



University of HUDDERSFIELD

University of Huddersfield Repository

Robson, Kathryn J.H., Dolo, Amagana, Hackford, Ian R., Doumbo, Ogobara, Richards, Martin B., Keita, Mamadou M., Sidibe, Touman, Bosman, Andrea, Modiano, David and Crisanti, Andrea

Natural polymorphism in the thrombospondin-related adhesive protein of *Plasmodium falciparum*

Original Citation

Robson, Kathryn J.H., Dolo, Amagana, Hackford, Ian R., Doumbo, Ogobara, Richards, Martin B., Keita, Mamadou M., Sidibe, Touman, Bosman, Andrea, Modiano, David and Crisanti, Andrea (1998) Natural polymorphism in the thrombospondin-related adhesive protein of *Plasmodium falciparum*. *The American Journal of Tropical Medicine and Hygiene*, 58 (1). pp. 81-89. ISSN 0002-9637

This version is available at <http://eprints.hud.ac.uk/id/eprint/21307/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

NATURAL POLYMORPHISM IN THE THROMBOSPONDIN-RELATED ADHESIVE PROTEIN OF *PLASMODIUM FALCIPARUM*

KATHRYN J. H. ROBSON, AMAGANA DOLO, IAN R. HACKFORD, OGOBARA DOUMBO, MARTIN B. RICHARDS, MAMADOU M. KEITA, TOUMANI SIDIBE, ANDREA BOSMAN, DAVID MODIANO, AND ANDREA CRISANTI
Medical Research Council Molecular Haematology Unit, Institute of Molecular Medicine, University of Oxford, Headington, Oxford, United Kingdom; Malaria Research and Training Centre, Department d'Epidemiologie des Affections Parasitaires, Ecole Nationale de Medicine et de Pharmacie et d'Odontostomatologie du Mali, Bamako, Mali; Service de Peditre, Hopital Gabriel Toure, Bamako, Mali; Istituto di Parassitologia, University di Roma La Sapienza, Rome, Italy; Department of Biology, Imperial College, London, United Kingdom

Abstract. We have developed a typing system using natural sequence variation in the thrombospondin-related adhesive protein (TRAP) gene of *Plasmodium falciparum*. This method permits a haplotype to be assigned to any particular TRAP gene. We have applied this method to a hospital-based, case control-study in Mali. Previous sequence variation and conservation in TRAP has been confirmed. Particular TRAP haplotypes can be used as geographic hallmarks. Because of the high level of conflict between characters, we have examined the phylogenetic relationships between parasites using a network approach. Having received patient samples from urban and periurban areas of Bamako, the majority of haplotypes were closely related and distinct from TRAP sequences present in other continents. This suggests that the structure of TRAP can only tolerate a limited number of sequence variations to preserve its function but that this is sufficient to allow the parasite to evade the host's immune system until a long-lived immune response can be maintained. It may also reflect host genetics in that certain variants may escape the host immune response more efficiently than others. For vaccine design, sequences from the major regional variants may need to be considered in the production of effective subunit vaccines.

The thrombospondin-related adhesive protein (TRAP) gene of *Plasmodium falciparum* was first described in 1988.¹ The gene derives its name from the deduced protein sequence that contains two amino acid motifs found in the adhesive host glycoprotein thrombospondin.² One of these motifs is based on the amino acid sequence Glu-Trp-Ser-Pro-Cys-Ser-Val-Thr-Cys-Gly-Lys-Gly-Thr-Arg-Ser-Arg-Lys-Arg, which has been demonstrated to allow TRAP to bind to sulfatide.³ The protein is present in sporozoites and erythrocytic-stage parasites.^{4,5} Antibodies against TRAP inhibit sporozoite invasion.³ Recent data demonstrate that TRAP is expressed in mature, salivary gland sporozoites but not in hemocoel sporozoites, suggesting that TRAP may be one of the unidentified accessory molecules required for infection of the vertebrate host.⁶

Data from a longitudinal and prospective study showed that the immune response to TRAP is long-lived in adults and older children but short-lived in young children.⁷ There were seasonal fluctuations of the levels of specific antibodies, as well as age-dependent quantitative differences.⁷ These data, together with that for the TRAP homolog sporozoite surface protein 2 (SSP2) from *P. yoelii*, suggesting that adoptive transfer of a CD8⁺ cytotoxic T cell clone recognizing SSP2 gave complete protection against *P. yoelii* infection, suggest that TRAP would be an important component of a multi-subunit malaria vaccine.⁸ Further evidence for this has been the identification of a cytotoxic T lymphocyte response to certain conserved epitopes in TRAP.^{9,10}

A major problem in the design of any malaria vaccine is which antigens to choose and which variants of these antigens are relevant. We have previously shown that TRAP is polymorphic and that all the nucleotide changes occurring outside what appeared to be a conserved core, containing the sulfatide-binding motif, resulted in amino acid changes.¹¹ The significance of this high degree of sequence variation is not known but it does suggest that host immune response may be a factor since the sequence of TRAP genes from

parasites in culture appear to be stable.¹¹ The parasites in this earlier study were from different geographic locations, apart from the three that were from children in the Gambia. The three Gambian sequences were closely related to each other, but nevertheless different. The most closely related gene sequences were from the parasite clones Dd2 (Indochina) and HB3A (Honduras). This report describes the genetic variation of the TRAP gene in Mali from samples collected during the 1992 malaria season from children (age range = 1–12 years) with mild and severe malaria.

PATIENTS, MATERIALS, AND METHODS

Study area and population. The study was carried out in Bamako, the principal city of Mali. All patients were recruited in the pediatric ward of the Gabriel Toure Hospital, which is located in the center of the city. An average of 3,200 children are admitted annually to this ward. Severe malaria cases account for 8% of all children admitted to this hospital; 23% of these cases prove fatal. The age range of the children in the study was 1–12 years. Blood samples were collected from September to November 1992. Malaria is seasonal in Mali, and the rainy season is from June to October, which corresponds to the period of maximal incidence of severe malaria in the study area. Malaria inoculation rates range one to 10 per person per year in urban areas, and from 50 and 200 in the surrounding rural villages.^{12,13} *Anopheles gambiae*, *A. arabiensis*, and *A. funestus* are the main malaria vectors in this part of Africa. The great majority of the population belongs to the Bambara ethnic group, with Fulani and Senoufu being in the minority. Parental consent was obtained for participation of all the children in the study, which was approved by the National Ethical Committee of the Ministry of Health of Mali. The modified Glasgow coma scale to be applicative to children, including those who have not learned to speak,¹⁴ was used in assessing severe malaria. Children having a coma score between 0 and

