

Plasmodium falciparum var gene expression is modified by host immunity

George M. Warimwe^a, Thomas M. Keane^b, Gregory Fegan^{a,c}, Jennifer N. Musyoki^a, Charles R. J. C. Newton^{a,d}, Arnab Pain^b, Matthew Berriman^b, Kevin Marsh^{a,e}, and Peter C. Bull^{a,e,1}

^aKenya Medical Research Institute-Wellcome Trust Research Programme, 80108 Kilifi, Kenya; ^bWellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; ^cInfectious Disease Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; ^dInstitute of Child Health, University College of London, London WC1N 1EH, United Kingdom; and ^eNuffield Department of Clinical Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, United Kingdom

Edited by Louis H. Miller, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, Rockville, MD, and approved October 21, 2009 (received for review July 9, 2009)

Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) is a potentially important family of immune targets, which play a central role in the host-parasite interaction by binding to various host molecules. They are encoded by a diverse family of genes called *var*, of which there are ≈ 60 copies in each parasite genome. In sub-Saharan Africa, although *P. falciparum* infection occurs throughout life, severe malarial disease tends to occur only in childhood. This could potentially be explained if (i) PfEMP1 variants differ in their capacity to support pathogenesis of severe malaria and (ii) this capacity is linked to the likelihood of each molecule being recognized and cleared by naturally acquired antibodies. Here, in a study of 217 Kenyan children with malaria, we show that expression of a group of *var* genes “cys2,” containing a distinct pattern of cysteine residues, is associated with low host immunity. Expression of *cys2* genes was associated with parasites from young children, those with severe malaria, and those with a poorly developed antibody response to parasite-infected erythrocyte surface antigens. *Cys-2 var* genes form a minor component of all genomic *var* repertoires analyzed to date. Therefore, the results are compatible with the hypothesis that the genomic *var* gene repertoire is organized such that PfEMP1 molecules that confer the most virulence to the parasite tend also to be those that are most susceptible to the development of host immunity. This may help the parasite to adapt effectively to the development of host antibodies through modification of the host-parasite relationship.

antigenic variation | escape | malaria | PfEMP1 | virulence

Children living in malaria endemic areas develop significant naturally acquired immunity to severe malaria during the first 5 years of life (1). The clonally variant surface antigens called *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) are strong candidate targets for this immunity. These multidomain variant antigens are encoded in a mutually exclusive fashion by about 60 *var* genes per parasite genome and exported to the infected erythrocyte surface where they are exposed to host antibodies (2).

PfEMP1 are also implicated as virulence factors. Through interactions with host molecules such as ICAM1, CD36, CR1, and CD31, PfEMP1 plays a central role in mediating cytoadherence of infected erythrocytes to host cells. This is believed to be responsible for the severe pathology associated with *P. falciparum* malaria (3). PfEMP1 molecules undergo clonal antigenic variation meaning that a single genotype can evade host antibodies by switching between *var* genes (4, 5). After repeated exposure to infection, a repertoire of variant-specific antibodies that can recognize the variant surface antigens expressed by most parasite isolates builds up. Piecemeal acquisition of such antibodies could help explain the development of naturally acquired immunity to malaria (6, 7).

The relatively rapid rate of acquisition of immunity to severe malaria compared to mild malaria (8) may suggest a limitation in the diversity of important immune targets in *P. falciparum*. This is supported by the fact that immunity to malaria is accompanied by changes in the serological properties of the variant surface antigens.

Parasite-infected erythrocytes from young children and those with severe malaria tend to be better recognized by antibodies from semi-immune children than those from older children or those with nonsevere malaria (9, 10). This may suggest the existence of a group of relatively conserved PfEMP1 variants associated with high pathogenicity and low host immunity. Such molecules could serve as potentially important targets of intervention.

Analysis of full-length sequences of the repertoires of *var* genes from several lab-adapted parasite lines supports genetic structuring of the variant antigen repertoire (2, 11, 12). For example, recombinant domains from PfEMP1 molecules carrying an UpsA promoter have been shown to have low affinity for CD36 binding relative to equivalent domains from *var* genes with UpsB or UpsC promoters (13). This structuring of the genomic *var* gene repertoire has been linked to the serological properties of the expressed variant surface antigens. Parasites selected in vitro for binding to IgG from semi-immune children have increased overall frequency of recognition by heterologous antibodies, reduced affinity for CD36 binding, and a bias toward expression of UpsA-associated *var* genes (hereafter called group A *var* genes) (14). Because of the association between commonly recognized variant surface antigens and severe malaria, group A *var* genes have been proposed to represent a pathologically significant group (14). However, direct evidence for a link between *var* expression, pathology, and naturally acquired immunity requires analysis of parasites from clinical malaria infections.

