J Mol Evol (1997) 45:154–160

JOURNAL OF MOLECULAR EVOLUTION

Sequence Diversity and Molecular Evolution of the Merozoite Surface Antigen 2 of *Plasmodium falciparum*

Ingrid Felger,[1,5] Vikki M. Marshal,[2] John C. Reeder,[2] John A. Hunt,[3] Charles S. Mgone,[4] Hans-Peter Beck[5]

[1]Institut für Zellbiologie, Universität Witten-Herdecke, Stockumer Str. 10, 58448 Witten, Germany
[2]The Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Melbourne, Victoria 3050, Australia
[3]Department of Genetics, University of Hawaii, 1960 East West Road, Honolulu, HI 96822, USA
[4]PNG Institute of Medical Research, P.O. Box 60, Goroka, Papua New Guinea
[5]Swiss Tropical Institute, Socinstrasse 57, CH-4002 Basel, Switzerland

Received: 8 February 1995 / Accepted: 24 March 1997

Abstract. Eleven new alleles of the *Plasmodium falciparum* merozoite surface antigen 2 (*MSA2*) from Papua New Guinea were analyzed by direct sequencing of polymerase chain reaction (PCR) products. We have used the sequence information to trace the molecular evolution of *MSA2.* The repeats of ten alleles belonging to the *3D7* allelic family differed considerably in size, nucleotide sequence, and repeat copy number. In the repeat region of these new alleles, codon usage was extremely biased with an exclusive use of NNT codons. Another new allele sequenced belonged to the *FC27* family and confirmed the family-specific conserved structure of 96 and 36 bp repeats. In order to assess sequence microheterogeneity within samples defined as the same genotype by restriction fragment length polymorphism (RFLP), we have analyzed single-strand conformation polymorphism (SSCP) of different samples of the most frequent allele (*D10* of the *FC27* family) in the study population. No sequence heterogeneity could be detected within the repeat region. Based on analysis of the repeat regions in both allelic families, we discuss the hypothesis of a different evolutionary strategy being represented by each of the allelic families.

Kew words: Merozoite surface antigen 2 — Nucleotide sequence comparisons — Molecular evolution

Introduction

The merozoite surface antigen 2 (*MSA2*) of the malaria parasite *Plasmodium falciparum* is recognized by antibodies present in 90% of the population of a malaria-endemic area (Al-Yaman et al. 1994). *MSA2* is also considered a promising candidate for a blood stage subunit vaccine against malaria. The *MSA2* gene is highly polymorphic, showing considerable size and sequence variation (Smythe et al. 1991; Marshall et al. 1994; Prescott et al. 1994; Felger et al. 1994). The different alleles fall into two allelic families, the *FC27* and the *3D7* family, as defined by family-specific unique sequences, flanking a central repeat region. The central tandem repeats differ considerably between both families. The *FC27* family of alleles shows varying numbers of a structurally conserved 96 bp and a 36 bp repeat unit. In contrast, the repeat units of the *3D7* family are less conserved, they are highly variable in length (ranging from 12 bp to 30 bp), in copy number, and in sequence.

During a study on genetic diversity in a parasite population in Papua New Guinea, 38 different *MSA2* alleles were identified by polymerase chain reaction (PCR) amplification of 184 positive blood samples (Felger et al. 1994). Of these alleles, 30 belonged to the *3D7* family and eight to the *FC27* family. Individual alleles and their frequencies were determined by a PCR-RFLP genotyping scheme (Felger et al. 1993). Most of the alleles (26/38) were found only three times or less in the study population. All but one of the infrequent alleles belonged

Correspondence to: I. Felger at the Swiss Tropical Institute.

into the *3D7* allelic family, whereas alleles of moderate or high frequencies (represented 4–41 times in the study population) generally belonged to the *FC27* allelic family. Here we report the nucleotide sequence and a detailed analysis of the repetitive regions of 11 new *MSA2* alleles all deriving from the above population study. Ten sequenced alleles belonged to the *3D7* family, seven of these were of low allelic frequency (<0.02). One new allele of the *FC27* family was sequenced and aligned for sequence comparison with four new alleles already described and deriving from the same study population (Felger et al. 1994).

