Copyright © 2003 by the Genetics Society of America

Evidence for Diversifying Selection on Erythrocyte-Binding Antigens of *Plasmodium falciparum* and *P. vivax*

Jake Baum,*[,1] Alan W. Thomas[†] and David J. Conway*

**Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom and [†]Department of Parasitology, Biomedical Primate Research Centre, 2280 GH Rijswijk, The Netherlands*

Manuscript received July 5, 2002
Accepted for publication December 19, 2002

ABSTRACT

Malaria parasite antigens involved in erythrocyte invasion are primary vaccine candidates. The erythrocyte-binding antigen 175K (EBA-175) of *Plasmodium falciparum* binds to glycophorin A on the human erythrocyte surface via an N-terminal cysteine-rich region (termed region II) and is a target of antibody responses. A survey of polymorphism in a malaria-endemic population shows that nucleotide alleles in *eba-175* region II occur at more intermediate frequencies than expected under neutrality, but polymorphisms in the homologous domains of two closely related genes, *eba-140* (encoding a second erythrocyte-binding protein) and ψ*eba-165* (a putative pseudogene), show an opposite trend. McDonald-Kreitman tests employing interspecific comparison with the orthologous genes in *P. reichenowi* (a closely related parasite of chimpanzees) reveal a significant excess of nonsynonymous polymorphism in *P. falciparum eba-175* but not in *eba-140*. An analysis of the Duffy-binding protein gene, encoding a major erythrocyte-binding antigen in the other common human malaria parasite *P. vivax*, also reveals a significant excess of nonsynonymous polymorphisms when compared with divergence from its ortholog in *P. knowlesi* (a closely related parasite of macaques). The results suggest that EBA-175 in *P. falciparum* and DBP in *P. vivax* are both under diversifying selection from acquired human immune responses.

INVASION of the erythrocyte by the malaria parasite *Plasmodium falciparum* is a complex process involving specific molecular interactions between the blood stage merozoite and the erythrocyte surface (Chitnis 2001). *P. falciparum* is able to utilize a number of different receptor-ligand interactions to successfully invade the erythrocyte (Mitchell *et al.* 1986; Dolan *et al.* 1994; Okoyeh *et al.* 1999). This is in contrast to the other common human malaria parasite, *P. vivax,* where erythrocyte binding depends on the interaction between the Duffy-binding protein (DBP) and the erythrocyte Duffy antigen (Chitnis 2001). Several *P. falciparum* invasion ligands that may play a role in erythrocyte invasion have been identified (Adams *et al.* 2001; Chitnis 2001). Of particular interest is a family of proteins that share homology with the DBP of *P. vivax* (Adams *et al.* 1992, 2001).

This erythrocyte-binding protein (EBP) family is defined by the presence of particular cysteine-rich regions (Adams *et al.* 1992, 2001). One is located near the N terminus of the extracellular part of the protein and the second is at the C terminus of the extracellular part, adjacent to a transmembrane domain. The N-terminal cysteine-rich region is termed region II (Adams *et al.* 1992) and consists of a duplicated Duffy-binding-like (DBL) domain with homology to the binding region in the DBPs of *P. vivax* and the related malaria parasite of macaques, *P. knowlesi.* The duplicated DBL domains are termed F_1 and F_2, respectively (Adams *et al.* 1992; Figure 1). Five divergent EBP genes in *P. falciparum* have been identified in the *P. falciparum* genome: *erythrocyte binding antigen* (*eba*)*-175* on chromosome 7 (Sim *et al.* 1990); *eba-140* on chromosome 13 (Mayer *et al.* 2001; Thompson *et al.* 2001; Narum *et al.* 2002); *eba-181* on chromosome 1 (Adams *et al.* 2001); *ebl-1* on chromosome 13 (Peterson and Wellems 2000); and ψ*eba-165*, a pseudogene on chromosome 4 (Triglia *et al.* 2001). EBA-175 was the first *P. falciparum* EBP to be characterized (Camus and Hadley 1985; Sim *et al.* 1990), and its binding is dependent on the sialic acid residues and peptide backbone of glycophorin A (GYPA; Sim *et al.* 1994), the major erythrocyte surface sialoglycoprotein. The F_2 domain of EBA-175 region II in particular has been shown to contain the region that binds to GYPA (Sim *et al.* 1994; Ockenhouse *et al.* 2001). Antibodies raised against EBA-175 have been shown to block invasion *in vitro* (Narum *et al.* 2000), and immunization with EBA-175 gives some protection in a nonhuman primate experimental challenge model (Jones *et al.* 2001). However, targeted genetic disruption of the *eba-*

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under the following accession numbers: *Plasmodium falciparum eba-175* sequences, AJ438799–AJ438828; *P. falciparum eba-140* sequences, AJ438830–AJ438853; *P. falciparum* ψ*eba-165* sequences, AJ438854–AJ438878; and *P. reichenowi eba-140,* AJ438829.

[1]*Corresponding author:* Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom.
E-mail: jakebaum@pobox.com

175 gene did not prevent invasion, demonstrating that EBA-175 is not essential (Reed *et al.* 2000). The EBA-140 protein (Mayer *et al.* 2001; Thompson *et al.* 2001; Narum *et al.* 2002) is ~30% identical to EBA-175 across the full protein sequence and plays a role in invasion, binding to the erythrocyte surface via the glycophorin C receptor (Mayer *et al.* 2001; Maier *et al.* 2002). The putative pseudogene ψ*eba-165* contains one or two stop codons (the second being polymorphic) and is transcribed but does not appear to be translated (Triglia *et al.* 2001). Functional characteristics of the EBA-181 and EBL-1 proteins have not yet been determined (Peterson and Wellems 2000; Adams *et al.* 2001).

Protective immune responses that block erythrocyte invasion might be targeted at the EBPs. If acquired immune responses select for polymorphic amino acids in the target antigens, then signatures of such selection ought to be detectable by molecular population genetic tests (Conway *et al.* 2000). A recent comparison of region II sequences from *eba-175* of different *P. falciparum* laboratory isolates with the orthologous region from the chimpanzee malaria parasite *P. reichenowi* [the closest known relative of *P. falciparum* (Escalante and Ayala 1994)] showed evidence for an excess of amino acid polymorphism in this domain within *P. falciparum* (Ozwara *et al.* 2001). Furthermore, antibodies to the region II domain of the protein are commonly detected in sera from individuals in endemic populations (Daugherty *et al.* 1997; Okenu *et al.* 2000).