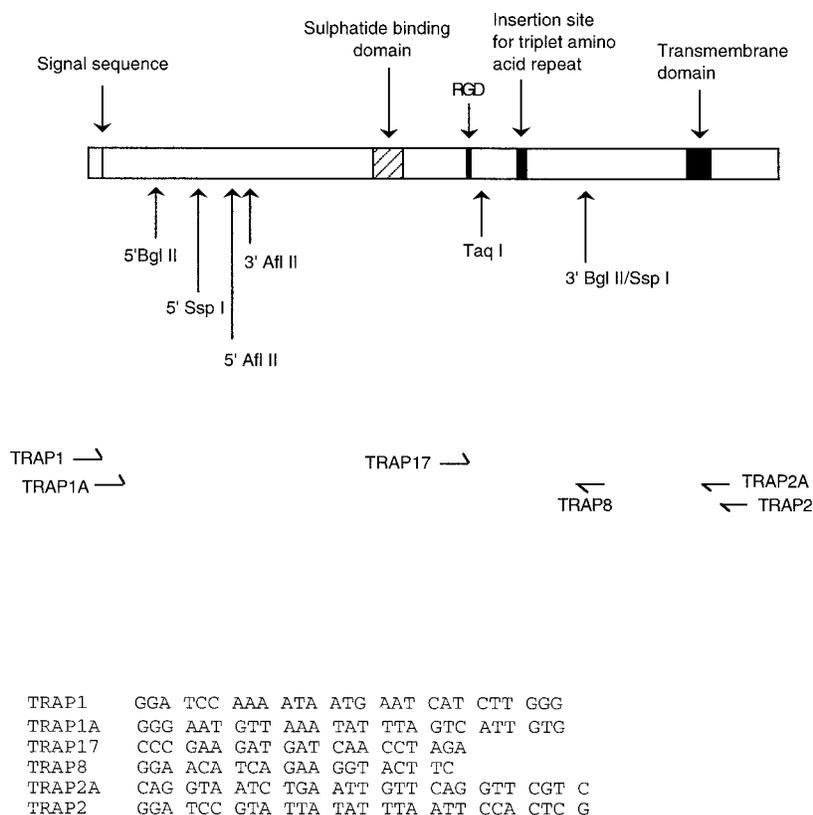


FIGURE 1. Schematic representation of the structure of the thrombospondin-related protein (TRAP) gene of *Plasmodium falciparum* and the strategy for the different polymerase chain reactions. The sequences of the oligonucleotide primers are given at the bottom of the Figure. TRAP1 and TRAP2 contain *Bam* HI sites at their 5' ends, TRAP1 corresponds to nucleotides (306-326), and TRAP2 to nucleotides (1979-2000) in the T9/96 sequence. The relative positions of the polymorphic restriction enzyme sites are shown. RGD = Arg-Gly-Asp.

2 were comatose. Children having a positive blood smear for *P. falciparum*, a body temperature greater than 37.5°C, and showing no signs of impaired consciousness were considered to have mild malaria. Children with patent bacterial infection and/or a history of repeated convulsions were excluded from the study. All children admitted to the study were treated according to World Health Organization (WHO) guidelines using drugs that were available free of charge as part of the study.

Blood samples and DNA collection. On entering the study, fingerpricks from each child were taken for thick and thin blood films and for DNA. The microscopic diagnosis of malaria was based on examination of 100 high-power fields of the thick film after staining with 3% Giemsa. Parasite counts were calculated according to the WHO method,¹⁵ assuming 6,000 leukocytes/ μ l. Every child was allocated a sheet of Whatman (Maidstone, United Kingdom) 3MM paper (13 cm \times 10 cm) on which the blood spots were stored prior to DNA extraction or hemoglobin isoelectric focusing. The name, age, code number, and date of collection were recorded in pencil on the appropriate sheet of Whatman 3MM paper. After collection, the blood spots were allowed to dry before the sheets of paper were transferred into individual resealable polythene bags for shipping and long-term storage at ambient temperature.

The DNA was extracted from the paper as previously described.¹⁶ Briefly, a small portion of blood spot (approximately 3 mm \times 3 mm) was soaked for 10–20 min at room

temperature in a 1.5-ml microfuge tube containing 1 ml of phosphate-buffered saline (PBS)/1% saponin (PBS = 0.14 NaCl, 0.01 M sodium phosphate, pH 7.2). The PBS/1% saponin was discarded and the DNA was extracted from the paper by the addition of 150 μ l of water and 50 μ l of Chelex-100 (Bio-Rad Laboratories, Hercules, CA) (made up to 20% [w/v] in water and adjusted to pH 9.5 with 1 M NaOH), followed by boiling for 8 min. The Chelex was removed by centrifugation and the supernatant was transferred to another microfuge tube. The DNA was then ready to use and was stored at -20°C prior to amplification.

Polymerase chain reaction (PCR) amplification. The sequences and relative positions of the primers are shown in Figure 1. All primers were used at a concentration of 100 ng/ μ l. Amplification was carried out using a programmable thermocycler (Omnigene; Hybaid, Teddington, United Kingdom).

Amplification was performed in two stages. The first round, in a final volume of 25 μ l, contained 5 μ l DNA, 200 μ M dNTPs, 1 \times *Taq* polymerase buffer containing 1.5 mM MgCl₂ and 1 unit of Amplitaq (Cetus, Branchburg, NJ). The primers were TRAP1 and TRAP2. The cycling conditions were as follows: two cycles of 1 min at 93°C, 1 min at 37°C, and 2 min at 72°C, followed by 35 cycles of 1 min at 93°C, 1 min at 55°C, and 2 min at 72°C, with a final cycle of 1 min at 93°C, 1 min at 55°C, and 10 min at 72°C. The reaction was then diluted to 250 μ l with sterile water and stored at -20°C until further use. If DNA had been purified from

TABLE 1
RFLP analysis of thrombospondin-related adhesive protein*

Restriction enzyme site Presence/Absence	Restriction fragment lengths size (bp) and order
<i>Bgl</i> II	
+ -	129, 1,509†
+ +	129, 1,090†, 419
- +	1,219†, 419
- -	1,638†
<i>Afl</i> II/ <i>Asp</i> 718	
+ -	351, 415, 872†
+ +	351, 33, 382, 872†
- +	384, 382, 872†
- -	766, 872
<i>Ssp</i> I	
+ -	13, 248, 1,377†
+ +	13, 248, 963†, 414
- +	13, 1,211†, 414
- -	13, 1,625†
<i>Taq</i> I	
+	926, 211†, 501
-	1,137†, 501

* For *Bgl*II, + - indicates that the 5' *Bgl*II site is present and that the 3' site is absent. The same rules apply in defining the presence or absence of the other restriction enzyme sites. Details can be obtained by email or fax from the corresponding author (KJHR).

