LINKAGE ANALYSIS OF BLOOD PLASMODIUM FALCIPARUM LEVELS: INTEREST OF THE 5q31-q33 CHROMOSOME REGION

ANDRE GARCIA, SANDRINE MARQUET, BRUNO BUCHETON, DOMINIQUE HILLAIRE, MICHEL COT, NADINE FIEVET, ALAIN J. DESSEIN, AND LAURENT ABEL

Institut Francais de Recherche Scientifique pour le Developpement en Cooperation (ORSTOM), Organisation de Coordination pour la Latte Contre les Enemies en Afrique Centraie (OCEAC), Yaoundé, Cameroon: Institut National de la Sante et de la Recherche Medicale (INSERM) Unite 436, Mathematical and Statistical Modeling in Biology and Medicine, Hospital Pitie-Salpetriere, Paris, France; INSERM Unite 399, Immunology and Genetic of Parasitic Diseases/Laboratory of Parasitology-Mycology, Faculty of Medicine, Marseille, France

Abstract. There is accumulating evidence for the involvement of genetic factors in the human response to malaria infection, mostly based on results obtained in studies of severe clinical malaria. The role of major gene(s) controlling blood parasitemia levels in human malaria has also been detected by means of segregation analysis. To confirm and to localize such gene(s), we performed a sib-pair linkage analysis investigating the role of five candidate chromosomal regions: 6p21 (HLA-tumor necrosis factor region), 2q13-q21 (genes coding for interleukin-1 α and β), 14q11 (locus coding for the α chain of T cell antigen receptor), 7q35 (gene cluster for the β subunit of T cell receptor), and 5q31-33, which includes several candidate genes and was recently linked to a locus controlling infection levels by Schistosoma mansoni, denoted as SM1. The analysis was carried out on nine families from a southern Cameroon village, and the phenotype under study was blood infection levels with Plasmodium falciparum. No linkage was found with any of the four markers outside the 5q31-q33 region. A trend in favor of linkage was observed in the distal part of the 5q31-q33 region, especially with the marker D5S636 (P < 0.05 using the Monte Carlo P value), which was the marker that provided the highest evidence for linkage with SM1. These results suggest that a locus influencing P. falciparum levels in malaria could be located in the same genetic region as that containing SM1, indicating that the 5q31-q33 region may be critical in the control of different parasite infections.

Malaria is a major cause of morbidity and mortality in tropical countries, especially in young children. The profound influence that the genetic makeup of the host has on resistance to malaria has been established in numerous animal studies, and there is also accumulating evidence for the involvement of such genetic factors in the human response to malaria infection. This genetic control can be investigated in humans through different malaria related phenotypes such as clinical phenotypes (e.g., severe malaria), immunologic phenotypes (e.g., levels of immune response induced by malaria antigens), or parasitologic phenotypes (e.g., levels of infection). Numerous population studies have focused on severe malaria and supported the important protective role of several genetic disorders of the red blood cell, such as abnormal hemoglobins, or glucose-6-phosphate dehydrogenase deficiency, and of certain HLA antigens. Similarly, homozygotes for a variant of the tumor necrosis factor-α (TNF-α) region were found to have an increased risk of cerebral malaria independently of their HLA alleles, and the functional consequence of this variant upon TNF-α gene expression has been recently established. More conflicting results exist with respect to the direct implication of HLA genes in the genetic regulation of immune responses induced by malaria vaccine antigens. Different studies have investigated the role of genetic factors in parasitologic phenotypes. It was demonstrated that erythrocytes of west Africans who are Duffy blood group negative cannot be infected by Plasmodium vivax. A segregation analysis provided evidence for the presence of a recessive major gene controlling malaria blood infection levels in African families mostly infected by P. falciparum, and indicated that about 20% of the population was predisposed to high parasitemia. Three of us (AG, MC, and LA) recently carried out another segregation analysis of blood infection levels in Cameroonian families exposed to higher vectorial transmission than the subjects of the previous analysis. The results confirmed the existence of complex genetic factors controlling parasitemia levels but were not consistent with the parent-offspring transmission of a single gene. This study also showed the dramatic effect of age on infection levels indicating that genetic-related differences are easier to detect in children than in adults.