The existence of divergent genes within the EBP family, including one that is a putative pseudogene, provides an opportunity to investigate in a comparative manner whether selection is operating at particular loci. Differences in the strength and type of selection on the different EBPs may reflect differences in their respective functional importance or immunogenicity. Here a molecular population genetic approach was undertaken to look at DNA sequence diversity in region II from the *eba-175*, *eba-140*, and ψ*eba-165* genes from a single malaria-endemic West African population. Nonsynonymous and synonymous nucleotide polymorphisms in *eba-175* and *eba-140* were also compared to divergence from orthologous genes in *P. reichenowi*. Results indicate that *eba-175* in particular is under diversifying selection in *P. falciparum*. An analysis of the polymorphism in the homologous domain of *P. vivax dbp* (compared to its *P. knowlesi* ortholog) also shows evidence for positive selection.

MATERIALS AND METHODS

DNA samples and DNA sequencing: DNA was obtained from 33 peripheral blood samples from individuals infected with *P. falciparum* malaria in Ibadan, southwestern Nigeria. These were a subset of samples previously used to study other genes (Conway *et al.* 2000; Polley and Conway 2001). Isolates that had previously been shown to have apparently single-clone infections were used, where still available, so that sequence haplotypes could be determined.

From each isolate, region II of the three erythrocyte-binding antigen genes (see Figure 1) was amplified by PCR using forward and reverse primers designed from the published sequences (GenBank) of each gene: *eba-175*, X52524, nucleotides 433–2280 from the start codon of the reference sequence; *eba-140*, AF332918, nucleotides 421–2268; and ψ*eba-165*, AY032735, nucleotides 394–2545. Primers used to amplify these regions were *eba-175*, Fwd 5-GGAAGAAATACTTCA TCTAATAACG-3 and Rev 5-CATCCTTTACTTCTGGACAC ATCG-3; *eba-140*, Fwd 5-CTGAAATATCTATTGGAAAGG-3 and Rev 5-CATTAATACTTATTGGCGTTC-3; and ψ*eba-165*, Fwd 5-CAATACGTTTAAGAGTATAGG-3 and Rev 5-CTTGAG AAGTCAGACTAAGG-3. PCR amplification was carried out in 20-μl volumes containing 1 unit of Expand high-fidelity enzyme (Roche Applied Science, UK), 1× Expand reaction buffer with 1.5 mm $MgCl_2$ (Roche Applied Science, Lewes, UK), 1 μm of each oligonucleotide primer, and between 10 and 50 ng of DNA (a mixture of human and parasite DNA). This was then run through the following temperature cycles where $a°$ represents the annealing temperature, which was 62°, 54°, and 48° for *eba-175, eba-140*, and ψ*eba-165*, respectively: 94° (2 min); 94° (30 sec), $a°$ (30 sec), 68° (2 min) for 10 cycles; 94° (30 sec), $a°$ (30 sec), 68° (2 min + 5 sec per cycle) for 25 cycles; and 5 min at 72° final extension.

Amplification of the *P. reichenowi* region II domain of *eba-140* and attempted amplification of ψ*eba-165* were undertaken with *P. falciparum* primers (as above) using *P. reichenowi* genomic DNA. The region II sequence from the *P. reichenowi eba-175* ortholog (GenBank no. AJ251848) was derived in a previous study (Ozwara *et al.* 2001), and the genomic DNA used here was from the same chimpanzee *P. reichenowi* isolate.

Purified PCR products [prepared with QIAGEN (Crawley, UK) spin columns] were ligated into pGem-T Easy Vector (Promega, Southampton, UK), which was then cloned and grown in JM 109 *Escherichia coli* high-efficiency competent cells (Promega). Purified plasmids containing the relevant inserts were sequenced using internal and plasmid sequencing primers by cycle sequencing with the 3′ BIG DYE dye terminator cycle-sequencing premix kit (Applied Biosystems, Warrington, UK). Sequencing products were run on an ABI Prism 377 DNA sequencer (Perkin-Elmer/Applied Biosystems), and sequences were checked and assembled using Sequence Navigator version 1.0.1 (Perkin-Elmer/Applied Biosystems). All nucleotide singletons were resequenced from new PCR products to confirm that they were not artifacts of amplification, cloning, or sequencing.

Statistical analyses of between- and within-species diversity: Population genetic tests of neutrality were applied to data on region II sequences. Tajima's (1989a) test was used to test for departure from neutrality as measured by the difference between π (observed average pairwise nucleotide diversity) and θ (expected nucleotide diversity under neutrality derived from the number of segregating sites, S). Under balancing selection rare alleles are selected and maintained at intermediate frequencies, elevating π above that expected under neutrality and making the value of the test statistic (D) positive. Fu and Li's (1993) test was used to test for excess or lack of singleton nucleotides by comparing estimates of θ based on the number of singletons *vs.* that derived from S (the D^* index) or π (the F^* index). An excess of intermediate frequency polymorphisms and a lack of rare variants (singletons) result in positive values for D^* and F^*. A third test of neutrality, the McDonald-Kreitman test (McDonald and Kreitman 1991), was used to compare inter- and intraspecific nucleotide changes in region II for *eba-175* and *eba-140*, using the ortholo-

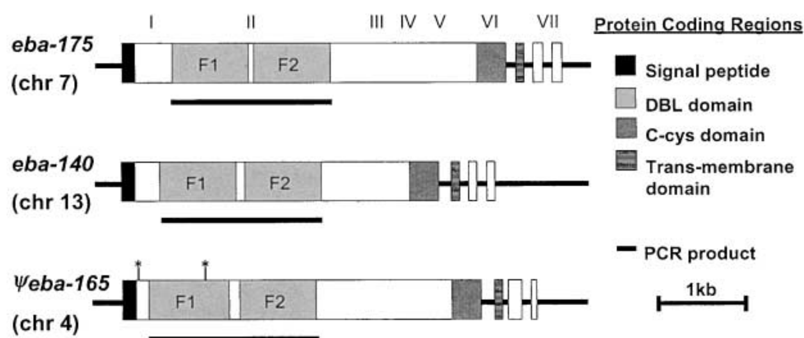

Figure 1.—Scheme of *eba-175*, *eba-140*, and *ψeba-165* genes showing regions studied here (black bars). *eba-175* regions (numbered at top) are as in Adams *et al.* (1992). Regions I–VI encode the extracellular domain with signal peptide, and region VII encodes the putative cytoplasmic domain. The asterisks (*) represent two positions where frameshifts have been observed in *ψeba-165* (Triglia *et al.* 2001).

gous sequences from *P. reichenowi*. A comparison of the ratio of nonsynonymous to synonymous polymorphisms within *P. falciparum* with the ratio for fixed differences between the species reveals if there is a skew in a particular direction (using a 2 × 2 table), tested using Fisher's exact test of significance. Analyses were carried out using DNAsp 3.5 (http://www.bio.ub.es/~julio/DnaSP.html). A McDonald-Kreitman test was also performed on 24 sequences of *P. vivax dbp* region II (GenBank accession nos. AF289480–483, AF289635–653, and AF291096) isolated from Papua New Guinea (Xainli *et al.* 2000), with *P. knowlesi dbpα* (Adams *et al.* 1990; GenBank accession no. M90466) used for the interspecific comparison. Tests based on nucleotide frequency distribution (*e.g.*, Tajima's *D* test) were not performed on the *P. vivax* data set since the possibility of artifactual singletons was not excluded from the published sequences (Xainli *et al.* 2000).