† These fragments can vary in size since they contain the region with repeats.

liquid blood as in the Gambian study,¹⁷ only a single round of the PCR was found to be necessary. When using DNA purified from blood or cultured parasites, we use primers TRAP1A and 2A to obtain the diagnostic fragment sizes shown in Table 1.

For restriction fragment length polymorphism (RFLP) analysis, a second round of PCR was performed using the primers TRAP1A and TRAP2A (Figure 1). Amplification was performed as before but using 2 µl of diluted DNA from the first round of amplification in a final volume of 100 µl. Cycling conditions were as for the first round except that the initial two cycles with annealing at 37°C were omitted.

For repeat analysis, the second round of the PCR was performed using primers TRAP17 and TRAP8 (Figure 1). Amplification was performed as before using 1 µl of diluted DNA from the first round of amplification in a final volume of 20 µl containing 2 µCi of ³⁵S-dATP (Amersham, Little Chalfont, United Kingdom). Cycling conditions were as for the RFLP second round except that the annealing temperature was decreased to 45°C.

Certain DNA samples proved refractory to PCR amplification under these conditions. First and second round amplifications were repeated using *Taq* extender (Stratagene, Cambridge, United Kingdom) using the conditions recommended by the manufacturer. The extension times were extended to 4 min for the first round PCR and the second round RFLP PCR. This modification permitted the majority of the remaining DNA samples to be amplified.

Amplification of merozoite surface protein 1 (MSP-1) by the PCR was performed as previously described.¹⁸

Analysis by RFLP. The product from the second round PCR amplification was digested with the restriction enzymes *Bgl* II, *Ssp* I, *Taq* I, and *Afl* II following the manufacturer's (New England Biolabs, Beverly, MA) instructions. To aid the *Afl* II analysis, the enzyme *Asp* 718 (Boehringer Mann-

heim, Lewes, United Kingdom) was added to those digests. The *Bgl* II and *Ssp* I digests were resolved on 1.7% agarose gels. The *Taq* I and *Afl* II/*Asp* 718 digests were resolved using 1% agarose 1% NuSieve gels (FMC Bioproducts, Rockland, ME). The sizes of the various restriction fragments (Table 1) were used to define particular haplotypes. The presence of a particular restriction enzyme site is represented by + and the absence by -. Each combination of + and - was assigned a number. It should be noted that the 3' *Bgl* II and *Ssp* I sites are mutually exclusive. Haplotype 1 is defined as *Bgl* II ++, *Ssp* I +-, *Taq* I +, *Afl* II ++ and so in abbreviated form haplotype 1 becomes (+++-+++). Further details can be obtained by email (krobson@hammer.imm.ox.ac.uk) or fax (44-1865-222-500) from the corresponding author (KJHR).

Repeat polymorphism analysis. The samples for repeat analysis were precipitated with ethanol. The DNA pellets were washed with ethanol, air-dried, and resuspended in 3 µl of formamide dyes (80% deionized formamide, 1× glycerol-tolerant gel buffer [0.089 M Tris, 0.0288 M taurine, 5.4 mM disodium EDTA], 0.1% [w/v] bromophenol blue, 0.1% xylene cyanol FF). The samples were heated at 100°C prior to loading on 5% Long RangerTM (Flowgen, Rockland, ME) gels containing 1× glycerol-tolerant gel buffer and 7 M urea. Gels (31 cm × 38.5 cm × 0.4 mm) were electrophoresed at 60 W until the xylene cyanol FF had migrated to within 10 cm of the bottom of the gel. The M13mp18 sequencing ladders were used as size markers on every gel. The PCR product (TRAP17-TRAP8) from the laboratory line T9/96 was used as an additional control on all gels, the product size being 235 basepairs (bp). Gels were allowed to cool completely before drying and exposure to x-ray film for 1-7 days.

Sequencing of DNA. Direct sequencing of DNA products was performed using a Sequenase PCR product sequencing kit (U.S. Biochemicals, Cleveland, OH). Primers internal to the PCR primers gave optimal sequencing results.

Experimental design and statistical analysis. Eighty-five children with cerebral malaria were recruited during the 1992 malaria season. Each of these children was then age matched with three children with mild malaria matched for age and place of residence (locality). A total of 337 samples were analyzed. Yates' chi-square test (or Fisher's exact probability test when needed) was used for the analysis of prevalence rates, and the Kruskal-Wallis test was used for the analysis of age, parasite counts, and antibody levels.

Phylogenetic analysis. Networks¹⁹ were constructed sequentially by splitting the data, restriction enzyme site by restriction enzyme site. This was performed manually. Splits are represented by branches between haplotypes, each of which represents a change in a restriction enzyme site. Each haplotype was represented as a circle containing the haplotype number. The haplotypes were simplified to the number only, with all versions with repeats being pooled.

Hemoglobin isoelectric focusing. Isoelectric focusing was performed using the Resolve System Hemoglobin kit (Isolab, Mechelen, Belgium). Hemoglobin (Hb) for analysis was eluted from a single dry blood spot from the same sheet of Whatman 3MM paper used for parasite DNA extraction. Samples were analyzed for the presence of Hb A, Hb S, and Hb C.

RESULTS

Sequence analysis of a small region (200 bp) of the TRAP gene from six isolates from Mali showed that we needed an approach that would allow us to group related parasites based on their TRAP gene sequence and that would save on the expense and workload associated with sequencing. We therefore chose to perform a haplotype analysis on the gene, examining what we believed to be four polymorphic restriction enzyme sites throughout the gene. Nucleotide changes at these polymorphic restriction enzyme sites all result in amino acid changes. The restriction enzymes chosen were *Bgl* II, *Ssp* I, *Taq* I, and *Afl* II.

Polymorphisms associated with the enzyme *Bgl* II. Our earlier detailed sequence analysis showed that there were two *Bgl* II sites in the TRAP gene (Figure 1).¹¹ The first, at the 5' end of the gene, appeared not to be polymorphic, whereas the second, towards the 3' end of the gene, clearly was. Analysis of these samples from Mali showed that both 5' and 3' sites were polymorphic. Direct sequencing across the region containing the 5' *Bgl* II site in samples that appeared to lack the site demonstrated that there was a nucleotide change resulting in the amino acid change Leu⁴⁹ to Val⁴⁹, which corresponds to C145G in the DNA sequence. This *Bgl* II site was present in 96.4% of the chromosomes analyzed.

The 3' *Bgl* II site and what we describe as the 3' *Ssp* I site are mutually exclusive. The presence of the 3' *Bgl* II site was detected only in 0.6% of our sample chromosomes, whereas the *Ssp* I site was much more common (37.2%). The most common situation was the intermediate nucleotide sequence in which neither site was present (62.2%).

