et al. 1995). However, the extent to which these genes influence the outcome of malaria infection is unclear, and important, as-yet-unknown genes, including genes controlling infection levels, are likely to be identified (Miller 1996).

Alternative approaches to the genetics of malaria in humans are segregation and linkage analyses. In one segregation analysis, a predominant recessive gene controlling blood infection levels was detected (Abel et al. 1992). However, more-recent segregation analyses showed the existence of complex genetic factors controlling blood infection levels and did not evidence a single predominant gene (Garcia et al. 1998; Rihet et al. 1998). The analyses revealed a strong interaction between genetic factors and age; genetics-related differences were much more prominent in children than in adults. Further nonparametric (NP) linkage analyses,
such as sib-pair analyses that do not require knowledge of the genetic model and that focus on children, seem to be helpful in investigating the genetic control of blood infection levels.

In a candidate-region approach to the human genetics of *P. falciparum* infection levels, chromosomal regions that contain genes involved in immune responses are of major interest. The chromosome 5q31-q33 region contains numerous candidate genes encoding immunological molecules such as cytokines, growth factors, and growth-factor receptors (Chandrasekharappa et al. 1990; Saltman et al. 1993), which are involved in the control of immunity to *P. falciparum* blood stages (Troye-Blomberg et al. 1994). Moreover, the importance of this region in immune regulation is highlighted by its linkage to plasma immunoglobulin E (IgE) levels (MIM 147061) (Marsh et al. 1994; Meyers et al. 1994), bronchial hyperresponsiveness (MIM 600807) (Postma et al. 1995), and schistosomiasis infection (MIM 181460) (Marquet et al. 1996; Müller-Myhsok et al. 1997). Therefore, we performed sib-pair linkage analyses of *P. falciparum* blood infection levels in a suburban area in Burkina Faso, and we present here evidence for linkage to chromosome 5q31-q33.

**Material and Methods**

**Families**

The study subjects live in an urban district of Bobo-Dioulasso, the second largest town in Burkina Faso, in an area in which malaria is endemic (Robert et al. 1986) and in which *P. falciparum* transmission occurs only during the rainy season (August–December). The family structures and the area of parasite exposure have been described extensively elsewhere (Rihet et al. 1998). Most of the individuals belong to the Mossi ethnic group (50%); the other ethnic groups include the Dafing (19%), Guian (5%), Bissa (15%), Samogo (1%), Bobo (6%), and Nounouma (4%). The Mossi and the Bissa are originally from the center of Burkina Faso, and the other groups are from the western part of the country. The families have been established there for >20 years, and the living habits of the different ethnic groups are very similar. The whole population volunteered to participate in the study, and all participants were clearly informed of the objective and the protocol. This protocol was approved by the national medical authorities of Burkina Faso.

From the 41 pedigrees studied in our previous segregation analysis, 30 informative pedigrees, corresponding to 34 nuclear families and containing at least three available sibs each, were selected for genotyping. Incompatibilities with Mendelian inheritance of marker alleles led us to detect incorrectly specified familial relationships. Nine subjects that failed to share an allele, at the marker locus, with one of their supposed parents were excluded. After exclusion of the nine subjects, all marker-allele segregation patterns were consistent with Mendelian transmission. A final total of 212 subjects (59 parents and 153 full sibs) were available for geno-

![Figure 1](link-to-image)
typing and were retained for linkage analysis. The number of families with 2, 1, and 0 genotyped parents was 27, 5, and 2, respectively. The 34 families led to a total of 285 possible sib pairs, with the following distribution of sibship sizes: 7, 8, 15, 3, and 1 sibships contained 3, 4, 5, 6, and 7 sibs, respectively. The mean age of the sibs was 12.1 ± 6.2 years (range 1–34 years).