Such studies are problematic. The immense architectural diversity of *var* genes, together with their capacity to undergo recombination (15), yields limited positions for PCR amplification and sequence sampling. Therefore we (16) and others (17–21) have relied on analysis of short, ≈ 350 nucleotide, expressed sequence tags amplified from a region corresponding to a domain that is present in most PfEMP1 variants, DBL α . To estimate PfEMP1 expression levels, reverse transcriptase PCR products are subcloned into *Escherichia coli*, bacterial colonies picked and sequenced, and the sequence tags classified and counted. Previous studies have varied in their depth of sequencing from each parasite isolate, the number of patients in different clinical categories, time of RNA sampling within the parasite asexual cycle, method of sequence analysis, and whether the emphasis is placed more on the dominantly expressed sequence(s) in each patient (19, 20) or on the

Author contributions: K.M. and P.C.B. designed research; G.M.W., J.N.M., A.P., M.B., and P.C.B. performed research; T.M.K., A.P., M.B., and P.C.B. contributed new reagents/analytic tools; G.M.W., G.F., C.R.J.C.N., and P.C.B. analyzed data; and G.M.W. and P.C.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: pbull@kilifi.kemri-wellcome.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907590106/DCSupplemental.

overall percentage representation within each patient of different groups of sequence [supporting information (SI) Table S1] (16, 22).

Despite varying approaches the results have been encouraging. First, although DBL α sequence tags contain only a small proportion of a *var* gene's sequence, specific sequence features present in DBL α tags isolated worldwide can be used to classify them (16, 22). The vast majority of DBL α tags carry either two or four cysteine residues. Although they are not exclusive to group A *var* genes, DBL α tags of all group A *var* genes contain two cysteine residues. A large proportion of DBL α tags with 2 cysteine residues (henceforth called *cys2 var* genes) also carry one of two motifs, MFK and REY, located at two different positions within the sequences but never found together within the same sequence (16).

Second, these broad classes of *var* genes appear to be differentially associated with host immunity. In a small pilot study of 12 isolates, children with poorly developed immune responses tended to express *cys2 var* genes with MFK motifs (16). More recently, Kyriacou et al. showed that *cys2 var* genes were the dominant sequence type expressed by parasites isolated from children with cerebral malaria in Mali (19). These studies support the usefulness of a sequencing approach to comparing PfEMP1 expression between clinical parasite isolates.

However, these studies were not designed to fully explore the relationship between expression of groups of PfEMP1, malaria pathology, and naturally acquired immunity (Table S1). To do this requires large numbers of patients and relatively deep sequencing of the expressed sequence tags. Moreover, data analysis needs to account for the fact that severe malaria in African children presents in three major overlapping clinical syndromes (i.e., impaired consciousness, severe malarial anemia, and severe respiratory distress) that tend to occur in children of different age groups (23, 24), meaning that the underlying mechanisms for the pathogenesis of these syndromes may be different.

To examine the relationship between *cys2 var* gene expression and host immunity we have extended our previous sampling, sequencing, and classification approach (16, 22, 25) to 217 parasites obtained from Kenyan children with clinical malaria. We relate these expression patterns to each child's clinical manifestation of disease, age, and level of antibodies to the surface of infected erythrocytes at the time of disease.

Results

We classified and counted 14,516 DBL α sequence tags amplified from parasite cDNA sampled from each of 112 children with severe malaria and 105 with nonsevere malaria (Table S2). The tags were assembled into 4,225 individual sequence "contigs" (EMBL accession nos FN588437–FN592661, see *Materials and Methods* and Dataset S1). As found previously (22), the majority of the DBL α sequence tags contained either two (42.8%) or four (53%) cysteine residues.

cys2 var Expression Is Associated with Severe Malaria and Young Host Age.