From these earlier studies on *MSA2* diversity in a natural parasite population, it remained unclear, due to limited resolution of RFLP, to what extent sequence heterogeneity existed within a certain RFLP-genotype. Therefore, in the present study we have analyzed the sequence variation within 36 different parasite isolates, all defined by PCR as *D10/FC27* genotype which is the most frequent one in the study area. As described earlier, interallelic sequence variation between different alleles of the *FC27* family is mostly contained within two *Hinf I* restriction fragments that comprise the repeat region (Felger et al. 1994). These two *Hinf I* fragments of samples genotyped as *FC27* were subjected to single-strand conformation polymorphism (SSCP) analysis, which is capable of detecting point mutations (Orita et al. 1989). A second question concerned the molecular evolution of the rare *MSA2* alleles which might throw light on the selective forces underlying the extensive allelic diversity. In our earlier cross-sectional population study, low allelic frequencies were mostly found in the *3D7* family of alleles. At the same time, this family is very diverse, comprising 3.8 times more members than the *FC27* family. This suggested differences in evolution rates and strategies or in molecular mechanisms of diversification between the two allelic families. Therefore, in the present study, the variable repeat regions of the newly identified and sequenced *3D7* alleles were analyzed in greater detail in order to trace their molecular evolution. For the analysis of the intragenic repeat region, we applied the model for repetitive pattern evolution proposed by Pizzi et al. (1990) and Frontali and Pizzi (1991). This method extracts underlying virtual repeats and allows the identification of supra repeats.

Methods

Parasite Collection

Parasites were derived from blood samples collected in the Wosera area, East Sepik Province, Papua New Guinea. The samples were collected in the course of the Malaria Vaccine Evaluation and Epidemiology Project, currently undertaken in the Wosera area, PNG (Alpers et al. 1992).

PCR-RFLP Genotyping

Isolation of parasite DNA from finger-prick blood samples was performed according to the rapid boiling method (Foley et al. 1992). PCR amplification of the *MSA2* gene, and genotyping of individual alleles were performed as described earlier (Felger et al. 1993, 1994).

DNA Sequencing

Alleles that showed a new PCR-RFLP pattern were sequenced as already described (Felger et al. 1994). The resulting 11 new nucleotide sequences have been submitted to EMBL/GenBank Data Libraries under accession numbers: U07001-2, U07004-5, U07008-9, U16695, U16696, and U16840-42.

Single-Strand Conformation Polymorphism (SSCP)

From 36 samples which were identified as *FC27* genotype by PCR-RFLP, 20% of the PCR product was digested with *Hinf I* (Promega) for 2 h at 37°C. Digests were run on 10% polyacrylamide gels. After ethidium bromide staining of the gel, two restriction fragments (96 bp and 137 bp in length) from each sample were excised from the gel and eluted overnight in 50 μl distilled water. Of the eluted fragments, 10% was mixed with 2 volume of a loading buffer containing 88% formamide, 10 mM EDTA (pH 8.0), and 0.01% bromphenol blue. The mixture was then incubated at 95°C for 5 min, snap cooled on wet ice for 5 min, and immediately loaded onto a 20% nondenaturing polyacrylamide gel containing 5% glycerol. Prior to sample loading, the gels were prerun for at least 1 h at 8 V/cm. The samples were then electrophoresed for 10 h at 10 V/cm using 0.5 x Tris Borate EDTA (TBE) buffer (pH 8.4), and bands were visualized by silver staining.