Recombination, linkage disequilibrium, and haplotype structure: The $|D'|$ (Lewontin 1964) and R^2 (Hill and Robertson 1968) indices of linkage disequilibrium were considered quantitatively between sites (including indels), excluding those sites where the rare nucleotide allele was represented less than five times in the population sample. For the single site with three variants, the two rare alleles were lumped together. All values were calculated using DNAsp 3.5, and the relationship between linkage disequilibrium and distance between nucleotide sites was plotted. The relationship between the level of linkage disequilibrium (using the R^2 and $|D'|$ statistics) and genetic distance was tested using the program *Permute* on LDHAT, a package for analyzing patterns of linkage disequilibrium within the framework of coalescent theory (McVean *et al.* 2002). The program is available freely on the Internet: http://www.stats.ox.ac.uk/?mcvean/LDhat/LDhat.html. Sites were included where the frequency of the rare allele was >10%, and 10,000 simulated permutations were carried out to test significance.

The minimum number of recombination events occurring throughout the aligned sequences was calculated, according to the method of Hudson and Kaplan (1985) using DNAsp 3.5. The population recombination parameter $C = 4Nc$, and the population mutation parameter $\theta = 4N\mu$, can be estimated from sequence polymorphism data under the assumption of neutrality (where N is the effective population size; c, the rate of recombination between adjacent base pairs per generation; and μ, the rate of mutation per base pair per generation). An estimate of the number of recombination events per mutation event can therefore be obtained using the ratio of C/θ; *i.e.*, $4Nc/4N\mu = c/\mu$ (Hey and Wakeley 1997; Andolfatto and Przeworski 2001). Two methods for estimating C were used (Hudson 1987; Hey and Wakeley 1997), which have different sensitivities to the number, size, and variability of the underlying sequence data (Hey and Wakeley 1997). θ was estimated on the basis of the proportion of segregating sites in the sample (Watterson 1975). All values were calculated using the program SITES (http://lifesci.rutgers.edu/~heylab) or DNAsp 3.5.

Coalescent simulations of the expected number of haplotypes (K), the haplotype diversity (H), and the frequency of the major haplotype (HP) in a population sample of n sequences with S diallelic polymorphisms were run to test whether the observed haplotype structure in the population sample fitted neutral expectations (Depaulis *et al.* 1999, 2001). The tests (the K-test, H-test, and HP-test, respectively) were run using a recombination rate (6×10^{-7} per site per generation) estimated from a genetic cross of *P. falciparum* (Su *et al.* 1999) and a conservatively estimated effective population size of 10,000 (Anderson *et al.* 2000; Hughes and Verra 2001; Polley and Conway 2001; Conway and Baum 2002). Simulations were carried out using the program ALLELIX available freely on the Internet: http://www.snv.jussieu.fr/moussant. The observed value of H was calculated using DNAsp 3.5 and multiplied by $(n - 1)/n$ in accordance with that used by Depaulis and Veuille (1998).

RESULTS

Analysis of region II sequences of *P. falciparum eba-175*: Sequence polymorphism in a 1848-bp region of *eba-175* [chromosome (chr) 7], a 1848-bp region of *eba-140* (chr 13), and a 2152-bp region of *ψeba-165* (chr 4) was analyzed (Figure 1). A total of 30 region II sequences from *eba-175* were sequenced from the Nigerian study population, among which were 16 different allelic sequences (haplotypes), with 17 segregating (polymorphic) nucleotide sites (Table 1, Figure 2). All substitutions at these polymorphic sites were nonsynonymous, and one site had 3 alleles (nucleotide position 1750, numbering from the start codon of the reference *eba-175* sequence as described in materials and methods). Except for position 676 all the polymorphic sites, including a two-codon deletion (nucleotide positions 1201–1206), have been seen in culture-adapted isolates of *P. falciparum* (Liang and Sim 1997).

The average nucleotide diversity index (π) for *eba-175* region II was 0.003 (*i.e.*, 0.3% nucleotide differences between pairs of alleles on average). The nucleotide frequency distribution was tested for statistical departures from neutral expectations. The overall value of Tajima's D for region II is positive ($D = 1.07$, Table 1), but not significantly different from zero. The overall

values of Fu and Li's D^* and F^* statistics are also positive ($D^* = 0.92$ and $F^* = 1.14$, Table 1), but again not significant. The positive values of both of these statistics indicate that nucleotide alleles occur at more intermediate frequencies than expected with few alleles being rare or near to fixation (Figure 2A). Such an observation is consistent with the action of balancing selection maintaining allelic variation in the population.

The presence of recombination influences the ability to detect selection since it breaks up the associations between sites under selection and linked variation (Charlesworth *et al.* 1997). Therefore, measures of recombination and linkage disequilibrium were investigated across the region of *eba-175* studied. Linkage disequilibrium (LD) as measured by $|D'|$ (Lewontin 1964) and R^2 (Hill and Robertson 1968), was plotted against nucleotide distance between polymorphic sites (Figure 3). Both $|D'|$ and R^2 decline with nucleotide distance, showing negative correlations that are significant ($P < 0.02$), with the high significant values visibly clustered in the top left-hand corner of each plot (Figure 3). This indicates that recombination commonly occurs between *eba-175* alleles as has been seen for the *ama1* (Polley and Conway 2001) and *msp1* (Conway *et al.* 1999) merozoite antigen genes within the same Nigerian population. However, there are some high pairwise values of LD between certain polymorphic sites separated by >800 bp (significant points in the upper right-hand corner of each plot; Figure 3). The LD between these sites is predominantly caused by the high frequency of the allelic sequence type number 11 within the population (Figure 2A), which is the most common allelic sequence and relatively distinct from the others in the population.

To investigate structure in the distribution and frequency of allelic sequences in the data set (considering each allelic sequence type as equivalent to a distinct haplotype), the probability of observing $K \leq 16$ haplotypes and a haplotype diversity of $H \leq 0.789$, given a sample of $n = 30$ sequences showing $S = 17$ diallelic polymorphisms, was determined using Depaulis and Veuille's (1998) *K*- and *H*-tests. The number of haplotypes did not depart significantly from neutral expectations ($P = 0.185$). The haplotype diversity was, however, significantly lower than expected ($P = 0.005$). The higher than expected frequency (13/30 sequences) of allelic sequence type 11 in the population (*HP*-test, $P = 0.008$; Depaulis *et al.* 1999) is the most likely cause of the significant reduction in haplotype diversity. A lower than expected haplotype diversity and higher than expected frequency of the major haplotype indicates that allelic sequence type 11 may have recently increased in the population through selection (Depaulis *et al.* 1999).