To provide a semiquantitative estimate of the proportion of parasites in each isolate expressing *cys2 var* genes we calculated, from each parasite isolate, the proportion of expressed sequence tags that contained two cysteine residues. If a group of *var* genes is associated with low host immunity the proportion of parasites expressing it within a clinical isolate should decrease with host age and increase with disease severity. In support of this, linear regression modeling showed that the level of expression of *cys2 var* genes was predicted by severity of malaria ($B = 0.17$, 95% CI 0.1, 0.3, $p = 0.0004$; adjusted for host age) and decreased with host age ($B = -0.03$, 95% CI -0.05 , -0.02 , $p = 0.0002$; adjusted for disease severity) (Fig. 1A and B and Figs. S1 and S2). For 53 of the parasite isolates we generated sequence tags from both cDNA and genomic DNA (EMBL accession numbers for gDNA contigs: FN592662–FN594512). Within this sample, comparable results were obtained for *cys2 var* gene sequences from cDNA but not genomic DNA,

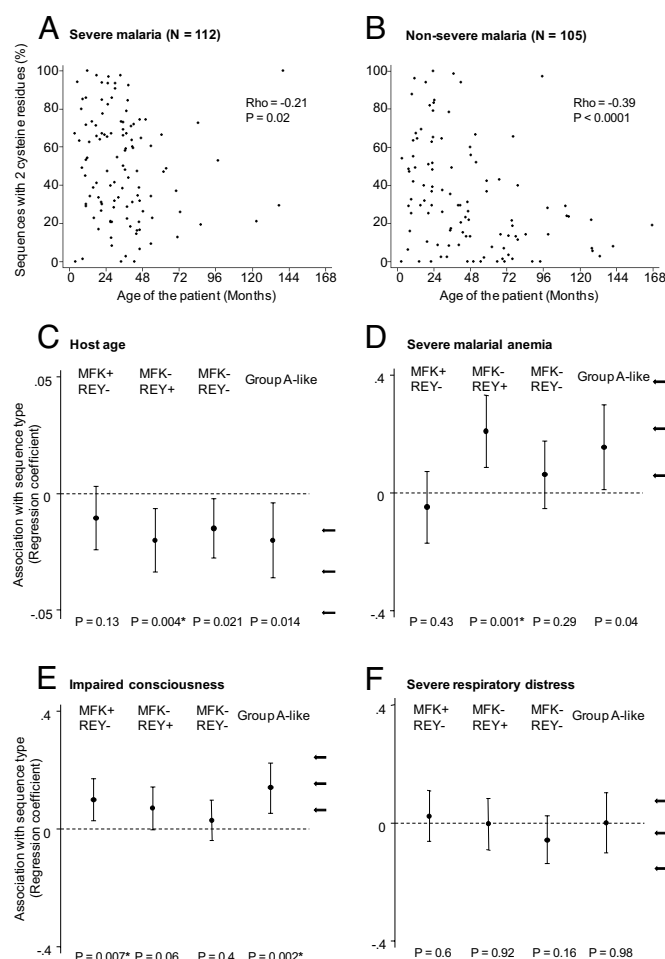


Fig. 1. Relationship between *cys2 var* expression, host age, and malaria syndromes. For both severe (A) and nonsevere (B) malaria, the proportion of expressed *cys2* tag sequences is plotted against age of the child from whom the isolate was obtained. Spearman's rank correlation ρ and p values are indicated at the bottom of each plot. (C–F) Results of four multiple regression models with one of the *cys2* subgroups (MFK⁺ REY[−], MFK[−] REY⁺, MFK[−] REY[−], and group A-like) as the dependent variable and each with host age and the three severe malaria syndromes as explanatory variables. For each of the four subgroup-specific expression profiles the respective regression coefficients, 95% confidence intervals, and p values for their relationship with host age (C), severe malarial anemia (D), impaired consciousness (E), and severe respiratory distress (F) are shown. These are compared to relationships with total *cys2* expression (black arrows). Statistically significant relationships are indicated (*). Parasitemia at presentation to hospital was not associated with *cys2* expression (Fig. S2) and hence was not included in the models.

suggesting that the association was the result of variation in *var* expression rather than differences in the repertoire of *var* genes between parasite isolates (Fig. S3).