The aim of the present study was to further investigate the genetic control of malaria blood infection levels in this Cameroonian population by linkage studies that facilitate the genetic dissection of complex phenotypes. With respect to the results of segregation analysis, we focused on young subjects and used a nonparametric method of linkage analysis, i.e., sib-pair analysis, which does not need to specify the mode of inheritance of the phenotype under study (the parasitemia levels). Five chromosomal regions containing genes that could be involved in the response against malaria infection were considered in the present study, including in particular the 5q31-q33 region, which has recently been linked to a locus controlling the intensity of infection by the parasite Schistosoma mansoni.

SUBJECTS AND METHODS

Family data and phenotype measurements. Family subjects lived in a small southern Cameroon village located in the tropical rain forest and were described previously. The population was clearly informed of the purpose of the study, and the protocol of the study, including blood samples, was approved by traditional (chef and village comity), local, and national government authorities (Public Health Ministry). From the family sample used in our previous segregation analysis, which included 44 pedigrees, 13 informative families were selected for genotyping. After refusal of four families, DNA was obtained from 71 individuals be-
longing to nine families. Since sib-pair method was used in this study, we retained for linkage analysis nuclear families with at least two available sibs and also one genotyped parent. After exclusion of five subjects, all marker allele segregation patterns were consistent with Mendelian transmission.

The determination of malaria infection intensities has been detailed previously12,13 and will be briefly summarized. During the one year follow-up, each family was visited every two months and parasitologic measurements were performed at each visit from thick blood smears. To study the control of infection levels by \textit{P. falciparum} alone, blood samples that exhibited other \textit{Plasmodium} species (20% of the positive blood samples) were removed from the analysis. Further analyses were conducted on a logarithmic transformation of the \textit{P. falciparum} density observed in 150 fields of thick blood films. To deal with a unique variable accounting for the intensity of global malaria infection, a mean \textit{P. falciparum} density (MPFD1) was determined. Since the parasite densities varied significantly across visit times with highest values during the rainy season, the MPFD1 was computed for each subject as the mean of his visit-adjusted parasite densities. Parasite densities were not significantly different according to hemoglobin genotype, and among five risk factors (gender, age, prophylaxis intake, use of protection against mosquito, area of residence) that were tested as potentially influencing the MPFD1 values, age was the only covariate that showed a significant effect.13 Therefore, MPFD1 values were adjusted on age by means of the polynomial regression estimated from the whole sample, and the standardized residuals, denoted as MPFD2, are the phenotypes that were used for the linkage analysis.

**Investigated chromosomal regions and genotyping.** The potential role of five regions containing genes that could be involved in the response against malaria infection were investigated by linkage analysis with polymorphic markers: 1) The HLA-TNF region (6p21) through the Genethon (Evry, France) microsatellite marker D6S276;17 2) the 2q13-q21 region including genes coding for interleukin-1 (IL-1) α and β through an IL-1α intragenic marker;17 3) the locus in 14q11 coding for the alpha chain of T cell antigen receptor through an intragenic T cell receptor (TCRα) marker;19 4) the gene cluster for the beta subunit of T cell receptor on chromosome 7q35 through an intragenic TCRβ marker;20 5) the last region was 5q31-q33, which contains several candidate genes as those coding for the granulocyte–macrophage colony-stimulating factor (CSF2), IL-3, IL-4, IL-5, IL-9, the immune regulatory factor 1 (IRF1), and the colony-stimulating factor 1 receptor (CSF1-1R). Six markers of this region were typed for the present analysis including four Genethon microsatellite markers (D5S393, D5S434, D5S436, and D5S636)17 and two intragenic markers IL-921 and CSF1R.21 The map of region 5q31-q33 is shown in Figure 1.

The DNA was extracted using standard methods and microsatellite analysis was performed as previously described.22 Polymerase chain reaction primers were obtained from Genethon,17 except for IL-1α,18 TCRα,19 TCRβ,20 IL-9,21 and CSF1R.21 Genotypes were determined from two independent readings of each autoradiograph.