The number of recombination events occurring throughout the aligned sequences was calculated according to the method of Hudson and Kaplan (1985) and revealed a minimum number of six recombination

TABLE 1

Genetic variation in region II of *P. falciparum* erythrocyte-binding protein genes

	Size (bp)	n	K	H	S	Singletons	Two variants	Three variants	π	θ	Tajima's D	Fu and Li's D^*	Fu and Li's F^*
eba-175	1848	30	16	0.789	17	1	15	1	0.003	0.002	1.07	0.92	1.14
eba-140	1848	24	8	0.732	7	4	2	1	0.001	0.001	−0.89	−1.65	−1.66
ψeba-165	2152	25	9	0.812	8	4	4	0	0.001	0.001	−1.23	−1.08	−1.31

n, sample size; K, number of haplotypes; H, haplotype diversity multiplied by $(n - 1)/n$ as in Depaulis and Veuille (1998); S, number of segregating sites; π, observed average pairwise nucleotide diversity; θ, the expected nucleotide diversity under neutrality derived from the number of segregating sites (Watterson 1975). All values were calculated using DNAsp 3.5.

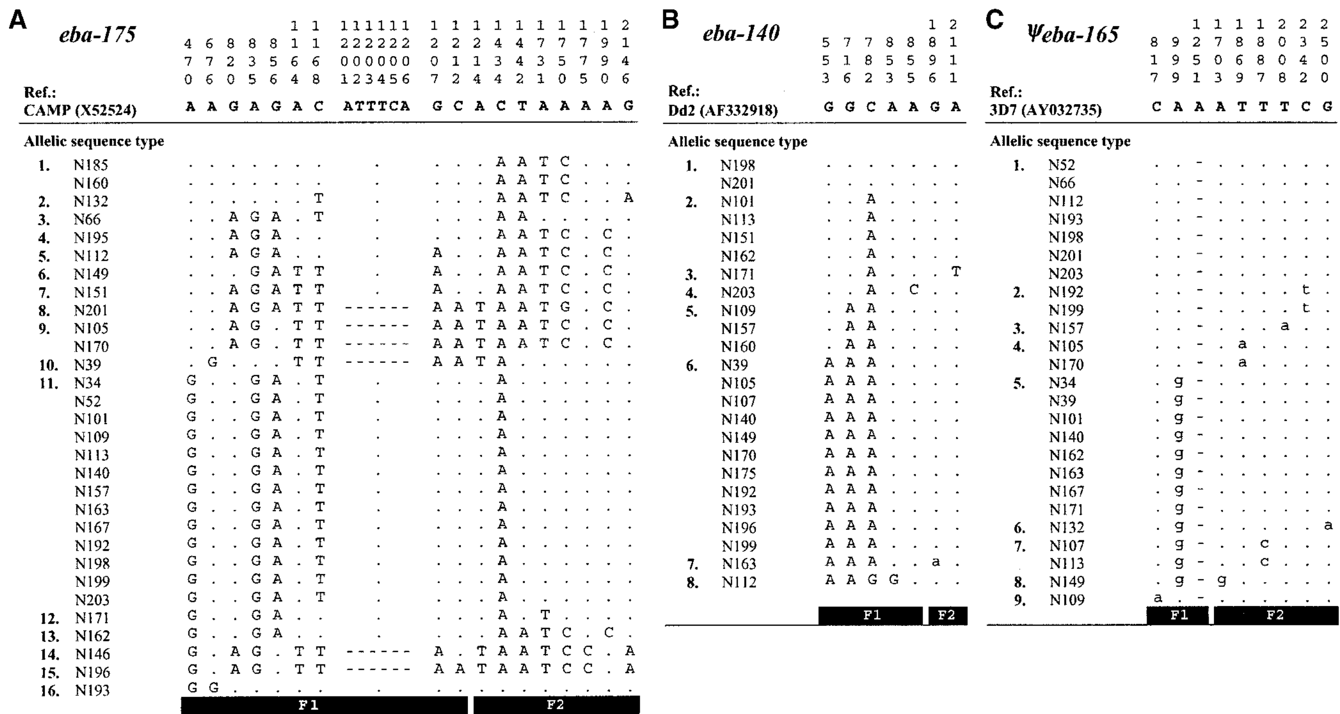

Figure 2.—Polymorphic nucleotide sites in region II of *P. falciparum* genes: *eba-175*, *eba140*, and *ψeba-165*. Nucleotide positions are numbered vertically (conserved positions are not shown), numbering from the start codon of *eba-175*, *eba-140*, and *ψeba-165* reference sequences (X52524, AF332918, and AY032735, respectively). Dot (.) indicates identity with a *P. falciparum* reference sequence. Dash (-) indicates a deletion. Uppercase letters indicate nonsynonymous differences; lowercase letters indicate synonymous (or noncoding in *ψeba-165*) differences. Solid boxes indicate which subdomain (F_1 or F_2) the nucleotides lie in for the respective genes.

events among all 30 *eba-175* region II sequences. Hudson's (1987) and Hey and Wakeley's (1997) estimators of the scaled recombination rate, *C*, are 0.006 and 0.013 between adjacent nucleotides and 10.7 and 24.9 across the whole gene sequence (1848 bp), respectively. The value of θ (Watterson 1975) for region II is 0.002. The ratio of the rate of recombination *vs.* the rate of mutation, c/μ, as estimated by C/θ using Hudson's estimate is 2.35 and using Hey and Wakeley's estimate is 5.81. This suggests that recombination is occurring between two and six times more often than mutation in region II. Together, the value of the C/θ ratio and the significant decline of LD across region II of *eba-175* indicate that recombination is common between *eba-175* alleles. The evidence of recombination suggests that phylogeny-based statistical approaches for detecting selection (Yang and Bielawski 2000) are not appropriate for analyzing intraspecific sequence variation here.

Two descriptive features of amino acid polymorphisms in *eba-175* are worth noting. First, following the two-codon deletion at nucleotide positions 1201–1206, the next three codons all contain nonsynonymous polymorphisms (GAA^{Glu} to AAA^{Lys}, AAC^{Asn} to AAA^{Lys}, and AAG^{Lys} to ATG^{Met}; Figure 2). Second, there is a concordant distribution of amino acid polymorphisms between cysteine residues 5 and 6 in both the F_1 and F_2 subdomains (Figures 2 and 4). The polymorphic nucleotide sites in the F_1 subdomain (at nucleotide positions 820, 835, and 856) are located at similar positions and result in amino acid changes similar to those in the F_2 subdomain (at nucleotide positions 1731, 1750, and 1775; Figure 4). These amino acid polymorphisms are conservative with respect to their resulting amino acid changes (Grantham 1974; with the exception of site 1775, where the substitution of a Glutamic acid residue (acidic) for an Alanine (nonpolar) may have moderate effects on protein function). The mutations in both subdomains occur in a region homologous to that known to be important for *P. vivax* and *P. knowlesi* DBP binding to the erythrocyte surface (Ranjan and Chitnis 1999), and the three in the F_2 subdomain overlap with synthetic peptides of *eba-175* that inhibit human erythrocyte binding and GYPA receptor recognition by EBA-175 (Ockenhouse *et al.* 2001).

