Epidemiology of multiple *Plasmodium falciparum* infections

2. Genotypes of merozoite surface protein 2 of *Plasmodium falciparum* in Tanzania

I. Felger*, A. Irion, S. Steiger and H.-P. Beck *Swiss Tropical Institute, Socinstrasse 57, 4002 Basel, Switzerland*

Abstract

The merozoite surface protein 2 (MSP2) of *Plasmodium falciparum* is extremely polymorphic: 82 different *msp2* alleles were found in 4 studies of molecular epidemiology conducted in Tanzania. This diversity renders *msp2* suitable as a marker gene for the genotyping of *P. falciparum* infections. Amplification of *msp2* by the polymerase chain reaction (PCR), and subsequent restriction digests of the PCR product (PCR–restriction fragment length polymorphism genotyping), has proved to be an informative tool for enumerating multiple concurrent infections in a blood sample, and distinguishing individual alleles. Depending on the specific questions asked in a genotyping study, analytical techniques of different degrees of complexity are employed. The restriction fragments resulting from a single *Hin*fI digest generally allow the enumeration of multiple concurrent infections and the determination of their allelic families. When a restriction pattern is too complex to be resolved, owing to the high number of concurrent infections, or due to the appearance of previously undescribed alleles, one or more additional digests (*Dde*I, *Rsa*I, *Scr*fI) may be necessary. To determine individual alleles unequivocally, in particular in longitudinal studies, when several consecutive samples need to be compared with each other, a more detailed analysis involving all 3 additional digests is applied. The methodological experience and results gained in 4 epidemiological field studies involving *msp2* genotyping are summarized. We also provide the *Hin*fI restriction patterns and some nucleotide sequences of the alleles found so far in our studies in Tanzania.

Keywords: malaria, *Plasmodium falciparum*, multiple infection, genotypes, *msp2* gene, restriction patterns, Tanzania

Introduction

Alleles of the merozoite surface protein 2 (MSP2) are grouped into 2 allelic families according to the dimorphic structure of the variable non-repetitive region (SMYTHE *et al.*, 1990). The extensive diversity is due to an allele-specific central region which comprises tandem repeats of varying size. Differences in the number of copies of these repeats result in length polymorphism. These characteristics of *msp2* have been exploited by several genotyping methods. Common to all methods of *msp2* genotyping is a polymerase chain reaction (PCR) amplification of a central part of the *msp2* gene, which comprises the allele-specific repetitive region and conserved flanking sequences. The major methods used for *msp2* genotyping are (i) analysis of size polymorphism with subsequent hybridization with family-specific probes (FOLEY *et al.*, 1992; BABIKER *et al.*, 1994; NTOUMI *et al.*, 1995; VIRIYAKOSOL *et al.*, 1995), or (ii) restriction fragment length polymorphism (RFLP) of the amplified product (FELGER *et al.*, 1993). The *msp2* gene has been used extensively as polymorphic marker gene in field studies in different geographical locations within either a multilocus analysis or single locus analysis. All studies showed extensive polymorphism at this locus, even in areas of low endemicity.

The use of *msp2* as marker has proven to be of great use when individual *Plasmodium falciparum* infections need to be identified. PCR–RFLP genotyping of parasites for *msp2* makes it possible to distinguish the individual parasite infections concurrently present in a blood sample. The possibility of tracing individual parasite clones over time allows detailed studies of infection dynamics. In immunological studies, genotyping can also provide important information on the diversity of antigenic challenge. Most importantly, genotyping makes it possible to determine the multiplicity of infection, which can be used as an outcome measurement of interventions such as drug trials, vaccine trials, or exposure-reducing interventions.

The *msp2* gene is not only an extremely polymorphic marker gene, but its product is also a vaccine candidate. MSP2, as part of a subunit vaccine, is currently under field trial in Papua New Guinea (PNG) (GENTON *et al.*, 1996). Two genotyping studies from PNG have already shown that the 2 allelic families of *msp2* are differently associated with morbidity (ENGELBRECHT *et al.*, 1995; AL-YAMAN *et al.*, 1997). Thus, using *msp2* for genotyping offers both genetic diversity for high resolution typing and additional information on the biological role and significance of MSP2 as an immunological target. In this paper we describe in detail the PCR–RFLP genotyping technique and its possible applications, as well as the diversity and RFLP patterns of the *msp2* alleles detected so far in blood samples from Kilombero District, Tanzania.

*Author for correspondence: phone +41 61 284 8116, fax +41 61 271 8654, e-mail <felger@ubaclu.unibas.ch>

Methods

Extraction of deoxyribonucleic acid

Infected erythrocytes (RBC) from samples of packed cells (stored at –20°C) were used for isolation of parasite deoxyribonucleic acid (DNA). After thawing the pellets, 5 µL of packed cells (equivalent to about 12·5 µL of whole blood) were added to 50 µL of 4 M guanidine isothiocyanate containing 25 mM sodium citrate (pH 7·0) and 0·5% sodium sarcosylsulphate; 5 µL of 2 M sodium acetate (pH 4·4) were then added. The mixture was left on ice for 10 min and then extracted once with 100 µL of phenol (pH 8·0):chloroform (1:1), precipitated with isopropanol and resuspended in 50 µL of distilled water.