***cys2 var* Expression Is Differentially Associated with Severe Malaria Syndromes.** If the pathogenesis mechanisms for the three severe malaria syndromes are linked to the interaction between PfEMP1 and the host, we would expect this to be evident in the *var* expression profiles of the infecting parasite population. We therefore compared expression of *cys2 var* genes between children with different clinical syndromes in a single multiple linear regression model with *cys2 var* expression as the dependent variable and each of the three syndromes and age as explanatory variables. This allowed us to include patients with overlapping syndromes (Table S2 and Fig. S4). Parasites from patients with severe malarial anemia ($n = 19$) and impaired consciousness ($n = 90$) both showed strong

Table 1. Models predicting disease severity while incorporating host age, *cys2* var expression, and antibodies to the surface of parasite-infected erythrocytes at the time of disease

Model	Variable	Odds ratio	95% CI	<i>p</i> value	LR χ^2 test*	LR χ^2 test†
1	Host age	0.9	0.76–0.96	0.006	—	—
2	<i>cys2</i>	4.5	1.91–10.61	0.0006	12.7, <i>p</i> < 0.0004	—
	Host age	0.9	0.80–1.01	0.08		
3	<i>cys2</i>	3.4	1.38–8.23	0.007	22.2, <i>p</i> < 0.0001	1.95, <i>p</i> = 0.16
	Host age	0.9	0.81–1.04	0.17		
	Antibodies to P6921	0.2	0.06–0.58	0.004		
4	<i>cys2</i>	4.2	1.76–10.11	0.001	13.3, <i>p</i> < 0.0001	2.28, <i>p</i> = 0.13
	Host age	0.9	0.81–1.03	0.14		
	Antibodies to P7148	0.8	0.49–1.38	0.46		
5	<i>cys2</i>	2.8	1.12–6.82	0.03	34.6, <i>p</i> < 0.0001	1.08, <i>p</i> = 0.3
	Host age	0.9	0.82–1.06	0.3		
	Antibodies to P7237	0.2	0.07–0.46	0.0003		
6	<i>cys2</i>	3.5	1.43–8.40	0.01	20.0, <i>p</i> < 0.0001	1.05, <i>p</i> = 0.31
	Host age	0.9	0.83–1.06	0.31		
	Antibodies to P7671	0.7	0.56–0.92	0.009		
7	<i>cys2</i>	4.4	1.85–10.41	0.001	20.0, <i>p</i> < 0.0001	2.86, <i>p</i> = 0.09
	Host age	0.9	0.80–1.02	0.1		
	Antibodies to P7860	0.2	0.05–0.65	0.009		
8	<i>cys2</i>	4.1	1.71–9.79	0.002	22.0, <i>p</i> < 0.0001	1.92, <i>p</i> = 0.13
	Host age	0.9	0.80–1.03	0.17		
	Antibodies to P7063	0.5	0.32–0.81	0.005		
9	<i>cys2</i>	3.6	1.50–8.73	0.004	22.9, <i>p</i> < 0.0001	2.33, <i>p</i> = 0.17
	Host age	0.9	0.80–1.03	0.13		
	Antibodies to P8073	0.1	0.01–0.50	0.01		
10	<i>cys2</i>	4.0	1.70–9.63	0.002	15.3, <i>p</i> = 0.0005	1.88, <i>p</i> = 0.17
	Host age	0.9	0.81–1.04	0.17		
	Antibodies to P7542	0.2	0.04–1.57	0.14		

Presented are odds ratios, 95% confidence intervals, and *p* values from multiple logistic regression models testing the independent association of severe malaria with host age, *cys2* expression, and IgG antibody levels (mean fluorescence intensity) to each of eight clinical isolates. In all models severe malaria is the dependent variable. The relationship between expression of *cys2* var genes and severe malaria is examined while controlling for host age alone (model 2) or host age and IgG antibody levels (models 3–10) as indicated. The parasite used for the flow cytometry IgG antibody assay (P6921, P7148, P7237, P7671, P7860, P7063, P8073, or P7542) in each model is indicated in the left column. The fit of each of the models is compared to the model with host age alone (model 1) or to the respective model with *cys2* expression and antibody levels only using the likelihood ratio test. The χ^2 and *p* values for each test are shown.

*LR χ^2 test when model is compared to the model with host age as the only explanatory variable.

†LR χ^2 test when model is compared to the model with host age removed as an explanatory variable.