Analysis of region II sequences of *P. falciparum eba-140* and *ψeba-165*: A total of 24 sequences were sampled for region II of *eba-140* (Table 1, Figure 2B). These consisted of 8 different allelic sequences (haplotypes) and contained seven polymorphic sites, with one site having three variants (nucleotide position 782 of the reference sequence as described in materials and methods). Of the seven sites, six had nonsynonymous

FIGURE 3.—Levels of linkage disequilibrium (LD) across *eba-175* region II. LD between all possible pairs of polymorphic sites measured using R^2 (top) and $|D'|$ (bottom) as a function of physical distance for EBA-175 region II is shown. Sites were excluded where frequency of the rarer allele was <10% in the sample. Solid circles indicate sites showing significant LD between them (Fisher's exact test, $P < 0.05$); open circles indicate nonsignificant LD. The highest significant values cluster in the top left-hand corner of each plot. Correlation coefficients (ρ) and significance values estimated by 10,000 permutations (see MATERIALS AND METHODS) are displayed on each plot.

```
    8                   8                         8
    2                   3                         5
    0                   5                         6   %
    GAA ATA AGT GAA CAT AAA ATT AAA AAT TTT AGA AAA GAA 17
    ... ... ... ... ... G.. ... ... ... ... ... ... A.. 53
    A.. ... ... ... ... G.. ... ... ... ... ... ... A.. 17
    A.. ... ... ... ... G.. ... ... ... ... ... ... ... 13

    Glu Ile Ser Glu His Lys Ile Lys Asn Phe Arg Lys Glu
     .   .   .   .   .  Glu  .   .   .   .   .   .  Lys
    Lys  .   .   .   .  Glu  .   .   .   .   .   .  lys
    Lys  .   .   .   .  Glu  .   .   .   .   .   .   .

  1                         1                                 1
  7                         7                                 7
  3                         5                                 7
  1                         0                                 5   %
AAA TAT GTT CAC AGG AAT AAA AAA AAT GAT AAG CTT TTT CGT GAT GAG 53
..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... 3
..T ... ... ... ... ... ... C.. ... ... ... ... ... ... ... ... 33
..T ... ... ... ... ... ... C.. ... ... ... ... ... ... ... .C. 7
..T ... ... ... ... ... ... G.. ... ... ... ... ... ... ... ... 3

Lys Tyr Val His Arg Asn Lys Lys Asn Asp Lys Leu Phe Arg Asp Glu
Asn  .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
Asn  .   .   .   .   .   .  Gln  .   .   .   .   .   .   .   .
Asn  .   .   .   .   .   .  Gln  .   .   .   .   .   .   .  Ala
Asn  .   .   .   .   .   .  Glu  .   .   .   .   .   .   .   .
```

FIGURE 4.—Alignment of three polymorphisms occurring between cysteine residues 5 and 6 in the F_1 (top) and F_2 (bottom) subdomains of *eba-175*. Polymorphic nucleotide positions are numbered above each region (numbering from reference sequence start codon). Dot (.) indicates identity with reference sequence (in boldface type). % indicates frequency of alleles in the Nigerian population.

polymorphisms and a single site had a synonymous polymorphism (site 1896). Except for a nonsynonymous mutation at position 855, the polymorphic sites within the F_1 region, including all three alleles at position 782, have been seen in isolates of *P. falciparum* from both wild and laboratory-maintained isolates (MAYER *et al.* 2002). A total of 25 sequences were sampled for region II of ψ*eba-165* (Table 1, Figure 2C), among which were 9 different allelic sequence types (haplotypes). None of the sequences contained the additional adenine at nucleotide position 1251 reported in an isolate of 3D7, in which it would lead to a second stop codon (TRIGLIA *et al.* 2001). The status of the reported stop codon that causes truncation upstream of region II was not investigated (TRIGLIA *et al.* 2001). Tajima's and Fu and Li's tests of neutrality for both *eba-140* and ψ*eba-165* region II give negative values in contrast to the positive values determined for *eba-175* (Table 1). This reflects the fact that the two loci have low-frequency polymorphisms, both having four singletons, whereas *eba-175* had only one singleton (Table 1, Figure 2). There is therefore no indication that either *eba-140* or ψ*eba-165* is under balancing selection that maintains polymorphism, but rather a weak trend in the opposite direction. The low level of polymorphic nucleotide sites in both *eba-140* and ψ*eba-165* precludes estimation of the recombination rate and linkage disequilibrium for these genes. The observed haplotype number, haplotype diversity, and the frequency of the major haplotype for *eba-140* and ψ*eba-165* were not significantly different from those predicted under neutrality ($P > 0.05$ for all tests).

Comparison between intraspecific nucleotide diversity in *P. falciparum eba-175* and *eba-140* and interspecific divergence from *P. reichenowi*: The ortholog of *eba-175* in *P. reichenowi* has been identified in a previous study (OZWARA *et al.* 2001) and shows 82% predicted amino acid identity with *P. falciparum* in region II. PCR amplification using primers for *P. falciparum eba-140* region II amplified an identically sized fragment from *P. reichenowi* genomic DNA. The predicted amino acid sequence of this fragment has a duplicated DBL domain and shows 92% overall deduced amino acid identity to the *P. falciparum* EBA-140 region II (see supplemental data at http://www.genetics.org/supplemental/). The duplicated DBL domains differ from *P. falciparum* EBA-140 F_1 and F_2 domains by 25/312 (8%) and 23/304 (8%) amino acids, respectively; all cysteine residues are conserved between the two species. No amplification product was obtained using *P. falciparum* ψ*eba-165*-specific primers on *P. reichenowi* genomic DNA, suggesting that

TABLE 2

McDonald-Kreitman test of neutrality for region II of *P. falciparum* and *P. reichenowi* EBPs and for region II of *P. vivax* and *P. knowlesi* DBPs

	Interspecific fixed nucleotide differences		Intraspecific (*Pf* or *Pv*) nucleotide polymorphisms		
	Syn	Nsyn	Syn	Nsyn	*P* value[a]
Pfeba-175 vs. *Preba-175*	47	112	0	17	0.007*
Pfeba-140 vs. *Preba-140*	18	50	1	7	0.671
Pvdbp vs. *Pkdbp*α	75	153	1	16	0.016*

Pf, *P. falciparum*; *Pr*, *P. reichenowi*; *Pv*, *P. vivax*; *Pk*, *P. knowlesi*; Syn, synonymous; Nsyn, nonsynonymous substitution.