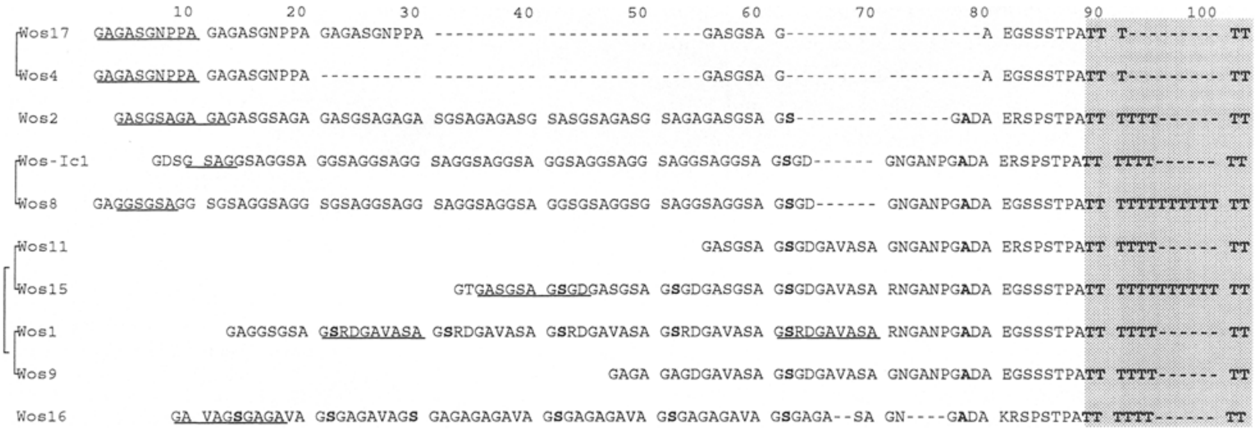
Data Analysis

Sequence alignments were done with Multiple Alignment Program (MAP) (Huang 1994). Codon usage was calculated by DNA Strider Program version 1.0. Synonymous versus non-synonymous substitutions were analyzed according to Li (1993).

Results

Codon Usage and Synonymous Versus Nonsynonymous Substitutions

The most striking feature of the new *MSA2* nucleotide sequences, found in all 10 alleles of the *3D7* family, was the unusual codon usage in the repeat region. The codon NNT was used exclusively in this region, whereas in nonrepeat regions of *MSA2*, usage of NNT codons was 46.8%. For coding regions of *P. falciparum,* in general, 39.6% NNT codons were determined (Saul and Battistutta 1988). Due to this exclusive use of NNT codons in the repeats of *3D7* alleles, no synonymous substitution was found in this region. But in the unique regions of the *3D7* alleles the ratio of synonymous versus nonsynonymous substitutions was about 1:1. By contrast, the alleles of the *FC27* family did not show any synonymous substitutions in neither the repeat nor the unique region.



```
                  10         20         30         40         50         60         70         80         90        100
Wos17    GAGASGNPPA GAGASGNPPA GAGASGNPPA ---------- ---------- ----GASGSA G--------- ---------A EGSSSTPATT T--------- TT

Wos4     GAGASGNPPA GAGASGNPPA ---------- ---------- ---------- ----GASGSA G--------- ---------A EGSSSTPATT T--------- TT

Wos2       GASGSAGA GAGASGSAGA GASGSAGAGA SGSAGAGASG SASGSAGASG SAGAGASGSA GS-------- ------GADA ERSPSTPATT TTTT------ TT

Wos-Ic1        GDSG SAGGSAGGSA GGSAGGSAGG SAGGSAGGSA GGSAGGSAGG SAGGSAGGSA GSGD------ GNGANPGADA ERSPSTPATT TTTT------ TT

Wos8     GAGGSGSAGG SGSAGGSAGG SGSAGGSAGG SAGGSAGGSA GGSGSAGGSG SAGGSAGGSA GSGD------ GNGANPGADA EGSSSTPATT TTTTTTTTTT TT

Wos11                                                               GASGSA GSGDGAVASA GNGANPGADA ERSPSTPATT TTTT------ TT

Wos15                                       GTGASGSA GSGDGASGSA GSGDGASGSA GSGDGAVASA RNGANPGADA EGSSSTPATT TTTTTTTTTT TT

Wos1                GAGGSGSA GSRDGAVASA GSRDGAVASA GSRDGAVASA GSRDGAVASA GSRDGAVASA RNGANPGADA EGSSSTPATT TTTT------ TT

Wos9                                                          GAGA GAGDGAVASA GSGDGAVASA GNGANPGADA EGSSSTPATT TTTT------ TT

Wos16            GA VAGSGAGAVA GSGAGAVAGS GAGAGAGAVA GSGAGAGAVA GSGAGAGAVA GSGAGA--SA GN----GADA KRSPSTPATT TTTT------ TT
```

Fig. 1. Alignment of 10 amino acid sequences translated from nucleotide sequences of new *Wos* alleles of the *3D7* allelic family of *MSA2.* A bracket indicates closest similarity between alleles. The repeat unit of each allele is underlined. The poly-Threonine stretch is highlighted by shading. Alternative codons for Serine and Alanine are printed in bold face.