### Determination of Blood Infection Levels

The determination of blood infection levels was described in our previous study (Rihet et al. 1998). In brief, each family was visited 20 times during the 21 mo of the study (April 1994–December 1995), and the mean number of parasitemia measurements per subject was 12.8 ± 5.1 (range 1–20). Peripheral blood samples were drawn from fingers of all family members present, and thick and thin blood films were stained with Giemsa stain. The parasite determination and enumeration were established blindly from two independent readings. More than 95% of parasites identified on thin blood smears were *P. falciparum*; the others were *P. malaria*. Only *P. falciparum* asexual forms were retained, to determine parasitemia. The parasite enumeration on thick and thin blood films showed a strong correlation ($r = .85; P < 10^{-5}$), and the parasite density (PD) was defined as the number of parasitized erythrocytes per microliter observed in thin blood films. The analysis was conducted on a logarithmic transformation of PD adjusted for seasonal transmission and for covariates that showed a significant effect on parasitemia. As reported elsewhere (Rihet et al. 1998), the parasitemia was not associated with sex ($P > .75$), area of parasite exposure ($P > .5$), ethnic group ($P > .2$), ABO blood group or Rh blood group ($P > .2$), or hemoglobin genotype ($P > .5$); in contrast, age was correlated strongly with parasitemia ($P < 10^{-4}$) and was retained for data adjustment. The standardized residual, termed “MAPDS” (mean adjusted parasite density standardized), was the phenotype used for linkage analysis. Blood infection level was determined, for the same population, during two periods of malaria transmission, so the MAPDS was calculated independently for the 2 years, and a strong individual correlation ($r = .75; P < .0001$) was observed. At the individual level, the blood infection phenotype is remarkably constant for the 2 consecutive years.

### Genotyping

DNA microsatellite analysis was performed according to the method described by Vignal et al. (1993), with DNA extracted from mononuclear cells separated by Ficoll-Hypaque density gradient. The DNA from the M134702 cell line was used as reference. The PCR primers (GENSET) consisted of 10 Genethon highly polymorphic markers of the 5q31-q33 region: D5S642 (Genome Database [GDB] accession number 199775), D5S2117 (GDB 614859), D5S393 (GDB 187866), D5S399 (GDB 187987), D5S658 (GDB 200268), D5S436 (GDB 188734), D5S2090 (GDB 613731), D5S636 (GDB 199700), D5S2012 (GDB 609486), and D5S487 (GDB 189396) (Dib et al. 1996) (fig. 1). Genotypes were determined from two independent readings of each autoradiograph.

### Linkage Analyses

Two-point and multipoint sib-pair linkage analyses were performed by the program MAPMAKER/SIBS, version 2.0, with use of two of the methods proposed, in this software, for analysis of quantitative traits. The first method was the Haseman–Elston (HE) (Haseman and Elston 1972) approach, and, for reasons discussed by Kruglyak and Lander (1995b), we used the expectation-maximization algorithm to perform regression of the squared phenotypic difference between two sibs on the actual distribution of the number of alleles shared identical by descent by these two sibs. The significance level