Polymorphisms associated with the enzyme *Ssp* I. In our study of the TRAP gene from different laboratory isolates and clones we found that there was a polymorphic 5' *Ssp* I site associated with DNA encoding amino acids 90, 91, and 92. The original sequence for the TRAP gene was from the Thai clone T9/96 in which the DNA encodes the amino acid motif Val⁹⁰Asn⁹¹Val⁹². In three samples taken from patients in the Gambia in 1988 the same motif was present.¹¹ Sequence analysis of three samples from our Mali survey also showed the same motif to be present. This motif is associated with the absence of the *Ssp* I site. Another related amino acid motif is also associated with the lack of this site, Ala⁹⁰Ser⁹¹Val⁹², and is found in two lines of African origin, 3D7A and 7901 (Palo Alto/Uganda). Two nucleotide changes are associated with this new amino acid motif.

The presence of this *Ssp* I site is associated with two related amino acid motifs, Ala⁹⁰Asn⁹¹Ile⁹² and Leu⁹⁰Asn⁹¹Ile⁹², which differ from the Val⁹⁰Asn⁹¹Val⁹² by two and three nucleotide changes, respectively. The presence of the 5' site was only detected twice in our survey. Since these samples were not sequenced, we do not know which of the two previously observed amino acid motifs were present. The presence of the 5' *Ssp* I site and the 3' *Bgl* II site seem to be more common in parasites of non-African origin.

Polymorphisms associated with the enzyme *Taq* I. A different history appears to be associated with the *Taq* I site. If the site is absent, the amino acid motif encoded by the DNA sequence is Ser³¹¹Ser³¹²Val³¹³Gln³¹⁴; if the site is present, then the amino acid motif is Phe³¹¹Ala³¹²Val³¹³Glu³¹⁴.

These sequences differ by three nucleotides. The *Taq* I site is absent in laboratory lines T9/96 and 7G8 but is present in all the others sequenced to date. The presence of the *Taq* I site appears to be in linkage disequilibrium with the Phe³¹¹Ala³¹²Val³¹³Glu³¹⁴ motif since no intermediary sequences have been seen. This cannot be said for the sequence changes associated with the 5' *Ssp* I site. The *Taq* I site was present in 66.7% of our sample chromosomes. These data suggest that either this mutation series of mutations may have arisen independently in different continents or else is a particularly ancient mutation that has been found in samples from Thailand, Mali, and Brazil. It is interesting to note that this series of amino acid changes follows the Arg³⁰⁷Gly³⁰⁸Asp³⁰⁹ motif.

Polymorphisms associated with the enzyme *Afl* II. There are two *Afl* II sites in the gene that are 33 bp apart. Loss of the second of these is associated with a change at amino acid position 134 of Ser to Thr. The first was not known to be polymorphic until we undertook this study. Direct sequencing across the region demonstrated that loss of the 5' site is associated with a Ser to Asn change at amino acid position 123 and corresponds to the nucleotide change G367A. All four combinations were observed in our survey. The frequencies were as follows: ++ = 6.2%, +- = 49.6%, -+ = 18.1%, and -- = 26.1%.

Haplotype analysis. Restriction enzyme analysis of the PCR-amplified TRAP gene permitted us to assign the gene to one of 96 possible haplotypes. It also allowed us to see whether mixed infections were present. It was not possible to get a complete result in all cases. Some samples (17 of 337) appeared refractory to PCR analysis regardless of whether human or parasite primers were tried. These samples were excluded from our analysis. In other cases in which more than two haplotypes were present, we could not always assign haplotypes to particular parasite chromosomes. In certain other samples, we were unable to generate data for all the restriction sites analyzed. We did not discard these data because they were still useful in assigning frequencies to particular RFLPs. We were able to assign numbered haplotypes to 319 chromosomes. Of the 96 possible numbered haplotypes, we observed 20. The results are given in Table 2.

Further analysis was possible due to our earlier observations¹¹ that the TRAP gene has a small repeating three amino acid motif inserted after amino acid 352. The PCR amplification across this region permitted a more detailed analysis of the parasites in our survey. Values varying from no extra repeats up to 16 were observed (Table 2). In many instances, these could be assigned to certain haplotypes (Table 2). We never observed a chromosome with either 11 or 13 extra repeat units. It should be noted that TRAP genes with repeats of more than 10 extra units were rarely observed (Table 2). In certain cases of mixed infections, we could not assign a repeat value to a given haplotype number; these have unassigned repeat numbers (Table 2). The repeat analysis proved very useful in analyzing mixed infections. The ability to leave the gels analyzing repeat values exposed to the x-ray film for several days permitted us to see mixtures that were not apparent using staining with ethidium bromide. Using repeat analysis alone we were able to examine the proportion of parasites present in mixed infections. Applying

TABLE 2
Summary of the observed thrombospondin-related adhesive protein haplotypes in Mali*

Haplotype number	Repeat letter and corresponding number of extra repeats																Unassigned	Total				
	A 0	B 1	C 2	D 3	E 4	F 5	G 6	H 7	I 8	J 9	K 10	L 11	M 12	N 13	O 14	P 15			Q 16			
3	1																			1	1	
25	2	1																			2	5
26	8	8	4	2	2	1	2	1			1										12	41
27	8	2																			5	15
28	2	8		1	1	1					1										4	18
29		1																			1	2
30	10	5																			6	21
31	1	1																			2	4
32	3	2																			2	7
41		3																			1	4
42	13	10	10	2	22	1	3	1	3		1										14	80
43	4	6	4				1	1	1	1											6	24
44	2	5	3		8		2	1			2										7	30
45	1	3																			2	6
46	6	8			2		1														8	25
47	10	3	3	1	1																1	19
48	1	3			2		1														3	10
74	1						1							1								3
75		1																				1
90															1	1					1	3
Total	73	70	24	6	38	3	11	4	4	1	5		1		1	1	1				77	319
Frequency of extra repeat units	118	111	42	12	67	5	26	8	13	5	5	0	1	0	2	3	1					419

* The haplotype number down the left hand side refers to the combination of restriction enzyme site. Haplotype 42 is +---+---. The row of letters across the top is the code applied to the number of repeating units, A being the baseline found in T9/96. Along the very bottom of the table is the frequency of the extra repeat units including those where it was not possible to assign a complete haplotype.

this approach, 180 children had single infections, 80 had two, 17 had three, three had four, two had five, and one had six infections. Combining the repeat analysis with the haplotype information, we were able to assign complete TRAP haplotypes for 242 chromosomes. We observed 74 of 1,440 of the possible combinations based on repeat and RFLP analysis (Table 2).