Sequence Comparisons

All but one of the *3D7 Wos* alleles contained novel repeat units which are underlined in the amino acid alignment shown in Figure 1. Only *WOS-Ic1* shared the same repeat with an earlier described allele, *Indochina 1* (Smythe et al. 1990), but both sequences differed by three nucleotide substitutions outside the repeats. The nucleotide sequences of the repeats were aligned with the virtual repeat sequence GGT GCT, proposed by Smythe et al. (1991) (alignment not shown). When a consensus sequence was extracted from the repeat tract of the *3D7 Wos* alleles by using the predominant nucleotide at each position of the alignment, this generally agreed with a tandem array of the proposed virtual repeat unit. Thus, on the nucleotide level, the hexamer units were still obvious in spite of few substitutions (mostly transitions from G to A at the first position of the hexamer), while at the amino acid level, new repeat units appeared, which differed considerably in length and sequence between alleles.

The poly-threonine stretch, which is characteristic for all alleles of the *3D7* family, is located downstream of the repeats and it represents a further example for sequence conservation at nucleotide level only. The nucleotide sequence of a stretch of 5–14 threonine residues revealed underlying 9 bp repeats of the sequence ACT ACC ACA, each of the three triplets encoding threonine. This could explain why the interallelic length variation in the poly-threonine stretch occurred always in multiples of 9 bp. A further particular site-specific codon usage was found in nonrepetitive regions next to *3D7*-type repeats. An alignment of these polymorphic sequences (shown in Fig. 1) was facilitated by the presence of a specific codon at equivalent positions in the alignment. These sites were conserved between alleles and codons used at these sites differed from those used within the repeat regions. For example, the codon GCA is used for alanine at position 78 (boldface in Fig. 1) as opposed to the codon GCT which is strictly used within the repeat region. Similarly, the codon AGT was frequently used for serine in the repeat region, while the codon TCT (bolded in Fig. 1, position 62) was always used downstream of the repeats in all *3D7 Wos* alleles sequenced. This conserved use of certain codons at specific positions can throw light on the homologies and evolution of the *Wos* alleles.

Analysis of the *Wos1* repeats provides an example of how repeat sequences may possibly have evolved. In the *Wos1* repeat unit (GSRDGAVASA), the first serine is encoded by the nucleotides TCT and the second by AGT. The complete nucleotide repeat unit (GGT TCT CGT GAT GGT GCT GTT GCT AGT GCT) is conserved throughout the Wos1 repeats and is also present in a relatively conserved region of 10 amino acids immediately 3′ to the repeats (positions 61–70 of the amino acid alignment, Fig. 1). The same conserved region, with the exception of a single amino acid substitution in Wos1, is also present in the alleles *Wos11, 15,* and *9.* Therefore, the *Wos1* repeat region may have evolved from this sequence unit by amplification, and consequently, the alternative codon for serine (TCT) is present in all repeats of *Wos1* at equivalent position. The fact that this conserved region is also found in other alleles indicates that this sequence might be ancestral.

The similarity in the region between the *3D7*-like repeats and the poly-threonine stretch allowed the subgrouping of the alleles into closely related pairs, indicated by brackets in Figure 1. The four alleles *Wos1, Wos9, Wos15,* and *Wos11* were grouped together because of their similarity in the region downstream of the repeats. *Wos4* and *Wos17* showed identical sequences except a 30-bp deletion in *Wos4,* but both differed from the other alleles considerably. *Wos4* could have evolved from *Wos17* by deleting one repeat unit, or *Wos17* could have derived from *Wos4* by repeat amplification.