To test these predictions, we used a flow cytometry assay on plasma sampled at the time of disease to measure each child's IgG antibody response (acquired as mean fluorescence intensity) to the surface of erythrocytes infected by each of eight parasite isolates. PfEMP1 is thought to be the predominant target for this assay (26). In support of prediction *i*, there was a negative, age-corrected association between *cys2* var expression and IgG antibody levels against six of the eight isolates (Fig. 2*F* and Fig. S5). To test prediction *ii*, we used a multiple logistic regression model to test whether *cys2* var expression predicts disease severity independently of host age. Expression of *cys2* var genes was strongly associated with severe malaria (Table 1, model 2). We then built eight multiple logistic regression models each predicting disease severity using *cys2* var expression, host age, and IgG antibody levels to one of the eight parasite isolates as explanatory variables. In all eight models *cys2* var expression predicted severe malaria whereas IgG responses to the infected erythrocyte surface was negatively associated with disease severity in six of the eight models (Table 1, models 3–10). In support of prediction *iii*, host age was not associated with severe malaria within these models. All models with *cys2* var expression and IgG antibody responses as explanatory variables were better at explaining severe malaria than was a model with host age alone. Finally, when tested using the likelihood ratio χ^2 improvement test, host age did not improve the fit of each of the eight models with *cys2* var expression and IgG antibody responses as explanatory variables (Table 1, right-hand column).

Overall, these results support the idea that *cys2* var genes are both adapted to infections of immunologically naïve children and directly associated with parasite pathogenicity. Rapid acquisition of immunity to this group of PfEMP1 variants might help explain why

children become less susceptible to severe malaria after relatively few infections (8).

Discussion

As with other infectious diseases, much discussion has been generated in the past about whether the malaria parasite population is structured into strains that have variable virulence (27). PfEMP1 presents us with a scenario in which a repertoire of molecules that play a central role in the host–parasite interaction, both through cytoadherence and immunogenicity, appear to be functionally and genetically differentiated within every parasite genome (13, 28), potentially giving each parasite line the ability to alter its pathogenicity depending on the combination of selection pressures experienced within the host (29). However, there is no direct evidence for links between the structure of the PfEMP1 antigen repertoire and a role for PfEMP1 in parasite immune evasion and pathogenicity.

A frequently cited study supporting a link between group A var expression and parasite virulence is based on the in vitro selection of a lab-adapted parasite isolate using pooled serum from semi-immune children (14). The rationale for this experiment was that the surface of parasite-infected erythrocytes from children with severe malaria had previously been shown to be more commonly recognized by antibodies from semi-immune individuals (9, 10). However, in these earlier studies, parasites from young children without severe disease also had a tendency to be better recognized by these antibodies, raising the possibility that group A var genes are merely an antigenically restricted group of genes associated with low exposure to infection. If some PfEMP1 variants have a direct role in pathogenicity we would

predict that the expression of these PfEMP1 variants would be associated independently with disease severity and young host age in clinical parasite isolates. This is a question that we have addressed here.

There is mounting evidence that *cys2 var* genes (including group A genes) may encode an antigenically restricted group of PfEMP1 variants (14, 30). Our data further suggest that, relative to other *var* genes, expression levels of *cys2 var* genes are independently associated with young host age, low anti-PfEMP1 antibody levels, and disease severity. The independent association with severe malaria suggests that this group of PfEMP1 variants may actively contribute to the pathogenesis of life-threatening disease, rather than passively reflect the host's immune status. However, because similar numbers of *cys2 var* genes are present in genomes of parasites sampled worldwide (22) our results beg the question of what initially causes the switch to PfEMP1 variants encoded by *cys2 var* genes. We have explored the hypothesis that, in the absence of a large repertoire of protective antibodies, PfEMP1 variants that are the most competitive will also confer the greatest pathogenicity to the parasite. However, this would not explain the actual mechanism of disease onset, because many children with high *cys2* levels did not have severe malaria. One possibility, suggested previously (19) is that many children with nonsevere malaria and high *cys2 var* expression would go on to get severe malaria if not treated. An alternative possibility is that high *cys2 var* expression merely increases the probability of an event that triggers the onset of severe malaria.

An example of the type of process that may trigger an episode of severe malaria has previously been suggested for ICAM1, whose expression is induced by inflammation (31, 32). ICAM1 induction could select for parasites expressing PfEMP1 variants that can bind ICAM1. If parasite binding were to lead to more inflammation, ICAM1 induction would be increased further. Such a positive feedback loop could trigger an episode of severe malaria.

The idea that a limited repertoire of antigenic molecules is associated with severe malaria is attractive because a vaccine that only needs to target a limited number of PfEMP1 epitopes may be a practical possibility. However a more sobering explanation for our data would come from the fact that most *cys2 var* genes encode long PfEMP1 molecules (12), meaning they have multiple domains. If each domain adds a functional component to the molecule, this would imply that longer PfEMP1 tend to (i) be more versatile in their ability to bind host cells, but also (ii) present more epitopes to the immune system, increasing their chance of being recognized. Such an idea is supported by the observation that severe malaria is associated with multiple binding phenotypes (33).