[a] *P* value from Fisher's exact test. *Significant.

either the pseudogene is absent or sequence differences at the primer-annealing sites prevented successful amplification.

The sequences of *eba-175* region II derived here differ from *P. reichenowi eba-175* by 158 fixed nucleotide differences (47 synonymous and 112 nonsynonymous). When the ratio of the synonymous to nonsynonymous fixed differences is compared to the ratio of polymorphisms within the population of Nigerian isolates (0 synonymous and 17 nonsynonymous) in a 2×2 table (McDonald and Kreitman 1991), there is significant evidence for an excess of nonsynonymous polymorphism within *P. falciparum* (Fisher's exact $P = 0.007$; Table 2). This is consistent with, and more highly significant than, a McDonald-Kreitman test previously applied to *P. falciparum* sequences derived from laboratory isolates (Ozwara *et al.* 2001) and indicates that there has been positive diversifying selection in *eba-175* region II of *P. falciparum.* In contrast, a comparison of polymorphism and interspecific divergence in *eba-140* region II shows no statistically significant difference in the ratios (Table 2), thus yielding no evidence of positive selection.

Comparison between intraspecific nucleotide diversity in *P. vivax dbp* and interspecific divergence from *P. knowlesi dbp*α: As a separate comparison, polymorphism in 24 sequences of region II for the *P. vivax* DBP from Papua New Guinea (Xainli *et al.* 2000) was compared with the divergence from region II of the closely related DBP of *P. knowlesi dbp*α (Adams *et al.* 1990) in a McDonald-Kreitman test. This analysis shows an excess of nonsynonymous polymorphism within *P. vivax* (Fisher's exact $P = 0.016$; Table 2). Analyses using region II from the other *P. knowlesi* Duffy-binding-like genes *P. knowlesi dbp*β and -γ, which are slightly less similar to *P. vivax* (Adams *et al.* 1990, 1992), were also significant ($P = 0.047$ and 0.047, respectively). The similarity between the significant results seen with *P. falciparum* EBA-175 and the *P. vivax* DBP suggests that both proteins are under positive diversifying selection within the species, in contrast to EBA-140 and ψ*eba-165* that do not show evidence of selection.

DISCUSSION

Analyses of sequence diversity in malaria parasite erythrocyte-binding antigens reveal signatures of intraspecific diversifying selection. Nucleotide polymorphisms in region II of the *eba-175* gene of *P. falciparum* have alleles with higher frequencies than expected under neutrality, and comparison with the orthologous region II domain of *P. reichenowi eba-175* reveals a significant excess of nonsynonymous polymorphism within *P. falciparum.* This indicates that balancing selection has operated on region II of the EBA-175 protein, to maintain alleles within the parasite population. A similar result is not seen with allelic sequences of region II from *eba-140*, a functional homolog of *eba-175*, or with the putative pseudogene ψ*eba-165*. The contrast between the frequency of allelic variants as measured by the direction of Tajima's and Fu and Li's statistics for *eba-175* (positive) and *eba-140* and ψ*eba-165* (negative) indicates that evidence for maintenance of variation in *eba-175* is not an artifact resulting from changes in population size during parasite history. Demographic changes, which can confound these statistics (Tajima 1989b), would be expected to affect variation across the genome and not at individual genetic loci.

The Duffy-binding protein gene (*dbp*) of *P. vivax* also shows an excess of nonsynonymous *vs.* synonymous polymorphism, when compared to divergence with its *P. knowlesi* ortholog (*dbp*α). These data suggest a similar type of selection on EBA-175 and DBP, reflecting the importance of DBP to *P. vivax* invasion [recognition of the Duffy antigen is essential for erythrocyte invasion (Miller *et al.* 1976)] and the primary importance of EBA-175 for *P. falciparum* invasion (Sim *et al.* 1994). The most likely agent driving intraspecific diversification of these antigens is the human acquired immune response.

New alleles of a parasite antigen that arise in the population would potentially be able to avoid immune detection (escape variants) and as such give the parasite a survival advantage leading to the allele's selection and increase in frequency within the population. The positive Tajima's *D* value for EBA-175 supports such a hypothesis of immune selection acting in a negative frequency-dependent manner. Additionally, the elevated frequency of allelic sequence type 11 in the population sample may indicate a relatively recent selective increase of a new variant, with the allelic sequence type containing (or being linked to) a site that determines an ability to avoid immune detection. The increase in its frequency is unlikely to be associated with the recent selective sweep of a chloroquine resistance allele of the *chloroquine resistance transporter* gene (*Pfcrt*) on chromosome 7 (Wootton *et al.* 2002) since the two genes are separated by a genetic distance of >900 kb [where 17 kb corresponds to 1 cM (Su *et al.* 1999) and the chromosome length is 1.4 Mb], between which recombination is very likely to disrupt any linkage. Further understanding of the process of immune-mediated selection of merozoite antigens will be important for strengthening this selective hypothesis. Insights from other pathogen models are likely to be of particular use, such as recent work showing that host cytotoxic T-lymphocyte (CTL) responses select for SIV viral escape variants during infection as evidenced by the preferential accumulation of amino acid replacements in viral CTL epitopes (Allen *et al.* 2000). In addition, computer models of malaria infection incorporating parameters such as parasite growth, mutation and recombination rates, and how host immunity develops will be important for investigating the emergence of escape variants in the parasite population and generating expectations for patterns of genetic variation seen in parasite surface antigens.

There is no evidence of selection on EBA-140. This protein, which is apparently involved in multiple invasion pathways (Mayer *et al.* 2001, 2002; Maier *et al.* 2002), might play a less important role in erythrocyte invasion or might be less exposed to the immune system, a factor that may affect its candidacy as a malaria vaccine antigen. The presence of an ortholog to *eba-175* and *eba-140* in *P. reichenowi* indicates that duplication and divergence of the EBP genes occurred before the ancestral split leading to *P. falciparum* and *P. reichenowi* and suggests that other orthologous EBPs may be identifiable within *P. reichenowi*. The inability to amplify a *P. reichenowi* ortholog of ψ*eba-165* here does not exclude the possibility that this might be identified with further efforts.