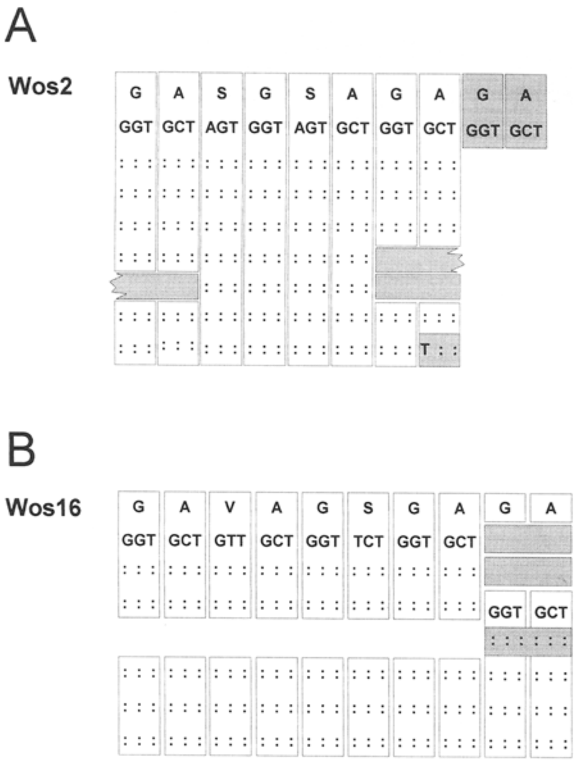

A
Wos2
G A S G S A G A G A
GGT GCT AGT GGT AGT GCT GGT GCT GGT GCT
T
B
Wos16
G A V A G S G A G A
GGT GCT GTT GCT GGT TCT GGT GCT
GGT GCT

Fig. 2. Repeat fidelity within two alleles of the *3D7* family, *Wos2* **(A)** and *Wos16* **(B).** Dots indicate identity in nucleotide sequence with the respective position in the first repeat unit. Sites of hypothetical mutation events are shaded. **(A)** *Wos2* contains several incomplete repeats. The scrambled pattern disappears by aligning the repeats with each other and hypothetical deletions and duplications visualize as gaps. The triplet at the end of the repeats encodes serine by the alternative codon ACT which is also found in other *Wos* alleles at equivalent position. **(B)** *Wos16* shows four long and two short versions of its repeat unit, which could reflect a transition state in repeat generation and homogenization.

When analyzing possible causes of repeat diversity, sequences with imperfect nucleotide repeats are of major interest. As an example, Figure 2 shows individual repeat units from the same allele aligned with each other. In *Wos2* (Fig. 2A), three out of seven repeat units showed elongation or deletion. Imperfect repeats can be caused by DNA polymerase slippage. Duplications of hexamers can be seen in *Wos16* (Fig. 2B). It is noteworthy that all rearrangements involved a complete hexamer (''virtual'' repeat unit) or a multiple of it, but never a single triplet. The observed modifications might be due to recent mutations, which disrupt originally perfect tandem repeats. This leads to the generation of a new repeat unit which then would be subject to sequence homogenization processes. *Wos16* in Figure 2B shows two versions of a repeat which differed in length by one hexamer. Four long and two short copies of the *Wos16* repeat unit were present. A further duplication of the additional hexamer was present following the first full-length repeat. This repeat structure might reflect a transient situation with incomplete homogenization, where either the short or the long repeat unit could eventually reach fixation.

Sequence Conservation in Repeats of the FC27 *Allelic Family*

The repeat regions of all compared alleles of the *FC27* family consist of one or more copies of 96 bp and 36 bp units, thus displaying a more conserved organization than in the *3D7* allelic family. We were interested in sequence diversity between repeat units within the same or between different alleles. For this comparison, the Wos repeats were aligned with repeat variants of other alleles of the *FC27* family (UNDP/WHO/TDR database). As a result, Figure 3 shows the few positions at the 5′ end of the repeats where nucleotide substitutions occurred, all leading to nonsynonymous changes. This observation of limited variability in the repeats of the *FC27* allelic family led to the question, whether any sequence variation exists within a group of samples which were determined by PCR-RFLP to be the same allele. Two adjacent *Hinf I* fragments containing the repeat region were isolated from 36 samples of the *FC27* genotype and were subjected to SSCP analysis. This technique allows the detection of mutated alleles. No mobility shift was observed in the samples tested.

Discussion

Analysis of other polymorphic *Plasmodium* surface antigens, such as MSA1 and CS protein, showed that repeats can be derived from an original ''virtual'' repeat unit (Frontali and Pizzi, 1991; Frontali 1994). We confirmed that this is also the case for the *MSA2* alleles of the *3D7* family, with a hexamer (GGT GCT) emerging as an underlying basic repeat unit. The exclusive presence of NNT codons in *3D7*-type repeats clearly causes a strong bias in codon usage. This can contribute considerably to antigenic diversity, but would not necessarily reflect selection acting at protein level to increase antigenic diversity. We suggest that the exclusivity of NNT codons results from homogenization mechanisms at nucleotide sequence level to maintain DNA sequence homology in the tandem repeats.

Homogenization mechanisms, such as unequal crossing over and biased gene conversion (Arnot et al. 1988), have been postulated to account for extreme fidelity in tandem repeats. These proposed mechanisms act at the DNA level only, and selection acting on the amino acid sequence would not maintain the observed sequence homogeneity between repeats. The position-specific use of different codons in the poly-threonine stretch might represent an operation of the same homogenizing mechanism (C. Frontali, personal communication).

The major sources of interallelic diversity in the *3D7* allelic family seem to be replication slippage causing either duplication or deletion of a hexamer or multiples of it, and mitotic or meiotic unequal exchange which can also increase or reduce the copy number of repeats. Such

96-base pair repeat

```
                  1       10         20         30         40         50
consensus         GCTGATACTA TTGCTAGTGG AAGTCAAAGG AGTACAAATA GTGCAAGTAC
substitution      :::A:::::G :::T:::G:T :::::::::T :::::::::: ::::::::::
numbers observed     1     6    1   5 6          9
                          60         70         80         90     96
consensus         TAGTACTACT AATAATGGAG AATCACAAAC TACTACTCCT ACCGCT
substitution      :::::::::: :::::::::: :::::::::: :::::::::: :T::::
numbers observed                                               1
```

36-base pair repeat

```
                  1       10         20         30    36
consensus         GAAAGTAATT CACCTTCACC ACCCATCACT ACTACA
substitution      A::::::T:: :::G:::::: :::::::::: ::::::
numbers observed  10     1      7
```

Fig. 3. Sequence comparison between repeats of the *FC27* allelic family. The consensus sequence, the positions of nucleotide substitutions (shaded), and the observed numbers of substitution are shown for the 96 bp and the 36 bp repeat. There are four positions of unique substitution, and six positions where a defined substitution appeared more than once (five to ten times). For analysis of the 96 bp repeat unit, 26 different repeats deriving from 13 *MSA2* alleles were aligned. Then, 24 repeats were analyzed for the 36 bp repeat unit. Dots indicate sequence identity between the consensus sequence and all other repeat sequences compared. The compared repeats derived from the following *MSA2* alleles: *Wos12* (this article), *Wos3, Wos6, Wos7,* and *Wos10* (Felger et al. 1994); *FC27-D10* (Smythe et al. 1988); *K1* (Smythe et al. 1991); *Col5* (Snewin et al. 1991); *Oks1* and *Oks11* (Marshall et al. 1994); and *N70, N71,* and *N92* (Prescott et al. 1994).

mutations occur at random, and the question has been raised whether the distribution of new repeat sequences in a population is a stochastic process, where each variant has the same chance of being eliminated or fixed. The concept of concerted evolution and molecular drive (Dover 1982) explains how neutral variants of repeat units can become abundant and be driven through a population by gene conversion, even in the absence of selection. But it also has to be considered that the intragenic repeat regions of *MSA2* are targets of the immune response, and, therefore, selective forces may act on these repeats, as for example, frequency-dependent selection. It has been discussed that frequency-dependent immune selection can favor rare alleles, thus maintaining antigenic polymorphism in *MSA2* (Conway 1997). Our findings do not answer the question whether changes in allele frequencies are due to frequency-dependent selection or random genetic drift.

In contrast to the situation in the *3D7* allelic family, a picture of strong conservation in repeat structure and nucleotide sequence emerged from the analysis of the *FC27* family. The lack of RFLP size variation within the two repeat units specific for the *FC27* family, already indicated sequence conservation. This was confirmed by SSCP analysis of fragments containing *FC27*-type repeats, where no variation was detected within nucleotide sequences from different samples of the same genotype. Because of both the importance of *MSA2* as a vaccine candidate molecule and the finding of immunodominant epitopes in the *FC27*-type repeats (Rzepczyk et al. 1990, 1992), interallelic sequence comparisons between different variants of the two *FC27* family repeat types were performed, and substitutions were located at few positions only. These restricted variable sites could reflect structural or selective constraints of the protein. Our results led us to speculate that purifying selection is acting on the repeat region, maintaining the two repeat structures. Diversification apparently can only take place by variation of repeat copy number of single-base changes at a few defined sites. The observed substitutions all resulted in nonsynonymous changes and thus might well be subject to positive selection. An increased rate of nonsynonymous nucleotide substitutions or the presence of only nonsynonymous substitutions, as found in both, repeats and nonrepetitive parts in the *FC27* family, is taken as evidence for positive selection (Hughes 1991). No biased codon usage was found in *FC27*-type repeats which could account for the exclusive nonsynonymous substitutions. But homogenization mechanisms also being active in *FC27*-type repeats have to be considered. Thus, the conclusion of positive selection cannot be drawn alone from the finding of only nonsynonymous substitutions in the repeats. Yet, the nonrandom distribution of substitution sites might indicate an immunological significance. In another polymorphic surface antigen of *Plasmodium falciparum,* the circumsporozoite protein, most nonsynonymous substitutions were found to be clustered in a region containing T cell epitopes (Good et al. 1988). As shown by epitope mapping, both repeat units of the *FC27* alleles contain immunogenic epitopes (Rzepczyk et al. 1990, 1992). Monoclonal antibodies against an epitope consisting of four residues, STNS, located in the center of the 96 bp repeat, can protect mice against plasmodial infection (Epping et al. 1988). This STNS epitope is not involved in any of the observed nucleotide substitutions. However, the T- and B-cell epitope *MSA2/2* (Rzepczyk et al. 1990, 1992) extends to the 5′ end of the 36 bp repeat and spans all three possible sites of substitutions in the 36 bp repeat. Thus, the presence of only nonsynonymous substitutions *FC27*-like alleles causes limited diversity in the amino acid sequence and might alter antigenicity, provided the repeats are indeed immunologically relevant in natural infections. The degeneracy at defined and clustered positions in the *FC27*-type repeats could be important for the parasite in order to evade the human immune response.

The SSCP analysis also resolved the question of whether the PCR-RFLP genotyping technique failed to detect additional sequence variation. Since in none of the tested samples a mobility change was detected, we assume that no sequence variation exists. However, it has to be remembered that although the SSCP method is a useful tool when screening for point mutations, identical mobility pattern cannot prove the absence of nucleotide substitutions. The final proof is given only by nucleotide sequencing of each individual allele.

The pronounced dichotomy in repeat organization between the two allelic families led us to speculate that the two allelic families could follow different evolutionary strategies with respect to their repeat regions. The *3D7* allelic family is highly diverse in sequence and structure of its repeats. The underlying virtual repeat units seem to allow high turnover and continuous rearrangements with consequent repeat homogenization. The strategy of the *3D7* family could be, therefore fast and random diversification resulting in a high number of infrequent alleles. This is consistent with our earlier findings in a cross-sectional population study. From our analysis it remains open, if and how selection acts on variants of the *3D7*-type repeats, since we were not able to find evidence for positive selection. In contrast, the repeats of the less diverse *FC27* family are likely to be subject to selective constraints. Diversification in *FC27*-type repeats seems to be limited and takes place within the boundaries of a conserved repeat structure at few hypervariable sites.

More immunological information (localization of epitopes) and molecular epidemiological studies might eventually help to clarify the selection modes and possibly diversifying mechanisms.

Acknowledgments. This article includes information obtained from the UNDP/World Bank/WHO-TDR malaria sequence database. We thank the people from Kunjingini for donating blood and L. Tavul and S. Kabintik for technical assistance. We are grateful to Drs. R. Anders and C. Frontali for stimulating discussion. The project was supported by the United States Agency for International Development Grant 9365967.89 and the European Community Research Program Grant TS3-CT92-0103.

References

Al-Yaman F, Genton B, Anders RF, Falk M, Triglia T, Lewis D, Hii J, Beck H-P, Alpers M (1994) Relationship between humoral response to merozoite surface antigen 2 and malaria morbidity in a highly endemic area of Papua New Guinea. Am J Trop Med Hyg 51:593–602

Alpers MP, Al-Yaman F, Beck H-P, Bhatia KK, Hii J, Lewis DJ, Paru R, Smith T (1992) The Malaria Vaccine Epidemiology and Evaluation Project of Papua New Guinea: rationale and baseline studies. PNG Med J 35:285–297

