MAJOR ARTICLE

# Longitudinal Assessment of *Plasmodium falciparum* *var* Gene Transcription in Naturally Infected Asymptomatic Children in Papua New Guinea

**Mirjam Kaestli,[1] Alfred Cortes,[3,a] Moses Lagog,[3] Michael Ott,[2] and Hans-Peter Beck[1]**

[1]Swiss Tropical Institute and [2]F. Hoffmann-La Roche AG, Basel, Switzerland; [3]Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea

**Sequestration and antigenic variation are essential for *Plasmodium falciparum* survival in vivo contributing to severe pathologic findings and, also, chronic infection. Both are conferred by *P. falciparum* erythrocyte membrane proteins encoded by ~60 *var* genes. To study the dynamics of *var* gene expression, we conducted a 4-month longitudinal study of semi-immune children from Papua New Guinea. By use of magnetic bead–anchored reverse-transcription polymerase chain reaction analysis performed over 5 *var* regions, as well as cloning and sequencing, the longitudinal distribution of full-length *var* transcripts was analyzed. We identified a dynamic picture of *var* gene expression with rapid switches but with identical *var* transcripts recurring for up to 10 weeks. The number of *var* transcripts was correlated to the number of infections, with a mean of 1.7 *var* transcripts identified per sample and infecting strain. Analysis of 158 different Duffy binding–like 1$\alpha$ sequences confirmed the recombinogenic nature of *var* genes. This is the first report of the dynamics of *var* gene expression in chronically infected children.**

Cytoadherence and sequestration of *Plasmodium falciparum*–infected red blood cells (RBCs) are considered to be among the most important factors associated with the pathogenicity and virulence of *P. falciparum* malaria. Cytoadherence is mediated by the polymorphic *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), which is located on the surface of infected RBCs. PfEMP1 is encoded by 1 of ~60 *var* genes, each of which is 8–14 kb in length [1–3]. These large proteins of 200–350 kDa mediate binding to various cell surface receptors (reviewed in [4]). Cytoadherence is thought to prevent spleen-dependent killing, but it also has been shown that binding of PfEMP1 to CD36 presented on dendritic cells down-regulates dendritic cell activity, suggesting an immune modulatory role for PfEMP1 [5]. PfEMP1 has also been implicated in rosetting [6], is a target of naturally occurring immune responses, and shows antigenic variation [7, 8]. The expression of different PfEMP1 variants is accompanied by changes in the adhesive phenotype of infected RBCs [7].

PfEMP1 proteins are structured into several semi-conserved domains—namely, an N-terminal segment (NTS); various Duffy binding–like (DBL) domains; a cysteine-rich interdomain region (CIDR); in some instances, a "constant 2" (C2) region; a transmembrane domain; and the conserved, C-terminal acidic terminal segment (ATS), which represents the intracellular part of PfEMP1 that anchors the protein to the cytoskeleton (reviewed in [9]). Different subclasses of DBL domains ($\alpha$–$\varepsilon$ and x) and CIDR domains ($\alpha$–$\gamma$) have been identified. The most N-terminal DBL1$\alpha$ and CIDR1$\alpha$ form the conserved head structure of the protein. This head structure is found in almost all PfEMP1 molecules. Binding has been associated with various domains of PfEMP1 [6, 10–15], such as DBL1$\alpha$ to CR1 (in rosette formation), CIDR1$\alpha$ to CD36, or DBL$\beta$-C2 to intracellular adhesion molecule 1 (ICAM1). Binding to chondroitin sulphate A in samples from placental ma-

Received 5 September 2003; accepted 26 October 2003; electronically published 29 April 2004.

Financial support: Swiss National Science Foundation (grants 031-059 064 and 031-67211.01).

Present affiliation: MRC National Institute for Medical Research, Mill Hill, London, United Kingdom.

Reprints or correspondence: Dr. Hans-Peter Beck, Swiss Tropical Institute, Socinstr. 57, CH-4051 Basel, Switzerland (hans-peter.beck@unibas.ch).

**The Journal of Infectious Diseases 2004;189:1942–51**



0022-1899/2004/18910-0020$15.00

laria has been shown to occur not only with DBL-γ but, also, with CIDRα [12–14].

Most *var* genes can be classified into 3 groups, on the basis of their different but, within each group, highly conserved 5′ upstream sequences [16, 17]. The majority of *var* genes are located subtelomerically and possess a upsB-type upstream region. These *var* genes are located in a region that is highly recombinogenic with various repetitive elements that support ectopic recombination by clustering of heterologous subtelomeric stretches (including *var* genes) [18]. Another small set of *var* genes is arranged in chromosome internal clusters; these genes possess upsC-type upstream regions. A third group of *var* genes consists of subtelomerically located *var* genes, which are transcribed toward the telomeres (upsA-type regions [17]). That *var* gene transcription of the upsB- and upsC-type regions is regulated differently [19] raises questions about the different functional features of these 2 *var* gene groups.

The detailed mechanism of *var* gene regulation is still unknown, but it has been shown that every *var* gene represents a single transcriptional unit that is capable of in situ activation involving epigenetic mechanisms [20, 21]. The *var* genes are transcribed for the first 24 h of the erythrocytic cycle [4]. Whereas relaxed transcription of multiple truncated *var* genes was found at the ring stage, only one full-length *var* transcript was found at the early trophozoite stage [22]. A switching rate of 2.4%/generation was calculated in vitro [23], but little is known about *var* gene expression and switching in vivo. In a longitudinal study involving nonimmune adults who were artificially infected with the 3D7 laboratory-adapted strain, Peters et al. [24] showed that the first transcribed *var* gene in the erythrocytic stage was identical in different adults. The initial switching rate was estimated to be 16%, but it decreased thereafter, suggesting different switching rates for different *var* genes and host conditions.

In the present study, we describe *var* gene expression in naturally infected semi-immune children from Papua New Guinea during a 4-month period. The longitudinal distribution and structure of the expressed *var* transcripts were analyzed by reverse-transcription polymerase chain reaction (RT-PCR), cloning, and sequencing. We also analyzed *var* transcripts, with respect to the chromosomal location using the conserved 5′ upstream regions for amplification. Using this approach, we describe the dynamic nature of *var* gene expression in several asymptomatic children. Many *var* genes were transcribed simultaneously with switches at short intervals, but some identical transcripts recurred in the same child, even after 10 weeks.

## MATERIALS AND METHODS

***Study area and collection of blood samples.*** From April to August 2001, the period of transition from the wet season to the dry season, we performed a longitudinal study at the Maiwara Primary School on the Madang North Coast in Papua New Guinea. In this region, where malaria with perennial transmission is endemic, infections with *P. falciparum* and *P. vivax* are common. *P. malariae* and *P. ovale* are also present in this area.

Written, informed consent was obtained from the children's parents or guardians. The study was approved and ethical clearance was given by the Medical Research Advisory Committee of Papua New Guinea. According to national treatment guidelines, children with parasites but without malaria symptoms were not treated.

After informed consent was obtained from parents, we obtained blood samples, by fingerprick (0.2–0.5 mL) or venopuncture (2 mL), from 11 children who had asymptomatic *P. falciparum* infections. The children were 8–10 years of age. Samples were obtained from 8 children every 2 weeks for 4 months and from 3 children every 5 days for 1 month.

***Assessment of* P. falciparum *infections.*** Giemsa-stained blood slides were analyzed by microscopy. For blood samples that were found to be positive, by microscopy, for *P. falciparum,* the number of *P. falciparum* infections was determined by *msp2* genotyping, as described elsewhere [25]. In brief, 30 μL of full blood was spotted on filter papers (Isocode Stix; Schleicher & Schuell) and was dried for 20 min at 80°C. After washing, *msp2* PCR was performed directly on the filter papers, and restriction fragment–length polymorphism (RFLP) analysis of nested PCR products was used to record the number of infecting strains.

***Isolation of full-length* var *transcripts and RT-PCR.*** Total RNA was extracted using TRIzol (Invitrogen), according to the manufacturer's instructions. Extraction with TRIzol was performed twice, to decrease DNA contamination. After RNA was treated with 3 U of RQ1RNase-free DNase (Promega), another extraction with TRIzol was performed. To obtain only full-length *var* transcripts, RNA was dissolved in binding buffer (0.5 mol/L LiCl, 1 mmol/L EDTA, 10 mmol/L Tris, pH 7.5), and 1 pmol of biotinylated oligonucleotide complementary to the ATS domain (Biotin-5′-GGTTC(A/T)A(A/G)TAC(C/T)ACTTC(A/T)AT(C/T)CCTGGT(A/G)CATATATATCATTAATATCCAATT-CTTCATA(C/T)TCACTTTC(T/G)GA(A/T/G)GA-3′) was added and was incubated at a temperature gradient from 65°C to 4°C over 30 min. One hundred fifty micrograms of Dynabeads M-280 streptavidin, washed according to the manufacturer's protocol and dissolved in 5.5 mol/L LiCl, was added to the RNA. After undergoing rotation for 30 min at 37°C, the beads were washed 3 times with washing buffer (10 mmol/L Tris, 1 mmol/L EDTA, 0.15 mol/L NaCl, pH 7.5) and 1 time with 10 mmol/L Tris. RT was performed on the captured hybrids, primed by 400 ng oligo$(dT)_{12–18}$, and was done by use of Sensiscript (Qiagen) reverse transcriptase, according to the manufacturer's protocol, in a final volume of 20 μL. An aliquot without reverse transcriptase was used as a negative control. After RT, cDNA was treated with RNase A, and 1 μL was used for each of the various

PCR amplifications (table 1) with Advantage cDNA polymerase (Clontech), by use of the primers shown in table 1. The PCR conditions were 35 cycles for 30 s at 95°C, for 1 min at the annealing temperature (table 1), and for 70 s at 64°C. One microliter of negative control (without reverse transcriptase) was amplified in parallel. If this negative control yielded a product, then the positive sample was discarded and was excluded from the analysis.

To exclude cross-contamination of RT-PCR products, sequence-specific primers were designed for those sequences that occurred in >1 child. Sequence-specific PCR was performed on the genomic DNA of the parasites of those children, to confirm the presence of each particular sequence.

***Cloning and sequencing of PCR products.*** PCR products were cloned into pGEM-T vector (Promega) or pGEM-3Zf(+) vector (Promega), according to the manufacturer's instructions, and they were transfected into *Escherichia coli* SURE cells (Stratagene). An average of 20 positive clones was processed for sequencing (Montage Plasmid Miniprep$_{96}$ Kit [Millipore]; 96 capillary ABI Prism automated sequencing system [Applied Biosystems]). Multiple alignment of identical sequences derived from the same probe allowed the exclusion of PCR-derived mutations. Two sequences were considered to be identical when <3 single-nucleotide polymorphisms (SNPs) were detected.

***Sequence analysis.*** DNA sequence analysis was performed using DNASTAR (version 4; http://www.dnastar.com/), BLAST (from the National Center for Biotechnology Information Web page [http://www.ncbi.nlm.nih.gov/BLAST/]), plasmoDB, CLUSTALW (http://searchlauncher.bcm.tmc.edu/multi-align/ or http://www.ebi.ac.uk/clustalw/), and BioEdit (version 5; http://jwbrown.mbio.ncsu.edu/BioEdit/bioedit.html). Phylogenetic analyses were performed using PHYLIP (version 3.6; http://evolution.genetics.washington.edu/phylip.html) or Molecular Evolutionary Genetics Analysis (MEGA, version 1.02; The Pennsylvania State University [http://evolgen.biol.metro-u.ac.jp/MEGA]), with neighbor-joining or maximum parsimony methods and accompanied by bootstrap analysis with 1000 replicates. Predictions of secondary structure were performed using PredictProtein (http://cubic.bioc.columbia.edu/predictprotein/). Population comparison analysis was performed using ARLEQUIN (version 2; analysis of molecular variance; Genetics and Biometry Laboratory, University of Geneva [http://anthropologie.unige.ch/arlequin/]). Nucleotide sequence data are available in GenBank (accession nos. AY462581–AY462851).

## RESULTS

**var *Gene transcription in individual children.*** We studied *var* gene transcription longitudinally in asymptomatic children in Papua New Guinea for 4 months. RT-PCR was performed on full-length *var* transcripts over 3 adhesive domains (DBL1$\alpha$, CIDR1, and DBL$\beta$) and 2 upstream regions (subtelomeric upsB-type 5′ untranslated region [UTR] to DBL1$\alpha$ and central upsC-type 5′UTR to DBL1$\alpha$) (figure 1*A*). RT-PCR was followed by cloning and sequencing of the PCR products. The DBL1$\alpha$ domain yielded the most-informative PCR product, because it contains conserved blocks that enable the design of universal primers and thus ensures the amplification of the majority of *var* transcripts with minimal bias [22]. However, we are aware that, with this approach, potential bias in the identification of sequences might occur.

We observed a highly dynamic and variable picture of *var* transcription, which, in the next section, is described in detail for 4 representative children (figures 1–4). The longitudinal dis-

**Table 1. Oligonucleotide primers used for amplification of *var* gene regions.**

| *var* Gene region | Length of amplified product | $T_{anneal}$ | Name | Primer sequence | Reference |
|---|---|---|---|---|---|
| upsB-type 5′UTR-DBL1$\alpha$ | 1 kb | 53°C | *var* 4A3-5′ [a] | 5′-CTCAT(A/T)TATAATTTTACAAAATATATAAAAC-3′ | [16] |
| | | | *var* 4A3-3′ [b] | 5′-CC(A/T)AT(A/G)GC(A/G/T)GCAAAACT(G/C/T)CG(A/T)GC-3′ | [16] |
| upsC-type 5′UTR-DBL1$\alpha$ | 1 kb | 54°C | *var* 5B1-5′ | 5′-CACATATA(A/G)TACGACTAAGAAACA-3′ | [16] |
| | | | *var* 4A3-3′ | 5′-CC(A/T)AT(A/G)GC(A/G/T)GCAAAACT(G/C/T)CG(A/T)GC-3′ | [16] |
| DBL1$\alpha$ | 400 bp | 52°C | DBL$\alpha$-5′ | 5′-GCACGAAGTTTTGCAGATAT(A/T)GG-3′ | |
| | | | DBL$\alpha$-3′ | 5′-AA(A/G)TCTTC(T/G)GCCCATTCCTCGAACCA-3′ | |
| | 400 bp | 51°C | $\alpha$AF | 5′-GCACG(A/C)AGTTTTGC-3′ | [22] |
| | | | $\alpha$BR | 5′-GCCCATTC(G/C)TCGAACCA-3′ | [22] |
| CIDR1$\alpha$ | 400 bp | 45°C | CIDR1-5′ | 5′-GGT(A/T/G)(A/C/T/G)(A/C)TGATATGTTA(A/C)A(A/C)GATTC-3′ | |
| | | | CIDR1-3′a | 5′-T(C/T/G)TAGTAATTTATC(A/T/C)ATTGT-3′ | [14] |
| | | | CIDR1-3′b | 5′-T(C/T/G)TAATAAGAATTCGATTGC-3′ | |
| DBL$\beta$ | 500 bp | 50°C | DBL$\beta$-5′ | 5′-CGACGT(C/G)AACA(C/T)ATGTGTACATC-3′ | |
| | | | DBL$\beta$-3′ | 5′-CA(C/T)TC(T/G)GCCCA(C/T)TC(A/T)GTCATCC-3′ | |

**NOTE.** CIDR, cysteine-rich interdomain region; DBL, Duffy binding–like sequence; $T_{anneal}$, annealing temperature; UTR, untranslated region.

[a] Forward primer.

[b] Reverse primer.

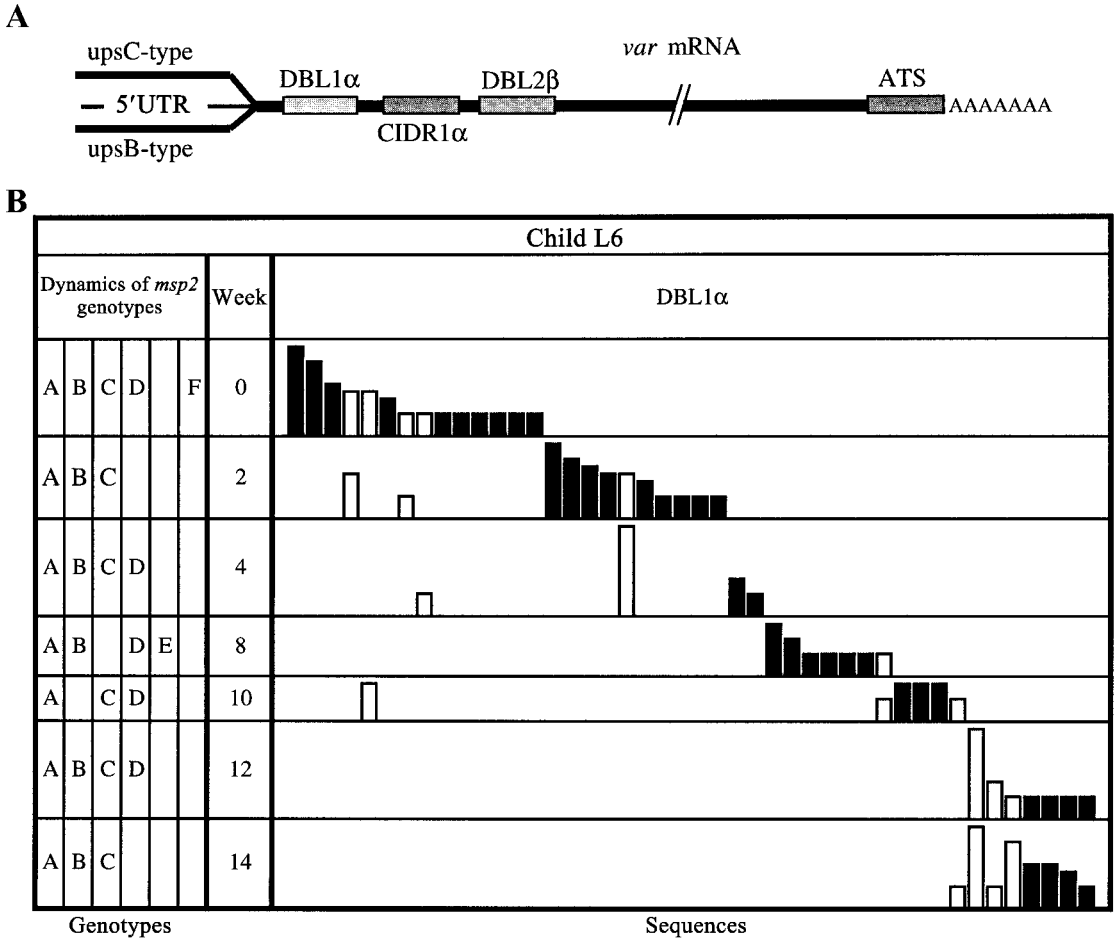


**Figure 1.** *A,* Schematic representation of sequenced *var* domains and upstream regions. *B,* Longitudinal distribution of Duffy binding–like 1α (DBL1α) sequences in child L6. The first major column shows the *msp2* genotyping data from the polymerase chain reaction–restriction fragment–length polymorphism analysis. The second major column indicates time points (in weeks) when samples were found to be positive for *Plasmodium falciparum* by microscopy. Different strains *(A–F)* are indicated in the minor columns. The third major column shows the longitudinal distribution of *var* gene sequences. *Bars,* individual sequences of DBL1α. Recurring sequences are indicated as white bars. The size of the bar reflects the relative frequency of a particular sequence ($\log_{10}$).

tribution of *var* sequences and *msp2* genotypes, for all time points and among all children, is shown in figure 5, which is available in the online version of this article, at the *Journal*'s Web site (http://www.journals.uchicago.edu/JID/journal/home.html).

In most children, a large number of different sequences were identified from single blood samples, and, at subsequent time points, mostly new sequences that had not been previously detected were identified. For example, for child L6, the majority of DBL1α sequences identified, which were isolated in 2-week intervals, were different at every time point (figure 1*B*). During 14 weeks, 48 different DBL1α sequences were identified. Some sequences were still present after 2 weeks or even recurred at later time points. It is important to note that child L6 was constantly infected with a large number of different *P. falciparum* strains, having 2–5 infections.

In contrast, some individuals, such as child L11, showed a less diverse *var* gene transcription, with only 1–3 different *var* transcripts identified per time point (figure 2). In child L11, all 5 domains were amplified from samples that were found to be positive for *P. falciparum,* according to microscopy, at weeks 0, 10, and 12, when the child was infected with only 1 or 2 strains. At week 10, 1 DBL1α, 1 CIDR1α, and 1 DBLβ sequence were found, suggesting that the origin of these sequences was in a single expressed *var* gene. The observed DBLβ sequence had already been observed 10 weeks earlier (figure 2). To test whether this reemerging sequence derived from 1 *var* gene or whether it was an identical domain of an otherwise different *var* gene, we amplified and sequenced genomic DNA with sequence-specific primers that targeted the domains 5′UTR-DBL1α and DBL1α-DBL2β. Identical sequences over this stretch were obtained, suggesting the presence of the identical *var* gene (data not shown).

In child L8, both upstream types (upsB-type and upsC-type

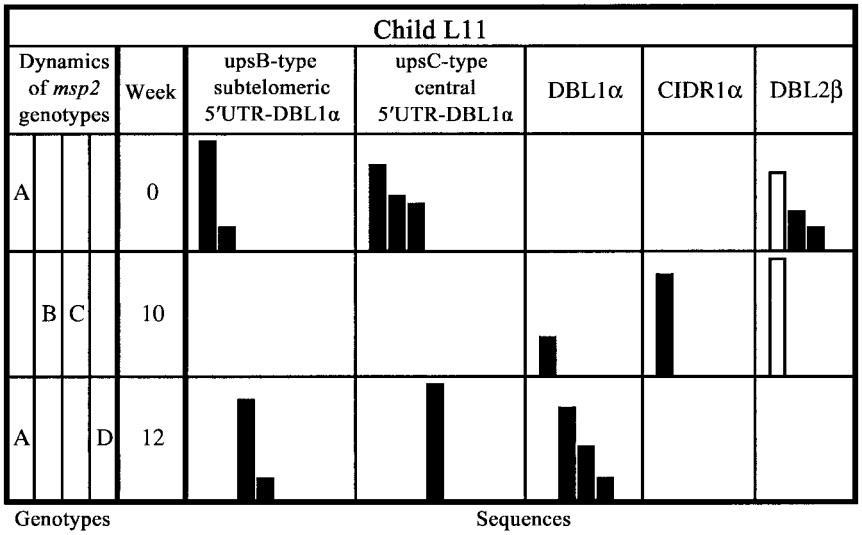


**Figure 2.** Longitudinal distribution of 5 different *var* gene regions in samples obtained from child L11. For details, see figure 1.

regions) were found at time points when samples were positive for *P. falciparum* by microscopy (figure 3). Between 1 and 11 different sequences of the upsB-type 5′UTR-DBL1$\alpha$ region were detected, whereas >1 upsC-type 5′UTR-DBL1$\alpha$ sequence was never found.

In child L12, identical sequences reemerged several times (figure 4). A upsC-type 5′UTR-DBL1$\alpha$ sequence recurred 3 times, at weeks 2, 4, and 12, and a CIDR1$\alpha$ sequence was found at weeks 0 and 4. Amplification of cDNA from week 2, by use of sequence-specific primers, also revealed the presence of this CIDR domain, and this indicates continuous transcription of this *var* gene for 4 weeks.

To test whether sequences from different domains of 1 sample were derived from the same *var* gene and to test the quantitative distribution of these sequences, the most abundant upsB-type 5′UTR and DBL1$\alpha$ sequences of child L6 (from the sample obtained at week 4) and child S12 (from the sample obtained at day 10) were linked with genomic DNA by PCR. Specific forward primers were designed in the upsB-type 5′UTR domain, and degenerated reverse primers were used for DBL1$\alpha$. Subsequent sequencing showed that, in both children, the previously identified most abundant DBL1$\alpha$ sequence was connected with the most abundant upsB-type 5′UTR-DBL1$\alpha$ sequence, indicating their origin in the same *var* gene.

***Distribution of* var *sequences in different children.*** Between 1 and 15 different sequences per domain were found in a single child. The average number of DBL1$\alpha$ sequences found per child was 5.2, and between 1.5 and 3.7 sequences were found for the other amplified domains (table 2).

Thirty-two identical sequences, which differed by no more than 1–3 SNPs, were found in >1 child (table 2). Twenty-three of those sequences were DBL1$\alpha$ domains. An additional 2 DBL1$\alpha$ sequences from child L3 and L8 differed by only 10 SNPs. DBL$\beta$ transcripts could be amplified from only 33% of samples in which we identified DBL1$\alpha$ transcripts.

Using the upstream sequence of transcribed *var* genes to determine their chromosomal location, we found that more *var* transcripts derived from subtelomerically located genes (upsB-type *var* transcript) than from centrally located genes (upsC-type *var* transcript). A mean of 1.3 upsB-type *var* gene transcripts was detected per infecting strain, compared with a mean of 0.7 upsC-type *var* gene transcripts (table 2).

***Multiple infections and* var *gene transcription.*** Using *msp2* genotyping, we found, in 86% of all cases, multiple infections (2–5 infections), with an average of 2.8 infecting strains. When the average of 5.2 DBL1$\alpha$ sequences per sample was adjusted to the number of infections present, we found an average of 1.7 DBL1$\alpha$ sequences per infecting strain (table 2). Linear regression analysis revealed a significant correlation between the number of infecting strains and the number of DBL1$\alpha$ sequences identified ($P = .0002$; $r^2 = 0.31$; confidence limits, $\pm 0.45$). In some instances, the number of *var* gene sequences observed was less than the number of infecting strains present. This finding might be ascribed either to technical limits, such as sensitivity or primer bias, which we cannot exclude while amplifying such a diverse gene family, or to the lack of the respective domain in a transcribed *var* gene (e.g., no DBL$\beta$ domain). At all time points when *var* sequences reemerged, identical strains were detected, indicating the presence of the same parasite. These identical strains were also detected in samples obtained between both time points (figures 1–4).

***Analysis of* var *domain sequences.*** All multiple alignments

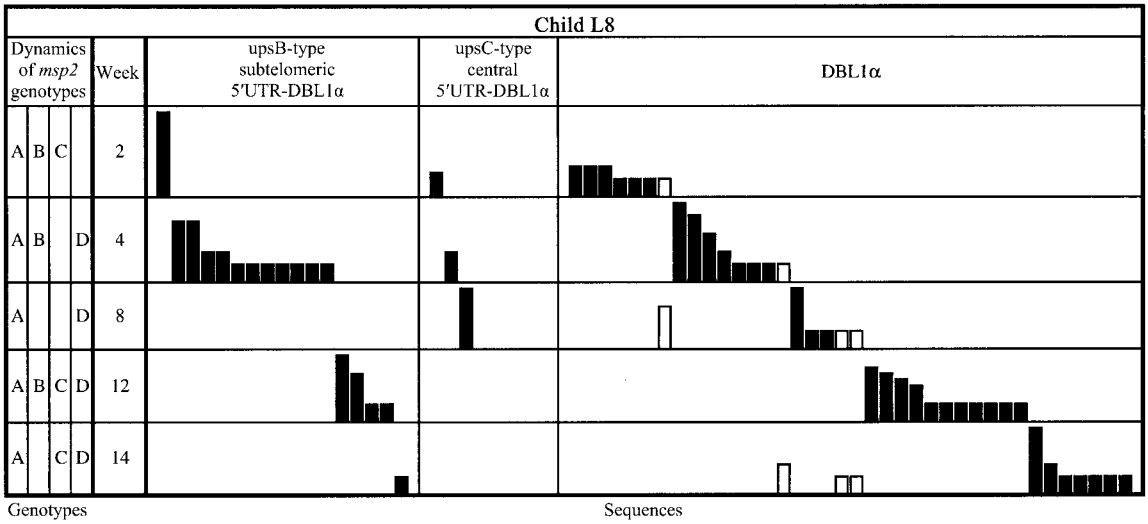


**Figure 3.** Longitudinal distribution of *var* gene sequences of both upstream regions and Duffy binding–like 1α (DBL1α) in samples obtained from child L8. For details, see figure 1.

of the different domains can be found in figure 6, which is available in the online version of this article, at the *Journal*'s Web site (http://www.journals.uchicago.edu/JID/journal/home.html). Multiple alignment of 150 DBL1α sequences confirmed the existence of "universally" conserved blocks (reviewed in [9]). We found clear evidence for recombination—for example, 2 sequences in samples from child S12 and child L3, which differed in the first 180 bp and which were identical after the universally well-conserved homology block F. Furthermore, for child L6, over the first 255 bp to homology block F, a DBL1α sequence from a sample obtained at week 12 was identical to another DBL1α sequence obtained at the same time point, whereas, after this 255-bp stretch, the sequence was identical to a DBL1α sequence isolated from a sample obtained from the same child 2 weeks later.

Sequences similar to the previously described DBL-$\alpha_1$ subtype [26, 27] were found 40 times (25%). Also, the $var_{COMMON}$ type [28], which has been shown to be constitutively transcribed in 60% of malaria-infected Gabonese children, was found 3 times (twice in samples from child L6 that were obtained 2 weeks apart and once in a sample obtained from child L8).

Alignment and phylogenetic analysis of the Papua New Guinea–derived DBL1α sequences with 50 previously sequenced DBL1α sequences from African *P. falciparum* strains did not show any separate clustering of PNG or African samples (data not shown). In contrast, stretches of 10–20 amino acids in the polymorphic region of DBL1α occurring in only 1 PNG sample recurred in 1 of the African samples. No geographic patterns could be detected, which is consistent with the results of other studies [29, 30].

Multiple alignments with 40 upsB-type 5′UTR-ATG sequences and 9 upsC-type 5′UTR-ATG sequences confirmed the conserved character of upsB- and upsC-type upstream sequences [16]. However, 150 bp upstream of ATG, some sequences showed deletions of up to 100 bp. One upsB-type 5′UTR-DBL1α sequence occurred in 1 child and in 3D7 parasites as *var* group B/C (PF08′0103), and it showed small differences from the characteristic upsB-type upstream sequence. It had a different length of poly(dA-dT) and homopolymeric (dA:dT) tracts and a deletion of 117 bp in an otherwise well-conserved stretch, located 250 bp upstream of the ATG.

In multiple alignments and in phylogenetic and protein-structure analysis of 38 upsB-type and 15 upsC-type *var* genes over the NTS region and the first 100 amino acids of DBL1α, no difference was observed between subtelomeric and central *var* genes. NTS and DBL1α sequences derived from either subtelomeric or central locations did not cluster separately in phylogenetic analyses or by computing population comparison tests, and the difference in this stretch of sequence between subtelomeric and central *var* genes was not significant (fixation index, 0.015; $P = .088$).

***Sequences identified in field isolates and from 3D7 in the genome project.*** We identified 1 upsB-type 5′UTR-DBL1α sequence and 3 DBL1α sequences that were identical to *var* gene domains in 3D7 parasites. One particular 3D7 *var* DBL1α sequence was detected 3 times (once in child L13 and twice in child L6). According to PlasmoDB, this DBL1α sequence represents a *var* pseudogene (PFL1970w) with a premature stop codon 4275 bp downstream of the ATG. However, when we designed PFL1970w-specific primers to sequence over this stop codon, PCR on genomic DNA of 3D7, followed by cloning and sequencing of 10 clones, revealed an insertion relative to the 3D7

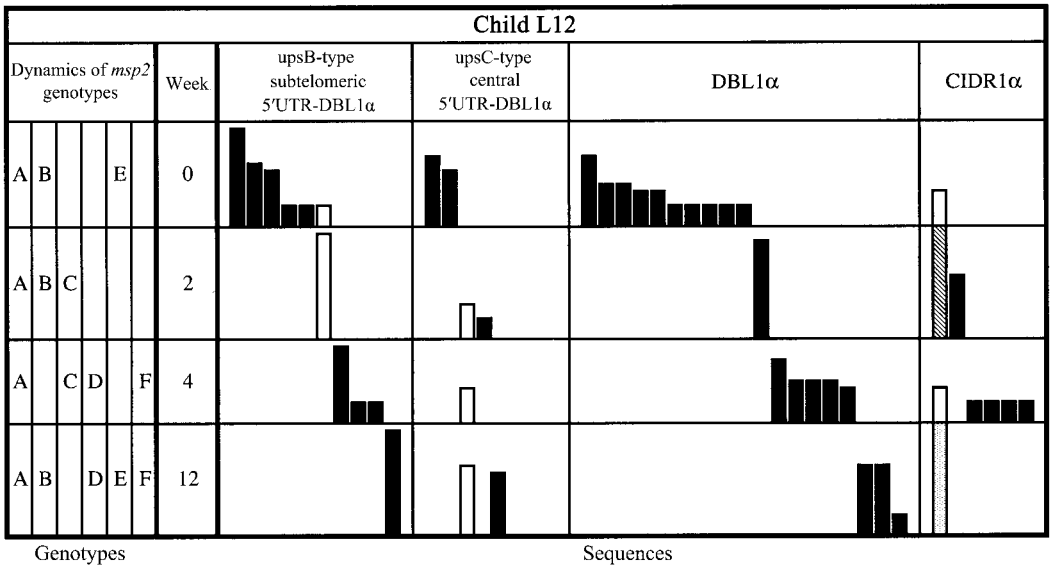


**Figure 4.** Longitudinal distribution of both upstream regions, Duffy binding–like sequence 1$\alpha$ (DBL1$\alpha$), and cysteine-rich interdomain region–$\alpha$ (CIDR$\alpha$) in samples obtained from child L12. The hatched and dotted bars in the "CIDR1$\alpha$" panel denote 1 CIDR1$\alpha$ clone that was detected at weeks 0 and 4 *(white bars)* and that was also detected by sequence-specific primers on cDNA at week 2 *(hatched bar)* and on genomic DNA at week 12 *(dotted bar).* For details, see figure 1.

sequence producing a frameshift mutation (4233 bp downstream of ATG), resulting in a continuous open-reading frame.

## DISCUSSION

During the past years, much information on *var* gene transcription has been gained; however, most of this information has been based on the findings of in vitro studies, and few studies have looked at *var* gene transcription in vivo [24, 26, 31]. To our knowledge, this is the first longitudinal study of *var* gene transcription in naturally infected children. We generated cDNA and cloned and sequenced 3 adhesive *var* gene domains (DBL1$\alpha$, CIDR1, and DBL$\beta$) and 2 *var* 5′UTR stretches (upsB-type and upsC-type regions) from blood samples obtained, over 4 months, from asymptomatic children living in an area where malaria is endemic. Using this approach, we identified a large number of different sequences, and we observed a dynamic picture of *var* gene transcription. However, despite such a dynamic transcription pattern, some sequences persisted or recurred for up to 10 weeks.

It has often been argued that RT-PCR could identify the smallest amounts of RNA, and, in the case of *var* gene transcription, it is indeed unclear whether cDNA represents only functional full-length mRNAs, because incomplete and 3′-truncated *var* transcripts have been observed [22]. By selecting *var* transcripts that contained the 3′ATS domain before RT-PCR was performed, we are confident that the number of incomplete mRNAs was reduced to insignificant levels. This was confirmed by the use of tags, other than the anti-ATS, that resulted in no product at all (data not shown). Furthermore, when Peters et al. [24] compared the number of *var* transcripts from 3D7 between the ring stage and the trophozoite stage, they observed no difference in the number and proportion of transcripts, and they even questioned the relaxed transcription in vivo. Moreover, by use of single-cell RT-PCR for trophozoite-stage cells, up to 5 different transcripts were observed in 3D7 [32], which questions the hypothesis of mutually exclusive *var* gene transcription that was previously suggested by results of cultures selected for receptor binding [20, 33].

We observed the largest diversity within the DBL1$\alpha$ sequences, with an average of 5.2 different sequences per blood sample. This finding is similar to previous reports of 3–15 different *var* gene transcripts in natural infections or from laboratory-adapted strains [22, 24, 31, 32, 34]. One explanation for the large number of observed *var* transcripts at 1 time point might be the presence of multiple concurrent *P. falciparum* infections. It is noteworthy that the mean number of *var* gene transcripts per sample very much resembles the mean number of concurrent infecting strains in this age group [35]. It has been speculated that multiple infections provide protection against hyperinfection by stimulating the immune system with a broad range of diverse antigens, such as PfEMP1 [35, 36]. Concomitantly expressed PfEMP1 variants could also explain the findings of Bull et al. [37], who showed that children with asymptomatic infections had a greater repertoire of variant-specific antibodies.

Most children had multiple infections, and multiple *var* transcripts were detected in these children. This clearly added to the complexity of the observed *var* gene expression dynamic.

**Table 2. Overview of analyzed sequences of different transcribed *var* gene regions.**

| General overview of the no. of sequences | *var* Gene regions | | | | |
|---|---|---|---|---|---|
| | Subtelomeric upsB-type 5′UTR-DBL1$\alpha$ | Central upsC-type 5′UTR-DBL1$\alpha$ | DBL1$\alpha$ | DBL$\beta$ | CIDR1$\alpha$ |
| No. of sequences (no. of positive blood samples) | 417 (20) | 133 (17) | 789 (39) | 97 (13) | 118 (8) |
| No. of different sequences | 66 | 23 | 158 | 22 | 17 |
| Average no. of different sequences per child and time point | 3.7 | 1.5 | 5.2 | 2.0 | 3.0 |
| Average no. of different sequences per infecting strain, child, and time point | 1.3 | 0.7 | 1.7 | 0.8 | 1.7 |
| No. of transcribed different sequences in 3D7 culture strain | 7 | 1 | 8 | 5 | ND |
| No. of recurring sequences in the same child | | | | | |
| After 10 days | 2 | … | 2 | … | … |
| After 2 weeks | 2 | 2 | 7 | … | … |
| After 4 weeks | … | … | 2 | … | 1 |
| After 5 weeks | 1 | … | … | … | … |
| After 6 weeks | … | … | 4 | … | … |
| After 8 weeks | … | 1 | … | … | … |
| After 10 weeks | … | 1 | 2 | 1 | … |
| No. of identical sequences in different children | | | | | |
| In 2 children | 3 | … | 16 | 4 | … |
| In 3 children | … | 1 | 4 | … | … |
| In 3D7 culture strain | 1 | … | 3[a] | … | … |

**NOTE.** CIDR, cysteine-rich interdomain region; DBL, Duffy binding–like sequence; ND, not done; UTR, untranslated region.

[a] One of these 3 sequences was detected in 2 children.

However, in samples obtained from children who had single infections, only 1 DBL1$\alpha$ sequence was found to be abundant or present. There was a linear correlation between the number of detected transcripts and the number of infecting strains present. The relatively low number of *var* transcripts in children who had a single infection suggests a tight regulation that allowed the transcription of only 1 or a few *var* genes at 1 time in the clonal parasite population. However, the observation of rapidly changing *var* transcripts suggests high switching rates. Peters et al. [24] reported different switching rates between initial and subsequent switching events in laboratory-induced infections. Our data suggest different switching rates for some *var* genes that were transcribed for only a short period, whereas others persist for weeks or recur. In the present study, we cannot distinguish between different switching rates or selection forces acting on previously expressed PfEMP1 molecules. We also cannot completely rule out that recurring *var* gene sequences are derived from a closely related parasite strain that expresses a highly similar *var* gene. However, the detection of these recurring *var* sequences at intermediate time points (figure 4) and the presence of the same *P. falciparum* strain suggest that these *var* sequences are derived from a constantly expressed *var* gene.

Although we have no quantitative data for the detected *var* transcripts, we were able to semiquantify the presence of *var* transcripts, to a certain degree, and to show that the most abundant sequences of 2 *var* regions in the same probe originated in 1 *var* gene (see the "*var* Gene Transcription in Individual Children" subsection in Results). We are currently addressing this question in ongoing case-control studies that use real-time PCR to quantify *var* transcripts, with respect to the various groups of upstream regions.

The previously described conserved character of dimorphic *var* gene upstream regions [16] and the different regulation mechanisms of these *var* gene groups [19] led us to speculate that they might be functionally different. Therefore, we amplified both 5′UTR stretches and the adjacent NTS-DBL1$\alpha$ sequences, and we then compared expression dynamics and coding sequence. Of all analyzed 5′UTR sequences, 26% were of the upsC type, which reflects the proportion of upsC-type *var* genes in the 3D7 genome, which is 22%. We never detected expression of >3 different centrally located *var* genes in a sample. We had speculated that centrally located *var* genes would be more conserved because of a location that was less prone to recombination. This is supported by the findings of Rubio et al. [38], who showed that centrally located *var* genes were more related to each other than to subtelomerically located *var* genes. Consequently, we hypothesized that upsC-type *var* genes would be recognized faster by the immune system. However, in the present study, upsC-type sequences were found to persist for 8–10 weeks, suggesting that these *var* genes can be also

expressed over a long time. Furthermore, we were unable to show any structural or phylogenetic differences within the short sequence of the NTS-DBL1$\alpha$ domain of both *var* gene groups. However, this stretch of sequence is known to be conserved, represents only a limited stretch of the whole *var* gene, and does not allow for further conclusions to be made.

In summary, we have shown that chronically infected children express several *var* genes simultaneously, with a mean of 1.7 different *var* genes per infecting strain. The repertoire of *var* genes circulating in a parasite population seems to be large and recombinogenic. We have shown a rapid change in *var* gene expression, but we also detected, for the first time, recurring *var* genes after 10 weeks.

Expression of many and rapidly changing *var* genes is expected in asymptomatic children, in whom parasite survival is a balance between antigenic escape and binding capacity to avoid splenic clearance. These children must be considered to be semi-immune, and the *var* gene repertoire of the infecting parasites might already be exhausted. Therefore, the parasite might be forced to rapidly switch to new *var* variants. To unequivocally understand *var* gene expression in vivo, studies involving naive individuals must be conducted. Furthermore, comparing *var* gene expression in subjects who have severe malaria with that in subjects who have mild malaria should shed more light on the complex cascade of *var* gene expression and an immunological escape mechanism.

## Acknowledgments

We thank the staff of the Papua New Guinea Institute of Medical Research, Madang, for their field assistance. We appreciate the cooperation of the parents and children of the Maiwara Primary School for participating in the study. We are very grateful to Dorothee Foernzler and Patricia Riegert for sequencing, Sophia Kossida for assistance in phylogenetic analysis, and Matthias Rottmann and Christian Flueck for critical reading of the manuscript.

## References

1. Su XZ, Heatwole VM, Wertheimer SP, et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*–infected erythrocytes. Cell **1995**; 82:89–100.
2. Smith JD, Chitnis CE, Craig AG, et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell **1995**; 82:101–10.
3. Baruch DI, Pasloske BL, Singh HB, et al. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell **1995**; 82:77–87.
4. Kyes S, Horrocks P, Newbold C. Antigenic variation at the infected red cell surface in malaria. Annu Rev Microbiol **2001**; 55:673–707.
5. Urban BC, Ferguson DJ, Pain A, et al. *Plasmodium falciparum*–infected erythrocytes modulate the maturation of dendritic cells. Nature **1999**; 400:73–7.
6. Rowe JA, Moulds JM, Newbold CI, Miller LH. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. Nature **1997**; 388:292–5.
7. Biggs BA, Anders RF, Dillon HE, et al. Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum.* J Immunol **1992**; 149:2047–54.
8. Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med **1998**; 4:358–60.
9. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. Mol Biochem Parasitol **2000**; 110:293–310.
10. Krych-Goldberg M, Moulds JM, Atkinson JP. Human complement receptor type 1 (CR1) binds to a major malarial adhesin. Trends Mol Med **2002**; 8:531–7.
11. Smith JD, Craig AG, Kriek N, et al. Identification of a *Plasmodium falciparum* intercellular adhesion molecule–1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci USA **2000**; 97:1766–71.
12. Buffet PA, Gamain B, Scheidig C, et al. *Plasmodium falciparum* domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. Proc Natl Acad Sci USA **1999**; 96:12743–8.
13. Reeder JC, Cowman AF, Davern KM, et al. The adhesion of *Plasmodium falciparum*–infected erythrocytes to chondroitin sulfate A is mediated by *P. falciparum* erythrocyte membrane protein 1. Proc Natl Acad Sci USA **1999**; 96:5198–202.
14. Degen R, Weiss N, Beck HP. *Plasmodium falciparum:* cloned and expressed CIDR domains of PfEMP1 bind to chondroitin sulfate A. Exp Parasitol **2000**; 95:113–21.
15. Baruch DI, Ma XC, Singh HB, Bi X, Pasloske BL, Howard RJ. Identification of a region of PfEMP1 that mediates adherence of *Plasmodium falciparum* infected erythrocytes to CD36: conserved function with variant sequence. Blood **1997**; 90:3766–75.
16. Voss TS, Thompson JK, Waterkeyn J, et al. Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5′ flanking sequences. Mol Biochem Parasitol **2000**; 107:103–15.
17. Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum.* Nature **2002**; 419:498–511.
18. Scherf A, Figueiredo LM, Freitas-Junior LH. *Plasmodium* telomeres: a pathogen's perspective. Curr Opin Microbiol **2001**; 4:409–14.
19. Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP. Identification of nuclear proteins that interact differentially with *Plasmodium falciparum var* gene promoters. Mol Microbiol **2003**; 48:1593–607.
20. Scherf A, Hernandez-Rivas R, Buffet P, et al. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum.* EMBO J **1998**; 17:5418–26.
21. Deitsch KW, Calderwood MS, Wellems TE. Malaria: cooperative silencing elements in *var* genes. Nature **2001**; 412:875–6.
22. Taylor HM, Kyes SA, Harris D, Kriek N, Newbold CI. A study of *var* gene transcription in vitro using universal *var* gene primers. Mol Biochem Parasitol **2000**; 105:13–23.
23. Roberts DJ, Craig AG, Berendt AR, et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature **1992**; 357:689–92.
24. Peters J, Fowler E, Gatton M, Chen N, Saul A, Cheng Q. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci USA **2002**; 99:10689–94.
25. Felger I, Tavul L, Beck HP. *Plasmodium falciparum:* a rapid technique for genotyping the merozoite surface protein 2. Exp Parasitol **1993**; 77:372–5.
26. Kirchgatter K, Portillo HA. Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol Med **2002**; 8:16–23.

27. Robinson BA, Welch TL, Smith JD. Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol **2003**; 47:1265–78.
28. Winter G, Chen Q, Flick K, Kremsner P, Fernandez V, Wahlgren M. The 3D7var5.2 (*var*COMMON) type *var* gene family is commonly expressed in non-placental *Plasmodium falciparum* malaria. Mol Biochem Parasitol **2003**; 127:179–91.
29. Kyes S, Taylor H, Craig A, Marsh K, Newbold C. Genomic representation of *var* gene sequences in *Plasmodium falciparum* field isolates from different geographic regions. Mol Biochem Parasitol **1997**; 87:235–8.
30. Kirchgatter K, Mosbach R, del Portillo HA. *Plasmodium falciparum:* DBL-1 *var* sequence analysis in field isolates from central Brazil. Exp Parasitol **2000**; 95:154–7.
31. Afonso NP, Wunderlich G, Shugiro TM, et al. *Plasmodium falciparum:* analysis of transcribed *var* gene sequences in natural isolates from the Brazilian Amazon region. Exp Parasitol **2002**; 101:111–20.
32. Duffy MF, Brown GV, Basuki W, et al. Transcription of multiple *var* genes by individual, trophozoite-stage *Plasmodium falciparum* cells expressing a chondroitin sulphate A binding phenotype. Mol Microbiol **2002**; 43:1285–93.
33. Chen Q, Fernandez V, Sundstrom A, et al. Developmental selection of *var* gene expression in *Plasmodium falciparum*. Nature **1998**; 394:392–5.
34. Noviyanti R, Brown GV, Wickham ME, Duffy MF, Cowman AF, Reeder JC. Multiple *var* gene transcripts are expressed in *Plasmodium falciparum* infected erythrocytes selected for adhesion. Mol Biochem Parasitol **2001**; 114:227–37.
35. Smith T, Felger I, Tanner M, Beck HP. Premunition in *Plasmodium falciparum* infection: insights from the epidemiology of multiple infections. Trans R Soc Trop Med Hyg **1999**; 93(Suppl 1):59–64.
36. al Yaman F, Genton B, Reeder JC, Anders RF, Smith T, Alpers MP. Reduced risk of clinical malaria in children infected with multiple clones of *Plasmodium falciparum* in a highly endemic area: a prospective community study. Trans R Soc Trop Med Hyg **1997**; 91:602–5.
37. Bull PC, Lowe BS, Kaleli N, et al. *Plasmodium falciparum* infections are associated with agglutinating antibodies to parasite-infected erythrocyte surface antigens among healthy Kenyan children. J Infect Dis **2002**; 185:1688–91.
38. Rubio JP, Thompson JK, Cowman AF. The *var* genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. EMBO J **1996**; 15:4069–77.