For region II of *eba-175* the ratio of C/θ (an estimate of the ratio of the biological parameters c/μ, the recombination *vs.* mutation rate) suggests that recombination is occurring between two and six times more often than mutation. This is comparable to that derived for another *P. falciparum* merozoite antigen gene *ama-1* ($C/\theta \sim 7$), using sequences from the same Nigerian population (Polley and Conway 2001). Both are higher than those found in *Drosophila melanogaster* (median for 24 loci, ~1.5; Andolfatto and Przeworski 2001), humans (~1.3; Hey and Wakeley 1997), and *Saccharomyces cerevisiae* (1.24; Jensen *et al.* 2001). However, they are lower than expected given the available laboratory estimates of the *P. falciparum* genome recombination rate (1 cM per 17 kb ≈ c as 6×10^{-7}; Su *et al.* 1999) and spontaneous mutation rates ($\mu = 2.5 \times 10^{-9}$; Paget-McNicol and Saul 2001), which give a ratio of c/μ of ~240. The incongruity between the two values is likely to represent the combined effect of a number of different evolutionary processes (Andolfatto and Przeworski 2001; Jensen *et al.* 2001). For example, moderate population subdivision and inbreeding will both reduce the frequency at which different alleles meet and recombine (Paul *et al.* 1995), increasing disequilibrium between alleles and therefore reducing the value of C/θ. The presence of recent positive selection (a selective sweep) is also likely to reduce local variation through genetic hitchhiking and therefore increase linkage disequilibrium (Charlesworth *et al.* 1997), reducing the value of C/θ. This is given some support by the high frequency of allelic sequence type 11 in the population, which has certainly increased the level of LD. Further laboratory estimates of c and μ and analysis of other loci will help to clarify the basis of this difference.

In summary we report significant evidence for positive diversifying selection on the region II domain of the *P. falciparum* invasion ligand EBA-175, which is at least as strong as evidence for selection on the *P. vivax* DBP ligand. Similar signatures are not seen in the paralogous *eba-140* or ψ*eba-165* genes. This suggests a greater importance of EBA-175 in the invasion process of the human erythrocyte and/or as a target of acquired immunity. As such this gives encouragement to development of EBA-175 as a component in a multivalent malaria vaccine (Jones *et al.* 2001). However, it also clearly identifies the need to understand the allele specificity of its antigenicity and function, in particular the importance of polymorphisms within the F_1 and F_2 domains, and to incorporate this understanding into vaccine design.

We are grateful to Prof. A. M. J. Oduola, Dr. O. A. T. Ogundahunsi, and Dr. C. H. M. Kocken who helped with provision of parasite samples and to Dr. G. A. T. McVean and Dr. S. Mousset for invaluable advice on coalescent simulations. We also thank Spencer Polley for helpful discussion and comments on the manuscript. This work was supported by the Wellcome Trust (Prize Studentship for J.B.) and by the UK Medical Research Council (grant no. G9803180).

LITERATURE CITED

Adams, J. H., D. E. Hudson, M. Torii, G. E. Ward, T. E. Wellems *et al.*, 1990 The Duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. Cell **63:** 141–153.

Adams, J. H., B. K. Sim, S. A. Dolan, X. Fang, D. C. Kaslow *et*

al., 1992 A family of erythrocyte binding proteins of malaria parasites. Proc. Natl. Acad. Sci. USA **89:** 7085–7089.

Adams, J. H., P. L. Blair, O. Kaneko and D. S. Peterson, 2001 An expanding ebl family of *Plasmodium falciparum.* Trends Parasitol. **17:** 297–299.

Allen, T. M., D. H. O'Connor, P. Jing, J. L. Dzuris, B. R. Mothe *et al.*, 2000 Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. Nature **407:** 386–390.

Anderson, T. J., B. Haubold, J. T. Williams, J. G. Estrada-Franco, L. Richardson *et al.*, 2000 Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum.* Mol. Biol. Evol. **10:** 1467–1482.

Andolfatto, P., and M. Przeworski, 2001 Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster.* Genetics **158:** 657–665.

Camus, D., and T. J. Hadley, 1985 A *Plasmodium falciparum* antigen that binds to host erythrocytes and merozoites. Science **230:** 553–556.

Charlesworth, B., M. Nordborg and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided, inbreeding and outcrossing populations. Genet. Res. **70:** 155–174.

Chitnis, C. E., 2001 Molecular insights into receptors used by malaria parasites for erythrocyte invasion. Curr. Opin. Hematol. **8:** 85–91.

Conway, D. J., and J. Baum, 2002 In the blood—the remarkable ancestry of *Plasmodium falciparum.* Trends Parasitol. **18:** 351–355.

Conway, D. J., C. Roper, A. M. Oduola, D. E. Arnot, P. G. Kremsner *et al.*, 1999 High recombination rate in natural populations of *Plasmodium falciparum.* Proc. Natl. Acad. Sci. USA **96:** 4506–4511.

Conway, D. J., D. R. Cavanagh, K. Tanabe, C. Roper, Z. S. Mikes *et al.*, 2000 A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nat. Med. **6:** 689–692.

Daugherty, J. R., C. I. Murphy, L. A. Doros-Richert, A. Barbosa, L. O. Kashala *et al.*, 1997 Baculovirus-mediated expression of *Plasmodium falciparum* erythrocyte binding antigen 175 polypeptides and their recognition by human antibodies. Infect. Immun. **65:** 3631–3637.

Depaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite sites model. Mol. Biol. Evol. **15:** 1788–1790.

Depaulis, F., L. Brazier and M. Veuille, 1999 Selective sweep at the *Drosophila melanogaster Supressor of Hairless* locus and its association with the *In(2L)t* invasion polymorphism. Genetics **152:** 1017–1024.

Depaulis, F., S. Mousset and M. Veuille, 2001 Haplotype tests using coalescent simulations conditional on the number of segregating sites. Mol. Biol. Evol. **18:** 1136–1138.

Dolan, S. A., J. L. Proctor, D. W. Alling, Y. Okubo, T. E. Wellems *et al.*, 1994 Glycophorin B as an EBA-175 independent *Plasmodium falciparum* receptor of human erythrocytes. Mol. Biochem. Parasitol. **64:** 55–63.

Escalante, A. A., and F. J. Ayala, 1994 Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. Proc. Natl. Acad. Sci. USA **91:** 11373–11377.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Grantham, R., 1974 Amino acid difference formula to help explain protein evolution. Science **185:** 862–864.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hughes, A. L., and F. Verra, 2001 Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum.* Proc. R. Soc. Lond. Ser. B Biol. Sci. **268:** 1855–1860.

Jensen, M. A., H. L. True, Y. O. Chernoff and S. Lindquist, 2001 Molecular population genetics and evolution of a prion-like protein in *Saccharomyces cerevisiae.* Genetics **159:** 527–535.

Jones, T. R., D. L. Narum, A. S. Gozalo, J. Aguiar, S. R. Fuhrmann *et al.*, 2001 Protection of *Aotus* monkeys by *Plasmodium falciparum* EBA-175 region II DNA prime-protein boost immunization regimen. J. Infect. Dis. **183:** 303–312.