<table>
<thead>
<tr>
<th>MARKER</th>
<th>β</th>
<th>t-Statistic</th>
<th>P</th>
<th>Z Score</th>
<th>P</th>
<th>MARKER</th>
<th>β</th>
<th>t-Statistic</th>
<th>P</th>
<th>Z Score</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5S642</td>
<td>−.30</td>
<td>−1.02</td>
<td>.1</td>
<td>.94</td>
<td>.1</td>
<td>D5S642</td>
<td>−.26</td>
<td>−.88</td>
<td>.1</td>
<td>.85</td>
<td>.1</td>
</tr>
<tr>
<td>D5S2117</td>
<td>−.23</td>
<td>−.82</td>
<td>.1</td>
<td>.08</td>
<td>.5</td>
<td>D5S2117</td>
<td>−.24</td>
<td>−.84</td>
<td>.1</td>
<td>−.02</td>
<td>.5</td>
</tr>
<tr>
<td>D5S393</td>
<td>−.84</td>
<td>−2.94</td>
<td>.002</td>
<td>2.39</td>
<td>.008</td>
<td>D5S393</td>
<td>−.83</td>
<td>−2.90</td>
<td>.002</td>
<td>2.37</td>
<td>.008</td>
</tr>
<tr>
<td>D5S399</td>
<td>−.44</td>
<td>−1.58</td>
<td>.058</td>
<td>1.90</td>
<td>.029</td>
<td>D5S399</td>
<td>−.32</td>
<td>−1.89</td>
<td>.030</td>
<td>2.04</td>
<td>.020</td>
</tr>
<tr>
<td>D5S658</td>
<td>−.96</td>
<td>−3.41</td>
<td>.0004</td>
<td>2.54</td>
<td>.005</td>
<td>D5S658</td>
<td>−.86</td>
<td>−3.04</td>
<td>.001</td>
<td>2.41</td>
<td>.008</td>
</tr>
<tr>
<td>D5S436</td>
<td>−.18</td>
<td>−.61</td>
<td>.2</td>
<td>.65</td>
<td>.2</td>
<td>D5S436</td>
<td>−.11</td>
<td>−.39</td>
<td>.3</td>
<td>.59</td>
<td>.2</td>
</tr>
<tr>
<td>D5S2090</td>
<td>−.34</td>
<td>−1.21</td>
<td>.1</td>
<td>1.21</td>
<td>.1</td>
<td>D5S2090</td>
<td>−.31</td>
<td>−1.08</td>
<td>.1</td>
<td>1.15</td>
<td>.1</td>
</tr>
<tr>
<td>D5S636</td>
<td>−.31</td>
<td>−1.10</td>
<td>.1</td>
<td>.19</td>
<td>.4</td>
<td>D5S636</td>
<td>−.34</td>
<td>−1.20</td>
<td>.1</td>
<td>.24</td>
<td>.3</td>
</tr>
<tr>
<td>D5S2012</td>
<td>−.13</td>
<td>−.45</td>
<td>.3</td>
<td>.64</td>
<td>.2</td>
<td>D5S2012</td>
<td>−.06</td>
<td>−.22</td>
<td>.4</td>
<td>.53</td>
<td>.2</td>
</tr>
<tr>
<td>D5S487</td>
<td>−.21</td>
<td>−.72</td>
<td>.2</td>
<td>1.35</td>
<td>.1</td>
<td>D5S487</td>
<td>−.21</td>
<td>−.72</td>
<td>.2</td>
<td>1.35</td>
<td>.1</td>
</tr>
</tbody>
</table>
of the test was assessed by use of a one-sided Student’s $t$-test, for which the number of df was calculated by computing, as proposed in the S.A.G.E. software (1994), an effective sample size equal to $\sum (s_i - 1)$, where the summation is over the number of families ($34$, in the present study), and $s_i$ is the number of full sibs in the $i$th family; consequently, the number of df, for the present study, was $116$. The second method was the NP quantitative trait locus analysis proposed by Kruglyak and Lander (1995b), which is based on a rank statistic robust to violations of normality distributions assumed by the HE approach. This analysis resulted in a $Z$ score asymptotically distributed as a standard normal deviate.

Allele frequencies for the $10$ markers were estimated from our data, since some alleles found in this African population were not found in the CEPH reference families. The analysis was also performed considering equal allele frequencies, to assess the influence of these frequencies on the analysis; note that only $9$ of $68$ parents were missing.

**Results**

We genotyped $212$ subjects, from $34$ nuclear families, for $10$ polymorphic markers in the $5q31$-$q33$ region. A total of $153$ sibs providing $285$ sib-pairs were analyzed, and marker data were available for $59$ of $68$ parents. Figure 2 presents the segregation of three markers and the MAPDS values for three families, and results of two-point analyses are shown in table 1. For estimated allele frequencies, the HE method yielded $P$ values of $.002$, $.058$, and $.0004$ for marker loci $D5S393$, $D5S399$, and $D5S658$, respectively. Close results were obtained with an NP method of linkage that does not require assumptions about the distribution of the phenotypic effects (table 1). Equal allele frequencies weakly influenced the $P$ values and led to the same conclusions. The seven remaining markers of the region showed no evidence of linkage ($P > .1$).