**Statistical methods.** Linkage of blood parasitemia levels with the five genotyped regions was investigated using a nonparametric approach since no simple monogenic model could be inferred from our previous segregation analysis in this population.13 We used the sib-pair method for quantitative traits originally proposed by Haseman and Elston,23 as implemented in the SIBPAL software from the computer package Statistical Analysis for Genetic Epidemiology.24 This method uses the regression of the squared sib-pair difference for MPFD2 values on the estimated proportion of marker alleles shared identical by descent by sib-pairs. The statistic for testing linkage, \( T \), is the estimated regression coefficient divided by its standard error, which has for large samples approximate normal distribution. In the present case of a small sample size, a Student’s \( t \)-distribution was used for \( T \) to determine significance levels. The number of degrees of freedom (df) for this distribution was determined as performed in SIBPAL by computing an effective sample size taking into account the covariances between sib-pairs coming from the same sibship. In the present study including only full-sib pairs, the effective sample size was computed as \( \sum s_i - 1 \) where the summation is over the number of families (nine in the present study), \( s_i \) is the number of full sibs in the ith family, and the number of dfs was equal to 13. When \( T \) was close to significant level, the \( P \)-value of the test was also computed using Monte Carlo methods by simulating and analyzing 100,000 replicates of our family sample under the null hypothesis of no linkage.25 In this method, the Monte Carlo \( P \) value, \( P_{\text{MC}} \), is the proportion of replicates providing a value of the \( T \) statistic more extreme than the \( T \) value obtained in the actual data. Furthermore, significance levels were adjusted for multiple comparisons using the Bonferroni correction, and since five independent regions were tested in this analysis, a type I error of 0.01 was considered.

The sib-pair analysis was dependent on marker allele frequencies for the families with one missing parent. These frequencies were estimated from our data since some alleles found in this African population were not referred in the
Centre d’Etude du Polymorphisme Humain (Paris, France) reference families. The analysis was also performed considering equal marker allele frequencies.

RESULTS

The final sample for linkage analysis is shown in Figure 2, and consisted in nine nuclear families of which five had both genotyped parents and four had one genotyped parent. The nine sibships included 25 children and represented a total of 26 possible sib-pairs. These 25 children had a mean age of 11.2 years (range = 1–31 years), and MPFD2 values ranging from 1.25 to 2.97 with a mean (SD) of 0.66 (1.10). The nine sibships included at least one child with a MPFD2 value over the 70th percentile, of whom six were over the 90th percentile, a design that increased the power of the analysis.26

No linkage was found with any of the four markers outside the 5q31-q33 region: D6S276 (P > 0.6), IL1α (P > 0.6), TCRα (P > 0.5), and TCRβ (P > 0.5). The results observed with the markers of the 5q31-q33 region are shown in Table 1. No significant evidence of linkage at the 0.01 level was obtained. However, a trend in favor of linkage was observed in the distal part of this region with in particular D5S636 (P < 0.06). The use of equal marker allele frequencies led to similar results. The Monte Carlo P value computed for D5S636 was smaller than the one obtained under the assumption of a Student’s t distribution with 13 dfs, and equal to 0.045. Figure 2 shows the D5S636 genotypes in the nine analyzed families.

DISCUSSION

Although no significant result at the 0.01 level was obtained, our study shows that 5q31-q33 is a region of major interest in the investigation of the genetic control of Plasmodium falciparum levels. If linkage with this region does exist, the significance levels obtained with our small sample indicate that strong significant results could be observed with a larger sample of a quite reasonable size. Furthermore, the significance value obtained for D5S636 by the Monte Carlo simulation, which took into account the structure of the present data, support the validity of the sib-pair analysis. Interestingly, D5S636 was also the marker that provided the most significant result in our previous linkage analysis demon-

---

**Figure 2.** Nine nuclear families used for linkage analysis. Age and D5S636 allele numbers are indicated for all known subjects; unknown subjects are indicated by a ?. Mean Plasmodium falciparum density (MPFD2) values are provided only for children (parental values are not used in sib-pair analysis).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Estimated allele frequencies</th>
<th>Equal allele frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T statistic</td>
<td>P†</td>
</tr>
<tr>
<td>IL-9</td>
<td>-0.09</td>
<td>0.54</td>
</tr>
<tr>
<td>D5S393</td>
<td>-0.66</td>
<td>0.26</td>
</tr>
<tr>
<td>D5S436</td>
<td>-0.01</td>
<td>0.49</td>
</tr>
<tr>
<td>D5S434</td>
<td>-1.40</td>
<td>0.09</td>
</tr>
<tr>
<td>CSF1R</td>
<td>-1.31</td>
<td>0.11</td>
</tr>
<tr>
<td>D5S636</td>
<td>-1.68</td>
<td>0.06</td>
</tr>
</tbody>
</table>

* IL-9 = interleukin-9; CSF1R = colony stimulating factor 1 receptor.

† P values were computed under the assumption of a Student’s t-distribution with 13 degrees of freedom for the T statistic.
strating that a locus controlling infection intensities by *S. mansoni*, denoted as SM1, was located in the 5q31-q33 region.\textsuperscript{15}

The presence of genetic factors involved in the control of malaria infection intensities was demonstrated by both findings in mice that showed the role of host genetic factors in the outcome of experimental malaria infections,\textsuperscript{1} and a segregation analysis performed on families from Cameroon.\textsuperscript{12} A more recent segregation analysis, conducted in a larger sample of families from the village in which the present linkage study took place, provided evidence for a more complex genetic model.\textsuperscript{13} However, this study showed the strong effect of age on infection levels and its interaction with a putative gene, indicating that genetic-related differences are much more important in young subjects than in adults and that a sib-pair study was the more appropriate design to perform linkage analysis. Although the role of genetic red blood cell defects and of certain HLA alleles against severe forms of malaria have been documented, discordant results exist concerning their influence on the levels of infection intensities.\textsuperscript{27, 28} This observation could be related to the fact that most suspected severe malaria resistance alleles, within the MHC region as well as those related to red blood cell genetic disorders, confer protection against severe forms, and do not seem to have a major influence on blood infection levels. Our results are consistent with this view by showing no significant linkage between parasitemia levels and the HLA region suggesting that a different genetic region in 5q31-33 may influence blood infection levels.

The 5q31-33 region contains several genes encoding for molecules that play important functions in the regulation of the immune response to pathogens. In particular, a cluster of candidate genes is located near D5S393 (Figure 1), including genes coding for the CSF2, IL-3, IL-4, IL-5, IL-13, and IRF1 that regulates interferon \( \gamma \) (IFN-\( \gamma \)) transcription. A large number of observations indicated that IFN-\( \gamma \) is critical in immunity against intracellular pathogens.\textsuperscript{29} Among candidates that map in the distal part of 5q31-q33 is IL12p40, which encodes the \( \beta \) chain of IL-12, and IL-12 has been shown to protect monkeys against *P. cynomolgi* induced infection.\textsuperscript{30} Finally, in addition to the linkage with SM1, the 5q31-q33 region has also been linked with a locus controlling bronchial hyperresponsiveness in asthma,\textsuperscript{31} and a locus regulating IgE levels.\textsuperscript{32, 33} Although high serum levels of total IgE as well as of anti-*P. falciparum* IgE antibodies in malaria-exposed individuals have been reported,\textsuperscript{33, 34} the significance of IgE for protection and/or pathogenesis in malaria remains to be established.\textsuperscript{35}

In conclusion, our study suggests that a locus influencing *P. falciparum* levels could be located in the same genetic region as that containing SM1, which regulates the intensity of infection by *S. mansoni*, indicating that the 5q31-q33 region may be critical in the control of different parasitic infections. Ongoing studies will investigate further the role of this region both in malaria on a larger population living in different vectorial transmission conditions and in other parasitic diseases.

Acknowledgments: Andre Garcia and Sandrine Marquet contributed equally to this work. We are deeply grateful to S. Cot, H. Dekker, G. Le Goff, J. Y. Le Hesran, E. Mommers, G. Mbang, F. Nkoubele, P. Richard, and V. Robert for assistance.

Financial support: Data were collected in southern Cameroon with financial and technical support of ORSTOM and Agence Franco-phone Pour l’Enseignement Supérieur et la Recherche (AUPELF-UREF).

Authors’ addresses: Andre Garcia, ORSTOM, Institut Pierre Richet, BP 1500, Bouake, Côte d’Ivoire. Sandrine Marquet, Bruno Bucheton, Dominique Hillaire, and Alain Dessein, INSERM Unite 399, Immunology and Genetic of Parasitic Diseases/Laboratory of Parasitology-Mycology, Faculty of Medicine, 27 Bd. Jean Moulin, 33850 Marseille Cedex 5, France. Michel Cot, ORSTOM, Antananarivo, Madagascar. Nadine Fievet, ORSTOM, OCEAC, Yaoundé, Cameroon. Laurent Abel, INSERM Unite 436, Centre Hospitalo-Universitaire Pitié-Salpêtrière, 91 Bd. de l’Hôpital, 75013 Paris, France.


REFERENCES