Lewontin, R. C., 1964 The interaction of selection and linkage. I. General considerations: heterotic models. Genetics **49:** 49–67.

Liang, H., and B. K. L. Sim, 1997 Conservation of structure and function of the erythrocyte-binding domain of *Plasmodium falciparum* EBA-175. Mol. Biochem. Parasitol. **84:** 241–245.

Maier, A. G., M. T. Duraisingh, J. C. Reeder, S. S. Patel, J. W. Kazura *et al.*, 2002 *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nat. Med. **9:** 87–92.

Mayer, D. C., O. Kaneko, D. E. Hudson-Taylor, M. E. Reid and L. H. Miller, 2001 Characterization of a *Plasmodium falciparum* erythrocyte-binding protein paralogous to EBA-175. Proc. Natl. Acad. Sci. USA **98:** 5222–5227.

Mayer, D. C., J. B. Mu, X. Feng, X. Z. Su and L. H. Miller, 2002 Polymorphism in a *Plasmodium falciparum* erythrocyte-binding ligand changes its receptor specificity. J. Exp. Med. **196:** 1523–1528.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

McVean, G. A. T., P. Awadalla and P. Feamhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:** 1231–1241.

Miller, L. H., S. J. Mason, D. F. Clyde and M. H. McGinniss, 1976 The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N. Engl. J. Med. **295:** 302–304.

Mitchell, G. H., T. J. Hadley, M. H. McGinniss, F. W. Klotz and L. H. Miller, 1986 Invasion of erythrocytes by *Plasmodium falciparum* malaria parasites: evidence for receptor heterogeneity and two receptors. Blood **67:** 1519–1521.

Narum, D. L., J. D. Haynes, S. Fuhrmann, K. Moch, H. Liang *et al.*, 2000 Antibodies against the *Plasmodium falciparum* receptor binding domain of EBA-175 block invasion pathways that do not involve sialic acids. Infect. Immun. **68:** 1964–1966.

Narum, D. L., S. R. Fuhrmann, T. Luu and B. K. L. Sim, 2002 A novel *Plasmodium falciparum* erythrocyte binding protein-2 (EBP2/BAEBL) involved in erythrocyte receptor binding. Mol. Biochem. Parasitol. **119:** 159–168.

Ockenhouse, C. F., A. Barbosa, D. P. Blackall, C. I. Murphy, O. Kashala *et al.*, 2001 Sialic acid-dependent binding of baculovirus-expressed recombinant antigens from *Plasmodium falciparum* EBA-175 to Glycophorin A. Mol. Biochem. Parasitol. **113:** 9–21.

Okenu, D. M., E. M. Riley, Q. D. Bickle, P. U. Agomo, A. Barbosa *et al.*, 2000 Analysis of human antibodies to erythrocyte binding antigen 175 of *Plasmodium falciparum.* Infect. Immun. **68:** 5559–5566.

Okoyeh, J. N., C. R. Pillai and C. E. Chitnis, 1999 *Plasmodium falciparum* field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. Infect. Immun. **67:** 5784–5791.

Ozwara, H., C. H. Kocken, D. J. Conway, J. M. Mwenda and A. W. Thomas, 2001 Comparative analysis of *Plasmodium reichenowi* and *P. falciparum* erythrocyte-binding proteins reveals selection to maintain polymorphism in the erythrocyte-binding region of EBA-175. Mol. Biochem. Parasitol. **116:** 81–84.

Paget-McNicol, S., and A. Saul, 2001 Mutation rates in the *dihydrofolate reductase* gene of *Plasmodium falciparum.* Parasitology **122:** 497–505.

Paul, R. E., M. J. Packer, M. Walmsley, M. Lagog, L. C. Ranford-Cartwright *et al.*, 1995 Mating patterns in malaria parasite populations of Papua New Guinea. Science **269:** 1709–1711.

Peterson, D. S., and T. E. Wellems, 2000 EBL-1, a putative erythrocyte binding protein of *Plasmodium falciparum*, maps within a favoured linkage group in two genetic crosses. Mol. Biochem. Parasitol. **105:** 105–113.

Polley, S. D., and D. J. Conway, 2001 Strong diversifying selection on domains of the *Plasmodium falciparum apical membrane antigen 1* gene. Genetics **158:** 1505–1512.

Ranjan, A., and C. E. Chitnis, 1999 Mapping regions containing binding residues within functional domains of *Plasmodium vivax* and *Plasmodium knowlesi* erythrocyte-binding proteins. Proc. Natl. Acad. Sci. USA **96:** 14067–14072.

Reed, M. B., S. R. Caruana, A. H. Batchelor, J. K. Thompson, B. S. Crabb *et al.*, 2000 Targeted disruption of an erythrocyte binding antigen in *Plasmodium falciparum* is associated with a switch toward a sialic acid-independent pathway of invasion. Proc. Natl. Acad. Sci. USA **97:** 7509–7514.

Sim, B. K. L., P. A. Orlandi, J. D. Haynes, F. W. Klotz, J. M. Carter *et al.*, 1990 Primary structure of the 175K *Plasmodium falciparum* erythrocyte binding antigen and identification of a peptide which elicits antibodies that inhibit malaria merozoite invasion. J. Cell Biol. **111:** 1877–1884.

Sim, B. K. L., C. E. Chitnis, K. Wasniowska, T. J. Hadley and L. H. Miller, 1994 Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum.* Science **264:** 1941–1944.

Su, X., M. T. Ferdig, Y. Huang, C. Q. Huynh, A. Liu *et al.*, 1999 A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum.* Science **286:** 1351–1353.

Tajima, F., 1989a Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tajima, F., 1989b The effect of change in population size on DNA polymorphism. Genetics **123:** 597–601.

Thompson, J. K., T. Triglia, M. B. Reed and A. F. Cowman, 2001 A novel ligand from *Plasmodium falciparum* that binds to a sialic acid-containing receptor on the surface of human erythrocytes. Mol. Microbiol. **41:** 47–58.

Triglia, T., J. K. Thompson and A. F. Cowman, 2001 An EBA175 homologue which is transcribed but not translated in erythrocytic stages of *Plasmodium falciparum.* Mol. Biochem. Parasitol. **116:** 55–63.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wootton, J. C., X. Feng, M. T. Ferdig, R. A. Cooper, J. Mu *et al.*, 2002 Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum.* Nature **418:** 320–323.

Xainli, J., J. H. Adams and C. L. King, 2000 The erythrocyte binding motif of *Plasmodium vivax* duffy binding protein is highly polymorphic and functionally conserved in isolates from Papua New Guinea. Mol. Biochem. Parasitol. **111:** 253–260.

Yang, Z., and J. P. Bielawski, 2000 Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15:** 496–503.

Communicating editor: D. Charlesworth