Multipoint analysis over the whole region (fig. 3) resulted in a maximum $t$-statistic (HE method) of $3.08$ ($P = .0013$) and a maximum $Z$ score (NP method) of $3.18$ ($P = .00013$). The two analysis methods provided peaks very close to the $D5S658$ marker. Equal allele frequencies led to very close curves (data not shown), with a maximum $t$-statistic of $2.99$ ($P = .0016$) and a maximum $Z$ score of $3.13$ ($P = .0008$). This indicates that inaccurate allele frequencies were unlikely to have influenced our linkage results. Furthermore, the information content of the inheritance pattern at each point of the $5q31$-$q33$ region, computed as described by Kruglyak et al. (1996), was high and within the range $85\%$-$97\%$ ($>90\%$ between $D5S393$ and $D5S658$).

On the basis of HE regression slopes, the heritability due to the locus (assuming a zero recombination fraction with $D5S658$) was $.48$ (95% confidence interval [CI] $0.20$-$0.76$) and $.43$ (95% CI $0.15$-$0.71$), according to two-point and multipoint results, respectively. This indicates that variation of the locus accounted for $\sim45\%$ of the variance of MAPDS.

**Discussion**

In the present context of a candidate-gene approach, our results provide strong evidence for the presence, on chromosome $5q31$-$q33$, of a locus, named “Pfiii” (P.
Figure 3    Multipoint quantitative trait linkage analyses of *P. falciparum* infection intensities (MAPDS). A, Results of HE quantitative trait linkage regression analysis. B, Results of NP quantitative trait linkage analysis. The horizontal axes represent the genetic distance, in centimorgans, according to genetic data from the CEPH pedigrees (Dib et al. 1996). The linkage analyses were based on marker allele frequencies calculated from the study sample.

*falciparum* infection level 1), controlling blood *P. falciparum* infection levels. Identification of genes in this region would be extremely helpful in understanding the mechanisms controlling malaria infection in humans. Since high blood infection level is an important factor in the pathogenesis of malaria (Miller et al. 1994), *Pf1l* is also a candidate gene for malarial disease.

In a small Cameroonian family sample (26 sibling pairs from nine nuclear families), Garcia et al. (in press) recently observed a trend in favor of linkage between *P. falciparum* parasitemia levels and a 5q31-q33 microsatellite marker (D5S636); however, the Cameroonian sample was too small to allow detection of significant linkage. Further studies are required, to determine whether, in African populations with various genetic backgrounds and living in different *P. falciparum* pressure areas, blood infection levels are under the control of the same locus; meta-analysis of several studies may
be useful in assessment of the importance of chromosome 5q31-q33 in malaria. The data from segregation analyses (Garcia et al. 1998; Rieth et al. 1998) nevertheless support the idea that other loci may also influence parasitemia. Human genes associated with severe malaria, such as HLA, TNFα, and G6PD (Hill et al. 1991; McGuire et al. 1994; Ruwende et al. 1995), or loci syntenic to mouse chromosomes 8 and 9 loci controlling P. chabaudi parasitemia (Foote et al. 1997; Fortin et al. 1997) should be tested in a candidate-gene approach. A further genomewide search may be useful in identification of other loci controlling blood parasitemia; mild malaria, which was recently linked to HLA (Jepson et al. 1997b); severe malaria; or immune response to P. falciparum antigens.