Nonetheless, cytoadherence phenotypes do appear to be differentially associated with the three major manifestations of severe malaria in African children (34), suggesting a role for distinct host-parasite interactions. The contrasting associations between the MFK motif and impaired consciousness, and the REY motif with severe malarial anemia and young host age may provide additional support for a distinction in the host-parasite interactions in these clinical syndromes. The extent to which this distinction is driven by functional or antigenic differences between *cys2 var* genes with the MFK motif and those with the REY motif remains to be determined. These results, together with the lack of an association between *cys2 var* expression and severe respiratory distress, reinforce the need to account for clinical syndrome in studies examining PfEMP1 expression in severe malaria. It is possible that further consideration of clinical presentation may help explain contrasting observations in the association between group A *var* genes and severe malaria in distant geographical regions (35, 36).

This study had several limitations. First, relatively few patients with severe malarial anemia were sampled. Future studies are therefore needed to determine whether the observed associations hold in regions where a larger proportion of patients have severe malarial anemia (37). Second, the PfEMP1 antibodies were measured in plasma collected at the time of disease. These responses do

not provide information about the dynamics of the host response to the actual parasites causing disease in each patient and cannot distinguish preexisting responses from those that are developing to the current infection. Longitudinal cohort studies may help overcome some of these limitations. Third, our analysis was restricted to transcription profiles of short sequence tags that contain only limited information about the *var* genes from which they were sampled. The development of new sequencing technologies and bioinformatics approaches that allow the assembly of whole *var* gene sequences from transcriptome data are likely to advance our understanding of the role of PfEMP1 in pathogenesis. Finally, what we have measured are potentially dynamic associations taken at a single time point. Establishing whether the observed relationships are stable over space, time, and with changes in malaria transmission is now a priority.

Materials and Methods

Study Site. The study was carried out at Kilifi District Hospital, situated on the coast of Kenya. The clinical spectrum of malaria admissions to the hospital have been described elsewhere (23).

Sample Collection and Clinical Classification of Patients. Following informed consent, children with a primary diagnosis of malaria and parasitemia of one or more trophozoites per 100 uninfected erythrocytes were recruited between August 2003 and September 2007. Because some patients had multiple episodes over the study period, only parasites and clinical data from the first episode were used in the analysis. Severe malaria was defined as hospital admission with any of three main overlapping clinical syndromes previously found to be indicative of life-threatening disease, i.e., impaired consciousness (Blantyre coma score <5), severe malarial anemia (Hb concentration <5 g/dL), and severe respiratory distress (abnormally deep breathing) (23). Patients admitted with none of these manifestations and those who did not require hospital admission were grouped as having nonsevere malaria. Patients with comorbidities [cerebrospinal fluid (CSF) leukocyte count ≥ 10 and/or a positive blood culture for bacteremia] were excluded from the study.

Sampling of Sequence Tags. The methods used to generate DBL α cDNA and genomic DNA sequence tags have been described elsewhere in detail (16).

Sequence Assembly and Classification. Base calling was performed on the sequencing reads using Phred software and low-quality ends of the reads were subsequently clipped using a Perl script we developed (*Method S1*). A second script was written to: (a) check that duplicate reads were correctly assembled into the same contig; (b) exclude sequences that, when translated, encoded a peptide of fewer than 100 aa long; (c) removed the constitutively expressed *var1* sequences from the analysis (*Method S1*); (d) classified the sequences into previously described groups (16, 25); and (e) counted the number of individual bacterial colonies that carried each sequence type.

The MFK/REY grouping system has been described elsewhere (16, 22). A second classification determined whether, in addition to having two cysteine residues, the sequences carried exact matches to a collection of 573 14-aa-sequence blocks within highly polymorphic regions of the sequence. This collection of sequence blocks were previously found to define a distinct group of sequences, block-sharing group 1, that carried most of the known group A *var* genes (38). Using this approach, 9/10 group A *var* genes from HB3 and IT4 are classified as block-sharing group 1-like whereas 0/52 nongroup A genes from these lines is classified as block-sharing group 1-like (25). This represents a sensitivity of 90% and a specificity of 100% for identifying sequence tags from group A *var* genes.