PCR amplification

Both a primary and a nested PCR were performed with each sample. For the primary PCR, 5 µL of the extracted DNA (corresponding to 0·5 µL of RBC pellet) were added to 95 µL of reaction mix including 1·5 units of *Taq* polymerase (Gibco BRL Life Technologies), 50 mM potassium chloride, 10 mM Tris–HCl (pH 8·8), 1·5 mM magnesium chloride, 0·2 mM of each deoxynucleotide (dNTP), and 0·5 mM of each primer. The primer pair for the primary PCR corresponded to nucleotides 3–23 and 789–811 from the 5′ and 3′ conserved region of the MAD71 sequence of *msp2* (FOLEY *et al.*, 1992). The PCR conditions were 5 min at 94°C followed by 30 cycles of 30 sec at 94°C, 2 min at 55°C, and 2 min at 70°C. Two µL of the primary PCR product were reamplified in a nested reaction (total volume 100 µL)

with nested primers (FOLEY *et al.*, 1992) corresponding to nucleotides 111–129 and 709–728 of the same sequence. The cycle conditions for the nested PCR were the same as those for the primary reaction. In earlier *msp2* genotyping studies in PNG, PCR conditions with a lower annealing temperature, according to FOLEY *et al.* (1992), were used. However, due to the complexity of Tanzanian *P. falciparum* infections, a higher annealing temperature was chosen to increase specificity, in spite of reduced yields of PCR products. The nested PCR products were separated on a 10% polyacrylamide (PAA) gel. Samples from which no detectable PCR product was obtained were re-examined at least once, starting from the DNA preparation. For quality control, DNA extraction, both rounds of PCR, and restriction digests were repeated with 10% of all samples.

Restriction digests

Nested PCR products were routinely digested with *Hin*fI for 2 h and then run on 10% PAA gels (BRL V16) using 1 mm spacers. The separation was improved if the gels were prepared at least one day before use. Gels were stained with ethidium bromide and photographed or electronically recorded. Fragment sizes were estimated by comparison with a DNA size marker (1 kb ladder, Gibco BRL Life Technologies). If samples from longitudinal studies were to be compared with each other, they were run side by side on the gel. When a higher discriminatory power of the RFLP pattern was desired, usually in the case of new alleles or ambiguities, additional *Dde*I, *Rsa*I and *Scr*FI restriction digests were performed. When direct comparison of RFLP patterns with those of already defined alleles was needed, restriction digests of both the known and the unknown PCR product were repeated, and run side by side on a PAA gel. Similarly, when consecutive blood samples from a longitudinal study had to be compared with each other, the digested PCR products of these samples were run on the same gel side by side, in order to identify size differences of a few nucleotides.

Results

Diversity of Tanzanian msp2 *alleles*

We performed several genotyping studies with blood samples from Ifakara or neighbouring villages in Tanzania. In all studies the same PCR–RFLP typing method was used. Taking the 4 studies listed in Table 1 together, 3506 individual infections with *P. falciparum* were genotyped. In these samples, 82 different *msp2* alleles were detected, of which 41 belonged to the FC27 family and 41 to the 3D7 family.

Table 1. Comparison of numbers of different *msp2* alleles of *Plasmodium falciparum* found in four studies of children from the Kilombero valley, Tanzania

Reference	Total no. of individual *P. falciparum* infections	No. of different alleles	No. of FC27-type alleles	No. of 3D7-type alleles
IRION *et al.* (1997)	1034	58	26	32
BECK *et al.* (1997)	292	51	18	33
FELGER *et al.* (1998)	365	50	24	26
FRASER-HURT *et al.* (1999)	1815	68	34	34
Total	3506	82	41	41

Principles of PCR–RFLP genotyping

To determine the number of concurrent infections per blood sample (multiplicity of infection), or to discriminate whether an allele belongs to the 3D7 or FC27 allelic family, only a single restriction digest (*Hin*fI) is performed on the nested PCR product of the *msp2* gene. The dimorphic region of *msp2* shows family-specific *Hin*fI restriction sites which are located in the variable non-repetitive region and which are shared by most alleles of the same family. Only rare point mutations within a *Hin*fI site, or the appearance of a new repeat type in the FC27 family, were found to cause deviations from the ubiquitous family-specific restriction fragments. In all other cases, size deviations were caused by crossing over between alleles of different families. The family-specific restriction fragments resulting from *Hin*fI digests of FC27-type alleles are 2 fragments with lengths of 137 bp and 115 bp, and of 3D7-type alleles, 2 fragments of 70 bp and 108 bp. Fragments of these sizes can easily be discriminated on a 10% PAA gel. Thus, either one pair of fragments or the other indicates the presence of at least one allele of the respective allelic family. If both fragment pairs appear in one sample, at least one allele of each family is present. The total number of different alleles found in a given blood sample is determined by the remaining *Hin*fI restriction fragments which form an allele-specific pattern. The restriction patterns of individual alleles are described below.

FC27 family of msp2 *alleles*

Table 2 lists all 41 FC27-type alleles found in these studies, their restriction fragments obtained after *Hin*fI digestion of the nested PCR product, and the Genbank™ accession numbers of those alleles already sequenced. Some alleles contained more than one 96 bp *Hin*fI fragment, as indicated in Table 2. A 96 bp *Hin*fI fragment is generated if an allele contains at least 2 copies of a 96 bp repeat unit in tandem array. Therefore, the actual number of repeats equals the number of 96 bp fragments +1.

*Hin*fI restriction digestion of a nested PCR product of FC27-type alleles normally produces 2 conserved fragments of 137 bp and 115 bp, which represent the 5′ and 3′ ends of the amplification product, respectively. A selection of *Hin*fI digested FC27-type alleles from Tanzania is shown in Fig. 1. The family-specific fragments (115 bp and 137 bp) are present in all alleles in lanes Wos6 to K1. Two repeat units (96 bp and 36 bp) are characteristic of all FC27-type alleles. At least one copy of each repeat is normally present in an allele of the FC27 family, while different alleles vary in the number of copies of 96 bp and 36 bp repeats present in the central region of the PCR product. Thus, this central region gives rise to allele-specific restriction fragments.

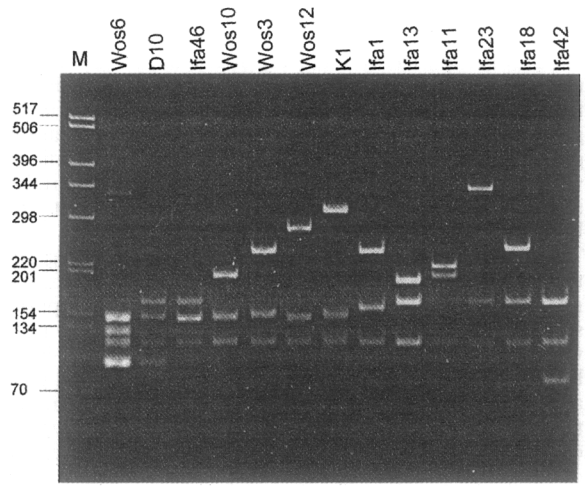

Fig. 1. *Hin*fI restriction patterns of selected FC27-type *msp2* alleles of *Plasmodium falciparum* from Tanzania. Restriction fragments were separated on a 10% PAA gel. The fragment sizes and Genbank™ accession numbers of these alleles are listed in Table 2. Lane M: DNA size marker (1 μg of 1 kb ladder, Gibco BRL Life Technologies); lanes Wos6 and D10: alleles with 4 copies (Wos6) or 2 copies (D10) of the 96 bp repeat; lanes Ifa46–K1: increasing copy numbers of the 36 bp repeat ranging from Ifa46 (1 copy) to K1 (5 copies); lanes Ifa1–Ifa23: increasing numbers of the 9 bp NAP repeat with 2, 6, 8 and 23 copies, respectively; lanes Ifa18 and Ifa42: recombined alleles.

Table 2. Details of 41 *msp2* genotypes of the *Plasmodium falciparum* FC27 allelic family from Tanzania

msp2 genotypes	Genbank™ accession no.		*Hin*fI restriction fragments (no. of base pairs)		
D10	JO3828	1×96	115	137	162
Ifa1	AF010462	–	115	146	234
Ifa2	–	–	115	308	162
Ifa4	–	–	115	137	378
Ifa10	–	–	115	164	162
Ifa11	AF010454	–	115	200	198
Ifa13	AF010455	–	115	182	162
Ifa14[a]	AF010456	2×96	115	131	162
Ifa15	–	–	115	218	162
Ifa16	–	–	115	146	270
Ifa17	–	1×96	115	146	126[b]
Ifa18	AF010457	–	115	239	162
Ifa23	AF010463	–	115	335	162
Ifa27	–	2×96	115	106+31	126[b]
Ifa28	–	–	115	137	342
Ifa30	–	1×96	115	191	162
Ifa31	AF010458	–	115	191	198
Ifa32	–	1×96	115	137	198
Ifa33	–	1×96	115	137	150[b]
Ifa34	–	–	115	191	234
Ifa36	–	–	115	218	198
Ifa38	–	–	115	137	126[b]
Ifa40	–	–	115	245	162
Ifa41	AF010459	–	115	272	198
Ifa42[a]	AF010460	–	115	31+82	162
Ifa43	–	–	115	–	371
Ifa44	–	–	115	146	306
Ifa45	–	3×96	115	137	162
Ifa46	AF010461	–	115	137	162
Ifa47[a]	–	1×96	115	~90	126[b]
Ifa49	–	–	115	254	162
Ifa50	–	–	115	173	198
Ifa51	–	1×96	115	137	234
Ifa52	–	–	115	263	162
Ifa54	–	2×96	115	137	126[b]
K1	M59766	–	115	137	306
Wos10	U07010	–	115	137	198
Wos12	U16696	–	115	137	270
Wos3	U07003	–	115	137	234
Wos6	U07006	3×96	115	137	126[b]
Wos7	U07007	2×96	115	137	162

[a]Cross-over.
[b]Deletion.

Due to variation in the copy number of a 36 bp repeat, between one and 7, the largest *Hin*fI fragment in FC27-type repeats varies from 162 bp in Ifa46 to 378 bp in Ifa4. Lanes Ifa46 to K1 in Fig. 1 represent a selection of alleles, which differ only in the number of the 36 bp repeats present. The smallest allele, Ifa46, has one copy of the repeat, and the longest allele, K1, has 5 copies. A 96 bp *Hin*fI fragment (as seen in Fig. 1, lanes Wos6 and D10) represents 2 or more copies of the 96 bp repeat unit.

In many FC27-type *msp2* alleles from Tanzania, a third repeat type was discovered, which increased the size of the family-specific 137 bp *Hin*fI fragment at the 3′ end of the amplification product (IRION *et al.*, 1997). Substituting for the 137 bp fragment, new fragment sizes were generated by different copy numbers of this 9 bp repeat encoding the amino acids asparagine (N), alanine (A) and proline (P). As a result, the size of this normally conserved fragment can range from 146 bp in Ifa1, 182 bp in Ifa13, and 200 bp in Ifa11 up to 335 bp in Ifa23 (see Fig. 1), which contains 23 copies of the NAP repeat.

3D7 family of msp2 *alleles*

*Hin*fI restriction digestion of a 3D7-type allele produces 2 conserved restriction fragments of 70 bp and 108 bp, which both derive from the 3′ end of the *msp2* nested PCR product. One further large *Hin*fI fragment is obtained, which varies between alleles. Individual alleles were named according to the size of this fragment, e.g. $3D7_{370}$. Other, unusual restriction patterns can also, though rarely, be found in this allelic family, mostly due to point mutations, creating new *Hin*fI restriction

Table 3. Details of 9 *msp2* genotypes of the *Plasmodium falciparum* 3D7 allelic family from Tanzania[a]

msp2 genotype	Genbank™ accession no.	*Hin*fI restriction fragments (no. of base pairs)
Ifa5	–	70, 108, ~550
Ifa6	–	51, 70, 108, ~310
Ifa7	AF033860	51, 70, 108, 386
Ifa8	–	51, 70, 108, ~400
Ifa20	–	70, 23, 85, ~370
Ifa24	AF033859	70, 108, 266
Ifa39	–	51, 70, 108, ~290
Ifa48	–	51, 70, 108, ~330
Ifa53	–	70, 108, ~250

[a]Only genotypes with unusual restriction fragments are listed.

sites. 3D7-type alleles with deviations from the conserved *Hin*fI pattern are listed in Table 3.

Recombinant forms between the two allelic families

Alleles which represent recombinations between the FC27 and the 3D7 families have also been found in Tanzanian blood samples. These recombinations are revealed in *Hin*fI restriction fragment analysis by their unusual fragment lengths. The *Hin*fI restriction digests of 2 recombinant alleles (Ifa18, Ifa42) are shown in Fig. 1. Nucleotide sequencing revealed that, in all examples analysed so far, the site of recombination was located at the 5′ end of the 96 bp repeat unit (IRION *et al.*, 1997). In all recombinations for which sequence data are accessible, the 5′ end was derived from a 3D7-type allele, and the 3′ end was of the FC27 type. PCR products of *msp2*, representing recombination between the 3D7 and the FC27 allelic families, are grouped with FC27-type alleles and listed in Table 2, because they contain mainly FC27-type sequences, in particular the FC27 family-specific 36bp and 96bp repeats.

Multiple infections in Tanzanian blood samples

To illustrate the use of our RFLP genotyping method for *P. falciparum* infection in Tanzanian blood samples, an example of a multiple infection (multiplicity of 4) is shown in Fig. 2. The photograph shows *Hin*fI restriction digests separated on a 10% PAA gel. The fifth lane represents a naturally-occurring mixed infection in a Tanzanian blood sample. The restriction patterns of all 4 individual alleles contained in this mixed infection are shown separately in lanes Wos7, Wos12, Wos10 and $3D7_{400}$.

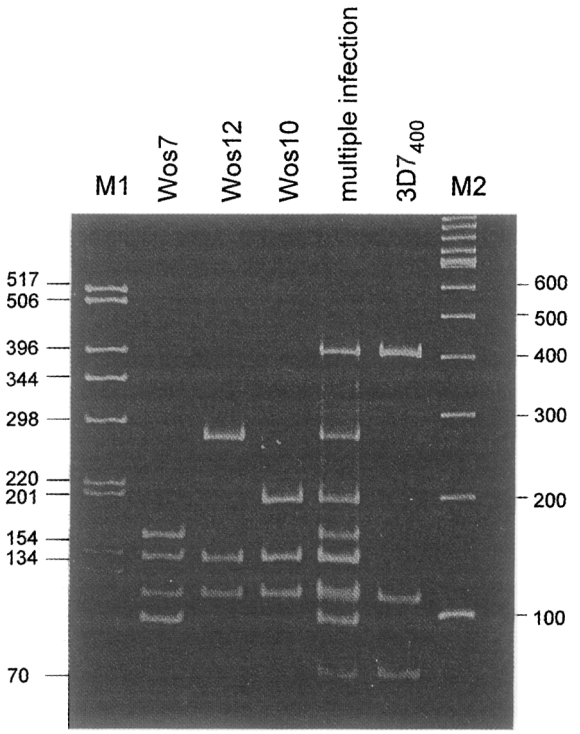

Fig. 2. *Hin*fI digest of *msp2* alleles of a *Plasmodium falciparum* multiple infection from a Tanzanian blood sample, which shows a multiplicity of 4 concurrent infections. In order to demonstrate how a complex restriction pattern is analysed, digests of corresponding alleles (Wos7–$3D7_{400}$) were loaded beside the RFLP pattern of the multiple infection. Lane M1: 1 μg of 1 kb ladder (Gibco BRL Life Technologies); lane M2: 1 μg of 100 bp DNA marker (Gibco BRL Life Technologies).

When analysing multiple infections, the intensity of ethidium bromide staining of individual bands in the PAA gel has to be considered. All *Hin*fI fragments of an allele show the same intensity, and can therefore be grouped together. Fragments which are shared by several alleles also appear as dominant bands, as can be seen in the mixed sample in Fig. 2. Here the 115 bp and 137 bp *Hin*fI fragments stood out as the strongest bands, because these 2 fragments were present in 3 of the 4 co-infecting alleles. Wos7 (Fig. 2) contained 3 copies of the 96bp repeat, resulting in two 96 bp *Hin*fI fragments, indicated by the increased intensity of staining of this fragment compared to the other *Hin*fI fragments of this allele.

Additional restriction digests increase the resolution of the genotyping scheme

In some studies, not only multiplicity and allelic family were to be determined, but it was also necessary to trace individual alleles in consecutive blood samples. For this purpose, a more complex genotyping technique was applied, involving 4 different restriction digests and side-by-side runs of the samples to be compared on the gel. Because the variable, allele-specific *Hin*fI fragment of 3D7-type alleles is too large (250–550 bp) to identify an allele unequivocally by gel electrophoresis, further restriction digests (*Dde*I, *Rsa*I, *Scr*FI) are necessary, which yield smaller and allele-specific fragment sizes. This allows further differentiation between 3D7-type alleles producing a *Hin*fI fragment of similar size (differences <10 bp). The conserved and variable restriction fragments resulting from *Rsa*I, *Dde*I and *Scr*FI digests of 3D7-type alleles have been described earlier (FELGER *et al.*, 1994). These additional restriction digests are generally done if infections from different blood samples need to be compared to each other. Further restriction digests, in addition to *Hin*fI, are also useful for the determination of recombinant, new, or 3D7-type alleles. These extra digests increase the discriminatory power of this genotyping scheme considerably and allow the identification of individual alleles.

Potential use of msp2 *genotyping by PCR–RFLP*

The potential of PCR–RFLP typing for direct comparison of blood samples is shown in Fig. 3. This example derives from a drug test, in which recrudescent and new infections were discriminated (IRION *et al.*, 1998). Each lane represents a consecutive blood sample from the same donor. In this example, baseline samples A and B (days 0 and 3 were defined as baseline) clearly showed RFLP patterns different from samples C and D. Sample A contained 4 different infections, 2 of which

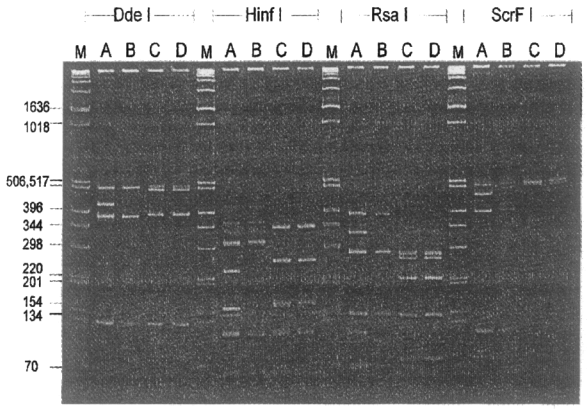

Fig. 3. Comparison of 4 subsequent blood samples (days A, B, C, and D) from a patient infected with *Plasmodium falciparum* and participating in a longitudinal drug efficacy study. A and B show genotypes of parasites at baseline, C and D represent parasites appearing after treatment. PCR products of *msp2* were digested with the restriction enzymes *Dde*I, *Hin*fI, *Rsa*I, and *Scr*FI. The resulting fragments were separated side by side on a 10% PAA gel. On days C and D, all baseline genotypes had disappeared, while 2 new genotypes were found. Lane M: 1 μg of 1 kb ladder DNA marker (Gibco BRL Life Technologies).

belonged to the FC27 family (Ifa1, Ifa4) and 2 were of the 3D7 type ($3D7_{320}$, $3D7_{430}$). The second baseline sample, B, showed the same alleles with the exception of Ifa1, which was missing. All baseline genotypes had disappeared from subsequent samples (C and D), which both contained 2 new and distinctively different infections (Ifa40, $3D7_{340}$). By conventional microscopy, these samples would have been identified as recrudescences and therefore treatment failures.

Comparison of PCR–RFLP with two-loci genotyping

In a previous study (BECK *et al.*, 1997), genotyping results were available for both loci, *msp2* genotyped by PCR–RFLP (BECK *et al.*, 1997) and *msp1* (W. Huber, personal communication), which was genotyped for allelic family by family-specific nested PCR amplification as described by HUBER *et al.* (1998). A comparison of multiplicity obtained from each locus made it possible to assess the gain in resolution by genotyping an additional locus. The analysis of *msp1* was based on the determination of allelic families only, size variations of the *msp1* PCR products not being considered. Only in 7 of 269 infections did the additional *msp1* genotyping improve the resolution obtained by the *msp2* PCR–RFLP method (Table 4). This increased resolution was due to infections defined as single clone infections by *msp2* PCR–RFLP typing, 7 of which were defined as mixed infections by *msp1* typing. In these cases, 2 different parasite clones shared the same *msp2* allele or, at least, produced the same RFLP pattern. However, in all cases of multiple simultaneous infections, additional *msp1* typing did not increase the multiplicity of infection determined by *msp2* PCR–RFLP genotyping.

Table 4. Numbers of concurrent infections detected by genotyping *msp1* and *msp2* of *Plasmodium falciparum* in 269 blood samples from Tanzania

	msp2								
	1	2	3	4	5	6	7	8	Total
msp1									
1	38	24	15	14	13	4	3	1	112
2	7[a]	15	20	17	32	13	16	23	143
3	0	0	0	1	3	4	3	3	14

[a]Samples in which genotyping for 2 loci resulted in a gain in the resolution of multiple infections.

Discussion

Genotyping *P. falciparum* field isolates is useful for a wide variety of applications. While we were mainly interested in detecting multiplicity of mixed clone infections and in studying infection dynamics, other research applications could include finding markers for virulence or drug resistance. Since only limited amounts of blood are available in epidemiological studies, amplification of a marker sequence by PCR is the method of choice. Potential genotyping markers are sequences showing polymorphism within a parasite population. Polymorphisms in both repetitive sequences and unique sequences have been used for genotyping (ROBERT *et al.*, 1996). When polymorphic repetitive sequences are amplified by PCR, a clone-specific pattern of multiple bands is seen after separating the PCR products on a gel. However, for the multiple concurrent infections normally found in areas of high malaria transmission, these banding patterns are superimposed upon each other and become too complex to analyse. A range of unique DNA sequences, suitable as polymorphic markers, is available in *P. falciparum* because of its highly diverse surface antigens. The most promising marker for genotyping is the one showing the most extensive polymorphism. We chose *msp2* as a marker gene for 2 reasons. (i) With 82 different alleles detected so far in Tanzania, *msp2* offers the genetic diversity necessary to allow single locus genotyping. The polymorphism is of such an extent that superinfection with the same genotype as the previous one is highly unlikely (SMITH *et al.*, 1999b). This fact allowed the use of *msp2* as a marker to discriminate between new *P. falciparum* infections and recrudescences in a drug trial (IRION *et al.*, 1998). (ii) Besides information on multiplicity of infection, further information about the biological role of MSP2 can be obtained from epidemiological studies which involve *msp2* as a marker gene, i.e., morbidity associations of certain *msp2* alleles or an allelic family.

When analysing genotyping data, it has to be remembered that there are intrinsic limitations of the PCR technique. It has to be noted that the number of concurrent infections detected by PCR is always a minimum estimate of the number of different parasite clones which are present in a carrier. Some clones might be missed owing to the detection limit of the PCR technique, which might miss infections of very low density, or to sequestration of parasites at the time of sampling, as shown by FÄRNERT *et al.* (1997). Furthermore, CONTAMIN *et al.* (1995) showed that, in multiple infections, the most abundant allele in a blood sample can suppress the amplification of an allele which is less abundant. However, statistical analyses of longitudinal data can be used to estimate the extent of such imperfect sensitivity (SMITH *et al.*, 1999b). Our knowledge of heterogeneities in densities of individual genotypes in multiple infections remains limited, and further development of quantitative genotyping techniques is needed to understand fully the interdependence in multiple clone infections.

We are also aware that alleles resulting from recombination between the allelic families remain undetected if no obvious alteration in the restriction patterns results. PCR amplification with mixed family-specific primers provided evidence for a high frequency of recombination (NTOUMI *et al.*, 1997). This was not observed in our studies. A further limitation of genotyping by RFLP consists in sequence diversity located outside the restriction sites. In order to test this, we have previously conducted single-strand conformation polymorphism (SSCP) experiments, using the FC27/D10 allele (defined by PCR–RFLP) from different isolates. We were not able to detect intra-allelic variation within 36 samples (FELGER *et al.*, 1997). In order to establish the full extent of sequence diversity in a marker gene, direct sequencing is necessary. However, this is not feasible in epidemiological studies, and direct sequencing is not possible if multiple infections are simultaneously present. A more detailed critical evaluation of PCR–RFLP genotyping has been made by SNOUNOU & BECK (1998).

In spite of the high resolution of PCR–RFLP genotyping of *msp2*, there is an obvious inherent limitation in typing only one locus, since 2 infections sharing the same *msp2* allele will be determined as one. However, by taking into account the allele frequencies, it is possible to estimate the true average number of concurrent infections with allowance for multiple infection with the same genotype (HILL & BABIKER, 1995). Analysing data from our Kilombero studies in this way, we estimated that the unadjusted infection multiplicity determined by PCR–RFLP genotyping of *msp2* only slightly underestimated the true number of infections. For in-

stance, in our largest study to date (FRASER-HURT *et al.*, 1999), the average number of genotypes detected was 2·93 while analysis with the method of HILL & BABIKER (1995) (assuming a conditional Poisson distribution for the multiplicity) gave an estimate of the true multiplicity of 3·01.

An alternative to the use of a very high resolution typing system for a single locus is to analyse several different loci with lower resolution for each individual locus. However, comparison of results obtained by PCR–RFLP genotyping of *msp2* with results from genotyping of *msp1* allelic families revealed that the additional locus did not contribute much to the determination of multiplicity. Typing *msp1* improved the resolution of PCR-based genotyping only in 2·6% of the samples, all of which appeared to be single infections with respect to *msp2*, but were shown to be double infections with respect to *msp1*. We therefore concluded that, in Tanzanian blood samples, with an average of 5 detected infections per child, the single locus PCR–RFLP typing using *msp2* has adequate discriminatory power.

There are some purposes for which multilocus typing is necessary, for instance studies of linkage disequilibrium, searching for markers of virulence or pathogenicity, or checking whether factors modifying parasite multiplicity or diversity selectively affect particular loci (e.g., locus-specific effects of a vaccine). HILL & BABIKER (1995) also described how the mean multiplicity in a population can be estimated from such multilocus typing data. Using this method, BABIKER *et al.* (1997) estimated the mean multiplicity in 53 samples from Michenga village in Tanzania to be 3·3, based on data generated by hybridization of oligonucleotide probes specific for allelic families to blotted PCR products of the 2 loci *msp1* and *msp2*. While this result was very similar to direct estimates of multiplicity for this area from PCR–RFLP genotyping of *msp2* alone (SMITH *et al.*, 1999a), the multilocus approach has several disadvantages when the objective is to relate the number of infections in a host to individual characteristics such as age, morbidity risks, or risk factors for infection. It is much more straightforward to analyse relationships of other variables with multiplicity when it is measured by a single number, and the HILL & BABIKER (1995) approach does not lend itself to providing estimates of multiplicity for individual carriers. Moreover, the estimates of multiplicity depend on assumptions about the population structure of the parasite.

The analysis of studies involving longitudinal tracking of individual infections is also much more straightforward if high resolution single-locus typing is used. Such applications include comparisons of baseline and recrudescent genotypes in vaccine or drug trials in order to distinguish treatment failure from reinfection (IRION *et al.*, 1998; SNOUNOU & BECK, 1998), and studies of infection dynamics (SMITH *et al.*, 1999b). High resolution typing of a single locus gives rise directly to patterns of appearance and disappearance of individual infections. In contrast, multi-locus genotypes of individual parasite clones cannot be identified in multiple infections, since the marker genes used to date are unlinked. Hence longitudinal studies using multilocus genotyping in areas of high parasite multiplicity can give rise to patterns which are difficult to interpret.

The extreme genetic polymorphism detected in particular in *msp2*, but also in other surface antigens of *Plasmodium* spp., raises further questions. For example, do individual alleles persist in time and space, or are new forms generated constantly *de novo*? A third possibility consists of a combination of both alternatives: a panel of defined conserved alleles plus newly generated ephemeral genotypes. Questions about the molecular evolution of *msp2* alleles have been addressed (CONWAY, 1997; FELGER *et al.*, 1997), but not yet fully answered. Nucleotide sequence comparisons between alleles of distant geographical origin indicated that alleles persist (DOBANO *et al.*, 1997). A large body of information on *msp2* genotypes is now available from the 4 studies summarized here, which we have conducted in Ifakara and the Kilombero valley, Tanzania, in different years, using the same PCR–RFLP method for genotyping. The diversity of *msp2* was similar in all 4 studies. This suggests that, though the genetic diversity is extensive, it is nevertheless limited by structural constraints or immune selection. This hypothesis has been previously proposed for the FC27-type alleles (FELGER *et al.*, 1997). Our molecular epidemiological studies (listed in Table 1) clearly showed that most of the alleles, in particular the most frequent ones, were found in all studies at comparable prevalence (data not shown). The fact that some rare alleles were not found in all studies might be due to sampling variation, which might miss alleles of low frequency. Alternatively, it could indicate that these rare alleles existed only for a short time. Even in studies comparing genotypes found in PNG and Tanzania, no major difference was observed, except for the group of new FC27-type alleles containing different numbers of a new 9 bp repeat unit, which was not detected in PNG. This extensive but restricted diversity might be due to functional constraints of MSP2 and might therefore help to elucidate the biological role of this molecule.

Acknowledgements

We thank Tom Smith for critical discussion, Jennifer Jenkins for her help in preparing the manuscript, and Werner Huber for providing *msp1* genotyping data. The investigation received financial support from the Swiss National Science Foundation grant no. 3200-045616.

References

Al-Yaman, F., Genton, B., Reeder, J., Anders, R., Smith, T. & Alpers, M. P. (1997). Reduced risk of clinical malaria in a highly endemic area in children infected with multiple clones of *Plasmodium falciparum*: a prospective community study. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **91**, 602–605.

Babiker, H. A., Ranford-Cartwright, L. C., Currie, D., Charlwood, J. D., Billingsley, P., Teuscher, T. & Walliker, D. (1994). Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitology*, **109**, 413–421.

Babiker, H. A., Lines, J., Hill, W. G. & Walliker, D. (1997). Population structure of *Plasmodium falciparum* in villages with different malaria endemicity in east Africa. *American Journal of Tropical Medicine and Hygiene*, **56**, 141–147.

Beck, H.-P., Felger, I., Huber, W., Steiger, S., Smith, T., Weiss, N., Alonso, P. L. & Tanner, M. (1997). Analysis of multiple *Plasmodium falciparum* infections in Tanzanian children during the trial of the malaria vaccine SPf66. *Journal of Infectious Diseases*, **175**, 921–926.

Contamin, H., Fandeur, T., Bonnefoy, S., Skouri, F., Ntoumi, F. & Mercereau-Puijalon, O. (1995). PCR typing of field isolates of *Plasmodium falciparum*. *Journal of Clinical Microbiology*, **33**, 944–951.

Conway, D. (1997). Natural selection on polymorphic malaria antigens and the search for a vaccine. *Parasitology Today*, **13**, 26–29.

Dobano, C., Khan, A., Robinson, J. V., Taylor, R. R. & McBride, J. S. (1997). Identical alleles of *Plasmodium falciparum* merozoite surface protein 2 found in distant geographic areas and times. *Parasitology International*, **46**, 137–142.

Engelbrecht, F., Felger, I., Genton, B., Alpers, M. & Beck, H.-P. (1995). *Plasmodium falciparum*: malaria morbidity is associated with specific merozoite surface antigen 2 genotypes. *Experimental Parasitology*, **81**, 90–96.

Färnert, A., Snounou, G., Rooth, I. & Björkman, A. (1997). Daily dynamics of *Plasmodium falciparum* subpopulations in asymptomatic children in a holoendemic area. *American Journal of Tropical Medicine and Hygiene*, **56**, 538–547.

Felger, I., Tavul, L. & Beck, H.-P. (1993). *Plasmodium falciparum*: a rapid technique for genotyping the merozoite surface protein 2. *Experimental Parasitology*, 77, 372–375.

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. *Experimental Parasitology*, **79**, 106–116.

Felger, I., Marshall, V. M., Reeder, J. C., Hunt, J. A., Mgone, C. S. & Beck, H. P. (1997). Sequence diversity and molecular evolution of the merozoite surface antigen 2 of *Plasmodium falciparum*. *Journal of Molecular Evolution*, **45**, 154–160.

Felger, I., Smith, T., Edoh, D., Kitua, A., Alonso, P., Tanner, M. & Beck,. H.-P. (1999). The epidemiology of multiple *Plasmodium falciparum* infections 6. Multiple *Plasmodium falciparum* infections in Tanzanian infants. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **93**, supplement 1, S1/29–S1/34.

Foley, M., Ranford-Cartwright, L. C. & Babiker, H. A. (1992). Rapid and simple method for isolating malaria DNA from fingerprick samples of blood. *Molecular and Biochemical Parasitology*, **53**, 241–244.

Fraser-Hurt, N., Felger, I., Edoh, D., Steiger, S., Mashaka, M., Masanja, H., Smith, T., Mbena, F. & Beck, H.-P. (1999). The epidemiology of multiple *Plasmodium falciparum* infections 9. Effect of insecticide-treated bed nets on haemoglobin values, prevalence and multiplicity of infection with *Plasmodium falciparum* in a randomized controlled trial in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **93**, supplement 1, S1/47–S1/51.

Genton, B., Al-Yaman, F., Reber, R., Anders, R., Brown, G., Saul, A., Mai, A., Stuerchler, D. & Alpers, M. (1996). Safety and immunogenicity of a three-component blood-stage vaccine (p[190, MSA-2, RESA) against *Plasmodium falciparum* malaria in adults living in a highly endemic area of Papua New Guinea (PNG). *Abstracts: XIVth International Congress for Tropical Medicine and Malaria, Nagasaki, Japan*, p. 76 (abstract no. B-21-4).

Hill, W. G. & Babiker, H. A. (1995). Estimation of numbers of malaria clones in blood samples. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **262**, 249–257.

Huber, W., Haji, H., Charlwood, J. D., Certa, U., Walliker, D. & Tanner, M. (1998). Genetic characterization of the malaria parasite *Plasmodium falciparum* in the transmission from the host to the vector. *Parasitology*, **116**, 95–110.

Irion, A., Beck, H.-P. & Felger, I. (1997). New repeat unit and hot spot of recombination in FC27-type alleles of the gene coding for *Plasmodium falciparum* merozoite surface protein 2. *Molecular and Biochemical Parasitology*, **90**, 367–370.

Irion, A., Felger, I., Abdulla, S., Smith, T., Mull, R., Tanner, M., Hatz, C. & Beck, H.-P. (1998). Distinction of recrudescences from new infections by PCR–RFLP analysis in a comparative trial of CGP 56 697 and chloroquine in Tanzanian children. *Tropical Medicine and International Health*, **3**, 490–497.

Ntoumi, F., Contamin, H., Rogier, C., Bonnefoy, S., Trape, J. F. & Mercereau-Puijalon, O. (1995). Age-dependent carriage of multiple *Plasmodium falciparum* merozoite surface antigen-2 alleles in asymptomatic malaria infections. *American Journal of Tropical Medicine and Hygiene*, **52**, 81–88.

Ntoumi, F., Mercereau-Puijalon, O., Ossari, S., Luty, A., Reltien, J., Georges, A. & Millet, P. (1997). *Plasmodium falciparum:* sickle-cell trait is associated with higher prevalence of multiple infections in Gabonese children with asymptomatic infections. *Experimental Parasitology*, **87**, 39–46.

Robert, F., Ntoumi, F., Angel, G., Candito, D., Rogier, C., Fandeur, T., Sarthou, J. L. & Mercereau-Puijalon, O. (1996). Extensive genetic diversity of *Plasmodium falciparum* isolates collected from patients with severe malaria in Dakar, Senegal. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **90**, 704–711.

Smith, T., Beck, H.-P., Kitua, A., Mwankusye, S., Felger, I., Fraser-Hurt, N., Irion, A., Alonso, P., Teuscher, T. & Tanner, M. (1999a). The epidemiology of multiple *Plasmodium falciparum* infections 4. Age dependence of multiplicity of *Plasmodium falciparum* infections and other malariological indices in an area of high endemicity. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **93**, supplement 1, S1/15–S1/20.

Smith, T., Felger, I., Fraser-Hurt, N. & Beck, H.-P. (1999b). The epidemiology of multiple *Plasmodium falciparum* infections 10. Effects of insecticide-treated bed nets on the dynamics of multiple *Plasmodium falciparum* infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **93**, supplement 1, S1/53–S1/57.

Smythe, J. A. Peterson, M. G., Coppel, R. L., Saul, A. J., Kemp, D. J. & Anders, R. F. (1990). Structural diversity in the 45 kilodalton merozoite surface antigen of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, **39**, 227–234.

Snounou, G. & Beck, H.-P. (1998). The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. *Parasitology Today*, **14**, 462–467.

Viriyakosol, S., Siripoon, N., Petcharapirat, C., Petcharapirat, P., Jarra, W., Thaithong, S., Brown, K. N. & Snounou, G. (1995). Genotyping of *Plasmodium falciparum* isolates by the polymerase chain reaction and potential uses in epidemiological studies. *Bulletin of the World Health Organization*, **73**, 85–95.