Our findings have important implications for understanding protective immune and physiopathological mechanisms, since the 5q31-q33 region contains several candidate genes implicated in the regulation of the immune responses to Plasmodium species and in malaria pathogenesis. These include genes involved in the TH1 and TH2 subset balance, such as interleukin (IL)–4 (IL4) (MIM 1477780), IL12 (MIM 161561), and interferon (IFN) regulatory factor 1 (IRF1) (MIM 147575) (Abbas et al. 1996). The production of anti–P. falciparum antibodies implicated in human antimalarial immunity is under the control of IL4 (Troye-Blomberg et al. 1990; Bouharoun-Tayoun et al. 1995). Injection of r-IL12 increases blood γIFN level and confers a sterile protection against P. cynomolgi in monkeys (Hoffman et al. 1997). Moreover, γIFN, the transcription of which is regulated by IRF1, activates neutrophils and macrophages to destroy P. falciparum (Ferrante et al. 1990; Bouharoun-Tayoun et al. 1995) and displays a direct parasite-killing activity (Naotune et al. 1991). The 5q31-q33 region contains other genes involved in the immune response, such as IL3 (MIM 147740), granulocyte-macrophage colony-stimulating factor (CSF2) (MIM 138960), IL9 (MIM 146931), IL13 (MIM 147683), β2-adrenergic receptor (ADRB2) (MIM 109690), and macrophage CSF1 receptor (CSF1R) (MIM 164770). IL3, CSF2, and CSF1, the CSF1R ligand, promote the growth and differentiation of neutrophils and macrophages (Metcalf 1991) that are toxic for asexual blood stages. IL9, IL13, and ADRB2 were close to the peaks provided by linkage analyses (fig. 1) and seemed of particular interest. IL9 is a T-cell growth factor; correlations between IL9 production and TH2 responses have been shown in vitro and in vivo (Nicolaides et al. 1997). IL13, which is essentially produced by TH2 cells, is involved in the production of antibodies and contributes to down-regulation of macrophage functions (De Vries 1996). ADRB2 is involved in TH1 and TH2 balance (Ramer-Quinn et al. 1997).

However, our results did not allow us to determine the exact location of the postulated gene or to exclude any of the candidate genes of the 5q31-q33 region (fig. 1). This is a common problem in linkage studies of complex traits (Lander and Shork 1994), especially when NP approaches, such as sib-pair methods, are used. It is difficult to localize a predisposing locus through sib-pair analyses because the allele-sharing proportion can fluctuate considerably and may not attain its maximum at the locus (Kruglyak and Lander 1995a). Polymorphism analysis of candidate genes and linkage disequilibrium can offer a powerful complement to traditional linkage studies and may play an important role in the identification of the gene.

The 5q31-q33 region was previously linked to a locus controlling the intensity of schistosoma mansoni susceptibility/resistance (Marquet et al. 1996; Müller-Myhsok et al. 1997), which suggests that resistance/susceptibility genes in this region may influence the outcome of different infectious diseases. Moreover, this region was also linked to several loci related to atopy (Marsh et al. 1994; Meyers et al. 1994; Postma et al. 1995) and might contain genes of critical importance in various pathological situations.

**Acknowledgments**

We acknowledge all volunteer families of Sarfalao, for their contributions, and the entomological team of Centre Muraz (Bobo-Dioulasso), for valuable technical assistance. We thank the medical authority of Burkina Faso for encouragement during the course of this work. We also thank Dr. Dominique Hillaire, for helpful technical advice, and Drs. Alain Bourgois and Luc Reininger, for critical readings of the manuscript. This work was supported by research grants from the French Ministry of Cooperation and Development and from the Association des Universités Partiellement ou Entièrement de Langue Française–Union des Réseaux d’Expression Française (AUPELF-UREF). Y.T. is supported by a studentship from the Ministry of Cooperation and Development and AUPELF-UREF, and C.A. is supported by a studentship from the Ministry of Research and Technology (France).

**Electronic-Database Information**

Accession numbers and URLs for data in this article are as follows:

Genome Database, http://www.gdb.org (for polymorphic markers used in genotyping [199775, 614859, 187866, 187987, 200268, 188734, 613731, 199700, 609486, and 189396])

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for P. falciparum (MIM 109270, 110750, 305900, 111300, 141900, 141800, 142830, 146631, 248310, and 613731), IgE levels [MIM 147061], bronchial hyperresponsiveness [MIM 600807], schistosoma mansoni susceptibility/resistance [MIM 609486, 181460], IL3 [MIM 147740], IL4 [MIM 147780], IL9 [MIM 146931], IL12 [MIM 161561], IL13 [MIM 147683].)
IRF1 [MIM 147575], CSF2 [MIM 138960], CSF1R [MIM 164770], and ADRB2 [MIM 109690])

References


Hill AVS (1996) Genetic susceptibility to malaria and other infectious diseases: from the MHC to the whole genome. Parasitology Suppl 112:S75–S84


Rihet et al.: Genetic Control in Malaria