Flow Cytometry. IgG levels (acquired as mean fluorescence intensity) against the surface of erythrocytes infected by each of eight clinical isolates were determined for each patient's plasma collected at the time of disease using flow cytometry as described elsewhere (39). The clinical isolates were selected on the basis of their level of *cys2* expression. *Cys2* expression levels were high in isolates P6921, P7063, P7671, and P7148, whereas levels in isolates P7860, P8073, P7542, and P7237 were negligible (Fig. 2F). For more details, see *Method S2*.

Statistical Analyses. The number of individual clones carrying each *var* sequence type was expressed as a percentage (*P*) of the total number of clones sequenced from each parasite isolate. The following analyses were performed on these percentage expression scores.

Univariate analysis. Spearman's rank correlation was used to assess the relationship between *cys2* var gene expression and host age (Fig. 1 A and B and Fig. S3) and PfEMP1 antibodies at the time of disease (Fig. S5). The Mann-Whitney *U* test was used to compare expression levels of *cys2* var genes between severe and nonsevere malaria in genomic and cDNA samples from 53 parasite isolates (Fig. S3).

Multiple linear regression analysis. To allow use of the percentage expression data in regression analyses the data were transformed by the arcsine transformation ($\sin^{-1} \sqrt{P}$) (40). Multiple linear regression models were used to examine the independent relationship of host age, disease severity, and antibodies to PfEMP1 with *cys2* expression (see text and Fig. 2F). These models were run with *cys2* expression as the dependent variable and host age, PfEMP1 antibodies, and disease severity as explanatory variables. Another set of four multiple regression models with severe malaria classified into the three syndromes and *cys2* sequences classified into the four subgroups (MFK⁺ REY⁺, MFK⁺ REY⁻, MFK⁻ REY⁺, and group A-like) were performed (Fig. 1 C–F). Each of the models was run with the respective *cys2* sequence subgroups as the dependent variable and with host age and severe malaria syndromes as explanatory variables. To allow for the fact that severe malaria syndromes were often overlapping, each syndrome was taken as an independent binary variable in the multiple linear regression models. Bonferroni correction was used to determine a nominal *p* value for statistically significant relationships in the multiple linear regression equations for each of

the four subgroups of *cys2* tag sequences. Thus a *p* value <0.0125 (i.e., (0.05)/4) was used as the cut-off for statistical significance.

Multiple logistic regression analysis. A set of 10 multiple logistic regression models was built to predict disease severity using host age, *cys2* var expression, and PfEMP1 antibodies measured at the time of disease as explanatory variables (Table 1). One model only had host age as the explanatory variable whereas another had both *cys2* var expression levels and host age explanatory variables. Eight models, each with PfEMP1 antibodies to each of the eight parasite isolates, host age, and *cys2* var expression as explanatory variables were also run. The fit of these models was assessed using the likelihood ratio test. Stata version 9.2 was used for all statistical analyses, with the exception of the Fisher's exact test for the mutual exclusivity of MFK and REY motifs which was done in R version 2.8 (<http://www.R-project.org>) (41).

ACKNOWLEDGMENTS. We thank the Wellcome Trust Sanger Institute sequencing operations staff for library construction, preparation of DNA, sequencing and sample loading. We thank B. Kitsao, T. Kazungu, and A. Marshall for help with clinical classification of patients; F. Menza and F. Kanyetta for sample collection and informed consent; and C. Newbold, J. Berkley, M. Mackinnon, B. Urban, and C. Buckee for helpful discussions. This paper is published with the permission of the director of the Kenya Medical Research Institute. P.C.B. and G.M.W. were supported by Wellcome Trust Program Grants 084535 and 077092 and Project Grant 076030.