The degree of observed variation in the TRAP gene precluded the numbers necessary to see whether parasites present at the start of our study period were absent at the end. It should be noted that the three TRAP genes sequenced from children in the Gambia with malaria infections in 1988 used the nomenclature 42B, 42C, and 42D. Haplotype 42 was the most common in our survey. The numbers present also precluded any correlation with age, hemoglobin status, or disease severity.

The haplotypes of certain, well-known, laboratory lines that are from different geographic locations are given in Table 3. With the exception of haplotype 30, we did not ob-

TABLE 3

Thrombospondin-related adhesive protein haplotypes of common laboratory clones/lines of *Plasmodium falciparum*

Line/clone	Haplotype	Geographic origin
T9/96	30A	Thailand
HB3A	34A	Honduras
3D7A	42F	West Africa
7G8	37B	Brazil
Dd2	33A	Indochina
7901	42A	Uganda
K1	1A	Thailand
FCR3 group	1A	?

serve any of the haplotypes of other known laboratory lines of non-African origin. Line FCR3, which has been thought to have been of African origin, is haplotype 1, as is K1, a well-known Thai strain. We did not observe haplotype 1 in our study. Line FCR3 differs from haplotype 42 by a minimum of three and from haplotype 3 (we observed haplotype 3 once in our study) by a minimum of one restriction enzyme sites, respectively. We have previously seen haplotype 30 in Thailand (T9/96); this may be a more global haplotype. It is interesting to note that one of the ways haplotype 30 differs from haplotype 42 is by the *Taq* I site, which differentiates the two sequences around the Arg-Gly-Asp motif. Both T9/96 and 7G8 (Brazilian) lack the *Taq* I site next to the Arg-Gly-Asp motif. Changes in this motif may reflect polymorphisms in the mosquito or human host. These data suggest that such a haplotyping system reflects the geographic differences between *P. falciparum* parasites.

Phylogenetic analysis. The network for the haplotypes in the TRAP gene is shown in Figure 2. Although networks are unrooted (having no obvious common ancestor), the neutral theory of molecular evolution implies that all else being equal, the more common haplotypes are likely to be older. With the immune pressure on this malaria gene, this may not be true.

DISCUSSION

Unlike other malaria vaccine candidates, TRAP has not been studied in detail in naturally occurring infections. If TRAP is to be considered seriously as a component of a subunit vaccine, we need to understand the basis of natural sequence variation so that this information is reflected in the

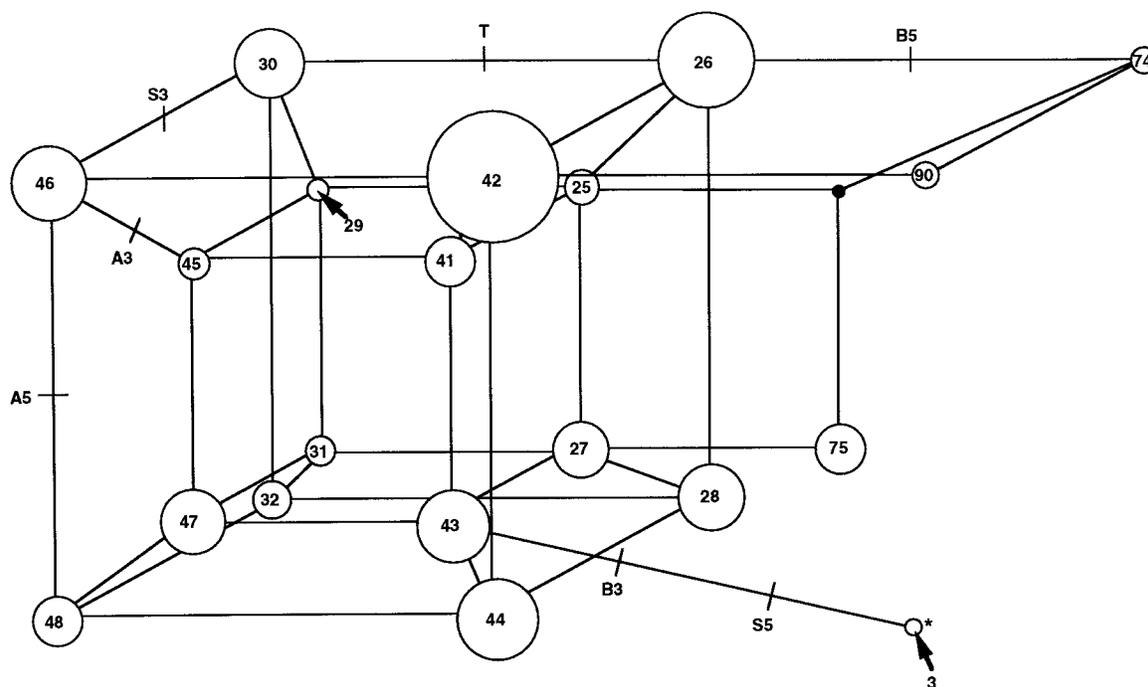


FIGURE 2. Phylogenetic network of thrombospondin-related protein haplotypes. Each numbered haplotype is represented by a circle containing the haplotype number, in which the area of the circle is proportional to the observed frequency of the haplotype. The small filled circle represents a theoretical intermediate node that could be removed to resolve B5 into two distinct events: the result would then be virtually a simple one-step connected network. Splits are represented by lines connecting the haplotypes, labeled for the restriction site changes that generated the split: A for *Afl* II, B for *Bgl* II, S for *Ssp* I, and T for *Taq* I. The subsequent 5 or 3 refers to whether it is at the 5' or 3' restriction enzyme site. The four-way conflict (S3T/A5/A3) is represented by placing haplotypes at the vertices of a four-dimensional cube, in which the fourth dimension (A3) is represented by connecting the vertices of two three-dimensional cubes. Haplotypes occur at all eight vertices. Certain haplotypes were observed 10 times or less (3, 25, 29, 31, 32, 41, 46, 48, 74, and 90). The majority of haplotypes occurring 10 or more times differ from haplotype 42 by one or two steps. Equally, some of the haplotypes that were observed infrequently differ from 42 by only one step. Therefore, certain evolutionary pathways may be preferred. * = B5+ B3+ S5+ S3- T+ A5- A3+.

sequences incorporated into such a vaccine. In this study, we have extended our earlier work on sequence variation, examining material from a case-control study in Mali. From our earlier work,¹¹ there was a suggestion that certain regions of the molecule could tolerate point mutations and insertions whereas others could not. We have detected two new RFLPs that corresponded to amino acid changes that have increased the power of our analyses. We have applied these observations to develop a PCR-based haplotyping system for TRAP that not only examines specific amino acid changes but also detects changes in the size of the small repeat region. This combinatorial approach allowed us potentially to identify up to 1,440 TRAP haplotypes. This method could easily be extended by examining more polymorphic restriction enzyme sites, e.g., *Fok* I. The restriction enzyme sites detected nucleotide changes that correlated with amino acid changes. We were able to assign complete haplotypes to 242 chromosomes and in doing so detected 74 of 1,440 possible haplotypes. Limited DNA sequence analysis suggests that this is an underestimate of the sequences in circulation. Together our data suggest that the molecule is under significant immunologic pressure. This may be due to a combination of factors including sporozoite exposure and the expression of TRAP both by mature sporozoites and by parasite infected red blood cells.

Using the numbered haplotypes, number 42 predominated. This haplotype had been observed in our earlier study using

samples from The Gambia taken in 1988.¹¹ Clones 3D7A and 7901, parasites also of African origin, are haplotype 42 by this system, but are known to differ both from each other and the isolates in this study at other positions in TRAP. The next most common haplotypes differed by one or two restriction enzyme sites from haplotype 42. Haplotype 3, at the other end of the spectrum, was observed only once and on its own. It differs from haplotype 42 at four restriction enzyme sites. It was found in a 12-year-old child with mild malaria living in an urban area. Where mixed infections of *P. falciparum* were present and haplotypes could be completely assigned, haplotype 42 was present with closely related haplotypes, and no particular combination was seen more frequently. This is in part due to the large number of haplotypes that were present in the study area, and therefore it was difficult to obtain statistically significant results for many of the observations. No particular TRAP haplotype or polymorphism was associated with parasites from children who died of severe malaria.

The system was used to analyze mixed infections in children with mild and severe malaria. It gives a minimum estimate of the number of parasites present in any mixture. When a TRAP haplotype was used as a marker to define mixed infections of *P. falciparum*, 49.3% of the children with severe malaria and 51.1% of the children with mild malaria had mixed infections of *P. falciparum*. The number of TRAP haplotypes present varied from one to six when

the repeat polymorphism was used on its own. Studies from Thailand,^{20,21} where parasites were cloned from particular patients, found six and six clones, respectively. Using monoclonal antibodies against three polymorphic blood-stage antigens, Conway and others were able to detect up to four clones per isolate.²² They found a mean of two *P. falciparum* clones in samples from patients with malaria in The Gambia. We were therefore not surprised to see this range of mixed infections of *P. falciparum* in children from Mali, another West African country. It is interesting to note that we were using a genetic marker that is predominantly associated with the sporozoite, whereas these other studies were looking at markers whose main biological activity was in blood-stage infections. These blood-stage proteins such as MSP-1 are much more subject to selective pressures by the host than a sporozoite protein such as TRAP.

Repeat length appears to be independent of TRAP haplotype. The significance of the small tandem repeat in TRAP is not understood. It occurs in the part of the molecule where significant repeating motifs are found in SSP2,^{23,24} *P. berghei* TRAP,²⁵ and TRAPs from *P. vivax*, *P. knowlesi*, and *P. gallinaceum*,²⁶ and is also the region where the TRAP molecules differ most from each other. Dual mixed infections based on TRAP analysis tended to be due to related parasite haplotypes differing either by repeat length or one restriction enzyme site. The high degree of sequence polymorphism precluded statistical analysis for this observation.

We have analyzed the data from the case-control study looking at levels of parasitemia, disease severity, sickle cell status, age, month of infection, and geographic location (urban, periurban, and rural) with relation to TRAP polymorphisms/haplotype. For statistical purposes, haplotypes were simplified and analyzed either for repeat letter or haplotype number. The sample size, together with the number of haplotypes and the number of comparisons, made were such that no statistically significant results were obtained. A further study would be required to increase the sample size and also to look at parasites occurring subsequently in this group of children.

The parasite samples came from a wide variety of children from one to 12 years of age living predominantly in urban and periurban regions of Bamako, Mali. Haplotype analysis revealed that 50% of the children in this study had mixed infections of *P. falciparum*, regardless of whether they had mild or severe disease. This suggests that if other markers are combined in a similar analysis, the vast majority of children will be shown to have mixed infections. It is difficult to say whether this is a characteristic of the children in our study or a country where malaria is hyperendemic or whether it relates to the number of infectious bites per child per year. Mixed blood feeding by mosquito vectors from more than one individual is unusual,²⁷ and therefore mating will mostly occur between parasites acquired from a single individual. This suggests that the children in our study who clearly had multiple infections had been bitten on more than one occasion by an infected mosquito and very possibly had failed to clear earlier infections. This result is not unexpected since the study was carried out during the main malaria transmission season in Bamako.

Based on our TRAP analysis, only two pairs of siblings had single infections, suggesting the possibility that they

might have been bitten by the same mosquito. A second gene, MSP-1, was analyzed in these four children. The two children infected with parasite TRAP haplotype 47A, a haplotype that was seen 10 times, had MSP-1 genes that were clearly different. The two children with a parasite having TRAP haplotype 42I, a haplotype that was observed three times, had MSP-1 genes that could not be distinguished using the block 3 repeat polymorphism. This was the only instance observed where there was any suggestion that children might have been infected by parasites from the same mosquito.

A series of studies looking at sequence variation have been carried out for other potential malaria vaccine constituents. These include the sporozoite stage of the parasite life-cycle, the circumsporozoite protein, and the two blood-stage merozoite surface proteins MSP-1 and MSP-2.²⁸⁻³³ These proteins differ from TRAP in that they all contain major repeating epitopes or blocks of repeats. The sequence variation in these antigens has been examined in a variety of ways that have included RFLP and hybridization analysis of PCR products, as well as sequencing of PCR products both cloned and uncloned.

The consequences of these studies are that for subunit malaria vaccines to be effective, the major variants of the different components for a given area may be required. This may be more important in Africa where more variants of the different protein components have been reported. This may also apply to regions of the world where there are large migrant populations such as on the Thai-Burmese border. The major influence of hemagglutinin and neuraminidase serotypes determines which major types are used in flu vaccines on an annual basis. As with influenza, it must be assumed that host immune responses are responsible for the observed sequence variation.

The structure of TRAP has yet to be determined so it is difficult to understand the significance of these sequence changes. From sequence alignment, the extracellular portion of TRAP can be divided into three domains. The first domain shares significant sequence similarity with the I domain of the a subunit of the integrin CR3 (CD11b/CD18).³⁴ The majority of RFLPs that we have examined in this study are in this region of the molecule. All of these amino acid substitutions can be regarded as conservative and as such would not alter the structure of this part of the molecule significantly and probably the function, but would be sufficient to alter the antigenicity. The second domain of TRAP contains the sulfatide-binding domain and has been found both in our earlier¹¹ and this study to be conserved. It is interesting to note that one of the major cytotoxic T cell epitopes for a large number of Europeans is found in this conserved part of the molecule.¹⁰ The third domain contains the polymorphism associated with the Arg-Gly-Asp motif, which can be considered to alter the hydrophobicity of the flanking sequence. The small tandem repeat is also found in this region as is the complex *Bgl* II/*Ssp* I polymorphism.

The fact that the same point mutations are found in parasites from different geographic locations suggests that these point mutations can arise independently and become fixed in different populations, which is suggestive of selective pressure. Each sporozoite is only exposed to the host's immune system for a short time but the numbers of sporozoites

an individual is exposed to over a lifetime is large. This is clearly sufficient pressure for TRAP variants to evolve and be selected. However, it should be pointed out that since the same point mutations are found in different geographic locations, this suggests that structural constraints are important and may in fact limit the repertoire of variants found. The majority of the point mutations give rise to conservative amino acid substitutions and such changes are unlikely to give rise to major alterations in particular epitopes. It is also possible that this variation in the form of point mutations occurred at a bottleneck during evolution of the parasite before the spread to different parts of the world.

In playing a game of chance such as a slot machine or throwing four dice, the chance of getting four identical symbols or throwing four sixes occurs infrequently but at a fixed odds. Most of the time a random combination of numbers or symbols occurs and this is what appears to be the case with the polymorphic sites in TRAP. At present we do not know whether certain combinations occur more often by chance alone or whether immunologic pressure is the selective force behind this. Thrombospondin-related adhesive protein is continually evolving. This is shown by applying phylogenetic analysis to the system (Figure 2).¹⁹ This is the first report of this method being applied to sequence variation in malaria. A two-dimensional approach has been tried with MSP-2 to try and identify geographic clustering.³⁵ Applying phylogenetic analysis to TRAP sequence variation suggests that there is fluidity in the system that would not be recognized using two-dimensional approaches. This method does permit clustering and is demonstrated by the fact that the most common haplotype, haplotype 42, differed from the majority by one restriction enzyme site using this classification system. Immunity may result as a consequence of being exposed to a certain number of permutations and combinations of TRAP and other antigens. This would explain why immunity develops with age after repeated exposure to parasite infections.

Acknowledgments: We thank Professor Mario Coluzzi for valued support and Professor Sir David Weatherall for continuing encouragement. We also thank Professor David Walliker and Dr. Alison Creasey for the gift of 7G8 DNA.

Financial support: This work was funded by a grant from the Commission of the European Communities, Program STD2 no. CT90-0320, and by the Italian Research Council. Amagana Dolo was supported by a TDR/WHO fellowship. The work in Mali was supported by TDR/OMS contract no. 910396.

Authors' addresses: Kathryn J. H. Robson and Ian R. Hackford, Medical Research Council, Molecular Haematology Unit, Institute of Molecular Medicine, Headley Way, Headington, Oxford OX3 9DS, United Kingdom. Amagana Dolo and Ogobara Doumbo, Malaria Research and Training Centre, Department d'Epidemiologie des Affections Parasitaires Ecole Nationale de Medicine, de Pharmacie et d'Odontostomatologie du Mali, Bamako, BP 1805, Mali. Martin B. Richards, Institute of Molecular Medicine, University of Oxford, Headley Way, Headington, Oxford OX3 9DS, United Kingdom. Mamadou M. Keita and Toumani Sidibe, Service de Pediatrie, Hopital Gabriel Toure, Bamako, Mali. Andrea Bosman and David Modiano, Istituto di Parassitologia, Universita di Roma La Sapienza, P.A. Moro 5, 00185 Rome, Italy. Andrea Crisanti, Istituto di Parassitologia, Universita di Roma La Sapienza, P.A. Moro 5, 00185 Rome, Italy and Department of Biology, Imperial College, Prince Consort Road, London SW7 2BB, United Kingdom.

REFERENCES

1. Robson KJH, Hall JRS, Jennings MW, Harris TJR, Marsh K, Newbold CI, Tate VE, Weatherall DJ, 1988. A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* 355: 79–82.
2. Lawler J, Hynes RO, 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J Cell Biol* 103: 1635–1648.
3. Muller H-M, Reckmann I, Hollingdale MR, Bujard H, Robson KJH, Crisanti A, 1993. Thrombospondin related anonymous protein (TRAP) of *Plasmodium falciparum* binds specifically to sulfated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. *EMBO J* 12: 2881–2889.
4. Cowan G, Krishna S, Crisanti A, Robson K, 1992. Expression of thrombospondin-related anonymous protein in *Plasmodium falciparum* sporozoites. *Lancet* 339: 1412–1413.
5. Rogers WO, Malik A, Mellouk S, Nakamura K, Rogers MD, Szarfman A, Gordon DM, Nussler AK, Aikawa M, Hoffman SL, 1992. Characterization of *Plasmodium falciparum* sporozoite surface protein 2. *Proc Natl Acad Sci USA* 89: 9176–9180.
6. Robson KJH, Frevert U, Reckmann I, Cowan G, Beier J, Scragg IG, Takehara K, Bishop DHL, Pradel G, Sinden R, Saccheo S, Muller H-M, Crisanti A, 1995. Thrombospondin related adhesive protein (TRAP) of *Plasmodium falciparum*: expression during sporozoite ontogeny and binding to hepatocytes. *EMBO J* 14: 3883–3894.
7. Scarselli E, Tolle R, Koita O, Diallo M, Muller H-M, Fruh K, Doumbo O, Crisanti A, Bujard H, 1993. Analysis of the human antibody responses to thrombospondin-related anonymous protein of *Plasmodium falciparum*. *Infect Immun* 61: 3490–3495.
8. Khusmith S, Sedegah M, Hoffman SL, 1994. Complete protection against *Plasmodium yoelii* by adoptive transfer of a CD8⁺ cytotoxic T-cell clone recognizing the sporozoite surface protein 2. *Infect Immun* 62: 2979–2983.
9. Aidoo M, Lalvani A, Allsopp GEM, Plebanski M, Meisner SJ, Krausa P, Browning M, Morris-Jones S, Gotch F, Fidock DA, Takiguchi M, Robson KJH, Greenwood BM, Druihle P, Whittle HG, Hill AVS, 1995. Identification of conserved antigenic components for acytotoxic T lymphocyte-inducing vaccine against malaria. *Lancet* 345: 1003–1007.
10. Wizen B, Houghton RA, Parker KC, Coligan JE, Church P, Gordon M, Ballou RP, Hoffman SL, 1995. Irradiated sporozoite vaccine induces HLA-B8-restricted cytotoxic T lymphocyte responses against two overlapping epitopes of the *Plasmodium falciparum* sporozoite surface protein 2. *J Exp Med* 182: 1435–1445.
11. Robson KJH, Hall JRS, Davies LC, Crisanti A, Hill AVS, Wellem TE, 1990. Polymorphism of the TRAP gene of *Plasmodium falciparum*. *Proc R Soc [Lond] B* 242: 205–216.
12. Doumbo O, Sangare O, Touri YT, 1989. Le paludisme dans le sahel: l'exemple du Mali. AUPELF-UREF, John Libbey, ed. *Maladies Tropicales Transmissibles*. Paris: Eurotex, 11–32.
13. Doumbo O, Ouattara NI, Koita MO, Touri YT, Troari SF, Quilici M, 1992. Approche ecogeographique du paludisme, en milieu urbain: la ville de Bamako au Mali. *Ecol Hum* 7: 3–15.
14. Warrell DA, Molyneux ME, Beales PF, 1990. Severe and complicated malaria. Second Edition. World Health Organization Malaria Action Programme. *Trans R Soc Trop Med Hyg* 84 (suppl 2): 1–65.
15. WHO, 1991. *Basic Malaria Microscopy. Part I and II*. Geneva: World Health Organization.
16. Kyes S, Craig AG, Marsh K, Newbold CI, 1993. *Plasmodium falciparum*: a method for the amplification of S antigens and its application to laboratory and field samples. *Exp Parasitol* 77: 473–483.
17. Hill AVS, Elvin J, Willis AC, Aidoo M, Allsopp GEM, Gotch FM, Gao XM, Takiguchi M, Greenwood BM, Townsend ARM, McMichael AJ, Whittle HC, 1992. Molecular analysis

- of the association of HLA-B53 and resistance to severe malaria. *Nature* 360: 434–439.
18. Kimura E, Mattei D, Di Santi SM, Scherf A, 1990. Genetic diversity in the major merozoite surface antigen of *Plasmodium falciparum*: high prevalence of a third polymorphic form detected in strains derived from malaria patients. *Gene* 91: 57–62.
 19. Bandelt H-J, 1994. Phylogenetic networks. *Verh Naturwiss Ver Hamburg (NF)* 34: 51–71.
 20. Thaithong S, Beale GH, Fenton B, McBride JS, Rosario V, Walker A, Walliker D, 1984. Clonal diversity in a single isolate of the malaria parasite *Plasmodium falciparum*. *Trans R Soc Trop Med Hyg* 78: 242–245.
 21. Pinswadi C, Thaithong S, Beale G, Fenton B, Webster HK, Pavanand K, 1987. Polymorphism of proteins in malaria parasites following mefloquine treatment. *Mol Biochem Parasitol* 23: 159–164.
 22. Conway DJ, Greenwood BM, McBride JS, 1991. The epidemiology of multiple-clone *Plasmodium falciparum* infections in Gambian patients. *Parasitology* 103: 1–6.
 23. Hedstrom RC, Campbell JR, Leef ML, Charoenvit Y, Garter M, Sedegah M, Beaudoin RL, Hoffman SL, 1990. A malaria sporozoite surface antigen distinct from the circumsporozoite protein. *Bull World Health Organ (suppl)*: 152–157.
 24. Rogers WO, Rogers MD, Hedstrom RC, and Hoffman SL, 1992. Characterization of the gene encoding sporozoite surface protein 2, a protective *Plasmodium yoelii* sporozoite antigen. *Mol Biochem Parasitol* 53: 45–52.
 25. Robson KJH, Naitze S, Barker G, Sinden RE, Crisanti A, 1997. Cloning and expression of the thrombospondin related adhesive protein gene of *Plasmodium berghei*. *Mol Biochem Parasitol* 84: 1–12.
 26. Templeton TJ, Kaslow DC, 1997. Cloning and cross-species comparison of the thrombospondin-related anonymous protein (TRAP) gene from *Plasmodium knowlesi*, *P. vivax*, and *P. gallinaceum*. *Mol Biochem Parasitol* 84: 13–24.
 27. Boreham PFL, Lenehan JK, Boulzaguet R, Storey J, Ashkar TS, Nambiar R, Matsushima T, 1979. Studies on multiple feeding by *Anopheles gambiae s.l.* in a Sudan savanna area of north Nigeria. *Trans R Soc Trop Med Hyg* 73: 418–423.
 28. Yoshida N, Di Santi S, Dutra AP, Nussenzweig RS, Nussenzweig V, Enea V, 1990. *Plasmodium falciparum*: restricted polymorphism of T cell epitopes of the circumsporozoite protein in Brazil. *Exp Parasitol* 71: 386–392.
 29. Lockyer MJ, Marsh K, Newbold CI, 1989. Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epitopes of the circumsporozoite protein. *Mol Biochem Parasitol* 37: 275–280.
 30. Doolan DL, Saul AJ, Good MF, 1992. Geographically restricted heterogeneity of the *Plasmodium falciparum* circumsporozoite protein: relevance for vaccine design. *Infect Immun* 60: 675–682.
 31. Felger I, Tavul L, Kabintik S, Marshall V, Genton B, Alpers M, Beck H-P, 1994. *Plasmodium falciparum*: extensive polymorphism in merozoite surface antigen 2 alleles in an area with endemic malaria in Papua New Guinea. *Exp Parasitol* 79: 106–116.
 32. Marshall VM, Anthony RL, Bangs MJ, Pumomo, Anders RF, Coppel RL, 1994. Allelic variants of the *Plasmodium falciparum* merozoite surface antigen 2 (MSA-2) in a geographically restricted area of Irian Jaya. *Mol Biochem Parasitol* 63: 13–21.
 33. Contamin H, Fandeur T, Rogier C, Bonnefoy S, Konate L, Trape J-F, Mercereau-Puijalon O, 1996. Different genetic characteristics of *Plasmodium falciparum* isolates collected during successive clinical malaria episodes in Senegalese children. *Am J Trop Med Hyg* 54: 632–643.
 34. Let J-O, Ritu P, Arnout MA, Liddington R, 1995. Crystal structure of the A domain from the α subunit of the integrin CR3 (CD11b/CD18). *Cell* 80: 631–638.
 35. Prescott N, Stowers AW, Cheng Q, Bobogare A, Rzepczyk CM, Saul A, 1994. *Plasmodium falciparum* genetic diversity can be characterised using the polymorphic merozoite surface antigen 2 (MSA-2) gene as a single locus marker. *Mol Biochem Parasitol* 63: 203–212.