Arnot DE, Barnwell JW, Stewart MJ (1988) Does biased gene conversion influence polymorphism in the circumsporozoite protein-encoding gene of *Plasmodium vivax?* Proc Natl Acad Sci USA 85:8102–8106

Arnot D (1989) Malaria and major histocompatibility complex. Parasitol Today 5:138–143

Conway DJ (1997) Natural selection on polymorphic malaria antigens and the search for a vaccine. Parasitol Today 13:26–29

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Epping RJ, Goldstone SD, Ingram LT, Upcroft JA, Ramasamy R, Cooper JA, Bushell GR, Geysen HM (1988) An epitope recognised by inhibitory monoclonal antibodies that react with a 51 kilodalton merozoite surface antigen in *Plasmodium falciparum.* Mol Biochem Parasitol 28:1–10

Felger I, Tavul L, Beck HP (1993) *Plasmodium falciparum:* a rapid technique for genotyping the merozoite surface protein 2. Exp Parasitol 77:372–375

Felger I, Tavul L, Kabintik S, Marshall V, Genton B, Alpers M, Beck HP (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

Foley M, Ranford-Cartwright LC, Babiker HA (1992) Rapid and simple method for isolating malaria DNA from fingerprick samples of blood. Mol Biochem Parasitol 53:241–244

Frontali C, Pizzi E (1991) Conservation and divergence of repeated structures in *Plasmodium* genomes: the molecular drift. Acta Leid 60(1):69–81

Frontali C (1994) Genome plasticity in *Plasmodium.* Genetica 94:91–100

Good MF, Pombo D, Quakyi IA, Riley EM, Houghten RA, Menon A, Alling DW, Berzofsky JA, Miller LH (1988) Human T-cell recognition of the circumsporozoite protein of *Plasmodium falciparum:* immunodominant T-cell domains map to the polymorphic regions of the molecule. Proc Natl Acad Sci USA 85:1199–1203

Huang X (1994) On global sequence alignment. Computer Applications Biosci 10(3):227–235

Hughes AL (1991) Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics 127:345–353.

Hughes AL (1992) Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum.* Mol Biol Evol 9:381–393

Hughes MK, Hughes AL (1995) Natural selection on *Plasmodium* surface proteins. Mol Biochem Paras 71:99–113

Li WH (1993) Unbiased estimation of the rate of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Marshall VM, Anthony RL, Bangs MJ, Purnomo, 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 Parasiol 63(1):13–21

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 86:2766–2770

Pizzi E, Liuni S, Frontali C (1990) Detection of latent sequence periodicities. Nucl Acid Res 18(3):3745–3752

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 Paras 63:203–212

Rzepczyk CM, Csurhes PA, Lord R, Matile H (1990) Synthetic peptide immunogens eliciting antibodies to *Plasmodium falciparum* sporozoite and merozoite surface antigens in H 2b and H 2k mice. J Immunol 145:2691–2696

Rzepczyk CR, Csurhes PA, Saul AJ, Jones GL, Dyer S, Chee D, Goss

N, Irving DO (1992) Comparative study of the T cell response to two allelic forms of a malarial vaccine candidate protein. J Immunol 148:1197–1204

Saul A, Battistutta D (1988) Codon usage in *Plasmodium falciparum.* Mol Biochem Parasitol 27:35–42

Smythe JA, Coppel RL, Brown GV, Ramasamy R, Kemp DJ, Anders RF (1988) Identification of two integral membrane proteins of *Plasmodium falciparum.* Proc Natl Acad Sci USA 85:5195–5199

Smythe JA, Peterson MG, Coppel RL, Saul AJ, Kemp DJ, Anders RF (1990) Structural diversity in the 45 kilodalton merozoite surface antigen of *Plasmodium falciparum.* Mol Biochem Parasitol 39(2): 227–234

Smythe JA, Coppel RL, Day KP, Martin RK, Oduola AMJ, Kemp DJ, Anders RF (1991) Structural diversity in the *Plasmodium falciparum* merozoite surface antigen 2. Proc Natl Acad Sci USA 88: 1751–1755

Snewin VA, Herrera M, Sanchez G, Scherf A, Langsley G, Herrera S (1991) Polymorphism of the alleles of the merozoite surface antigens MSA1 and MSA2 in *Plasmodium falciparum* wild isolates from Colombia. Mol Biochem Parasitol 49:265–276