- Marsh, K (1992) Malaria—a neglected disease? *Parasitology* 104 Suppl, S53–69.
- Gardner MJ, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419:498–511.
- Kyes S, Horrocks P, Newbold C (2001) Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol* 55:673–707.
- Roberts DJ, et al. (1992) Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* 357:689–692.
- Smith JD, et al. (1995) Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 82:101–110.
- Bull PC, et al. (1998) Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med* 4:358–360.
- Giha HA, et al. (2000) Antibodies to variable *Plasmodium falciparum*-infected erythrocyte surface antigens are associated with protection from novel malaria infections. *Immunol Lett* 71:117–126.
- Gupta S, et al. (1999) Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med* 5:340–343.
- Bull PC, et al. (2000) *Plasmodium falciparum*-infected erythrocytes: Agglutination by diverse Kenyan plasma is associated with severe disease and young host age. *J Infect Dis* 182:252–259.
- Nielsen MA, et al. (2002) *Plasmodium falciparum* variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. *J Immunol* 168:3444–3450.
- Lavstsen T, et al. (2003) Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J* 2:27.
- Kraemer SM, et al. (2007) Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: Comparisons of geographically diverse isolates. *BMC Genomics* 8:45.
- Robinson BA, Welch TL, Smith JD (2003) Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. *Mol Microbiol* 47:1265–1278.
- Jensen AT, et al. (2004) *Plasmodium falciparum* associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. *J Exp Med* 199:1179–1190.
- Freitas-Junior LH, et al. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407:1018–1022.
- Bull PC, et al. (2005) *Plasmodium falciparum* variant surface antigen expression patterns during malaria. *PLoS Pathog* 1:e26.
- Taylor HM, et al. (2000) A study of var gene transcription in vitro using universal var gene primers. *Mol Biochem Parasitol* 105:13–23.
- Kirchgatter K, Portillo Hdel A (2002) Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. *Mol Med* 8:16–23.
- Kyriacou HM, et al. (2006) Differential var gene transcription in *Plasmodium falciparum* isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol* 150:211–218.
- Normark J, et al. (2007) PfEMP1-DBL1alpha amino acid motifs in severe disease states of *Plasmodium falciparum* malaria. *Proc Natl Acad Sci USA* 104:15835–15840.
- Falk N, et al. (2009) Analysis of *Plasmodium falciparum* var genes expressed in children from Papua New Guinea. *J Infect Dis* 200:347–356.
- Bull PC, et al. (2007) An approach to classifying sequence tags sampled from *Plasmodium falciparum* var genes. *Mol Biochem Parasitol* 154:98–102.
- Marsh K, et al. (1995) Indicators of life-threatening malaria in African children. *N Engl J Med* 332:1399–1404.
- Marsh K, Snow RW (1997) Host-parasite interaction and morbidity in malaria endemic areas. *Philos Trans R Soc Lond B Biol Sci* 352:1385–1394.
- Bull PC, et al. (2008) *Plasmodium falciparum* antigenic variation. Mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol Microbiol* 68:1519–1534.
- Beck HP, Tetteh K (2008) Molecular approaches to field studies of malaria. *Trends Parasitol* 24:585–589.
- Gupta S, et al. (1994) Parasite virulence and disease patterns in *Plasmodium falciparum* malaria. *Proc Natl Acad Sci USA* 91:3715–3719.
- Kraemer SM, Smith JD (2003) Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum* var gene family. *Mol Microbiol* 50:1527–1538.
- Bull PC, et al. (2005) *Plasmodium falciparum* antigenic variation: Relationships between in vivo selection, acquired antibody response, and disease severity. *J Infect Dis* 192:1119–1126.
- Cham GK, et al. (2009) Sequential, ordered acquisition of antibodies to *Plasmodium falciparum* erythrocyte membrane protein 1 domains. *J Immunol* 183:3356–3363.
- Berendt AR, et al. (1989) Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for *Plasmodium falciparum*. *Nature* 341:57–59.
- Turner GD, et al. (1994) An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J Pathol* 145:1057–1069.
- Heddiini A, et al. (2001) Fresh isolates from children with severe *Plasmodium falciparum* malaria bind to multiple receptors. *Infect Immun* 69:5849–5856.
- Newbold C, et al. (1997) Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. *Am J Trop Med Hyg* 57:389–398.
- Rottmann M, et al. (2006) Differential expression of var gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infect Immun* 74:3904–3911.
- Kaestli M, et al. (2006) Virulence of malaria is associated with differential expression of *Plasmodium falciparum* var gene subgroups in a case-control study. *J Infect Dis* 193:1567–1574.
- Snow RW, et al. (1997) Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *Lancet* 349:1650–1654.
- Trimnell AR, et al. (2006) Global genetic diversity and evolution of var genes associated with placental and severe childhood malaria. *Mol Biochem Parasitol* 148:169–180.
- Bejon P, et al. (2009) Analysis of immunity to febrile malaria in children that distinguishes immunity from lack of exposure. *Infect Immun* 77:1917–1923.
- Sheskin, D (2004) *Handbook of Parametric and Nonparametric Statistical Procedures* (CRC Press, Cleveland, OH) 3rd Ed.
- R Development Core Team (2008) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna).