

Allele Frequency-Based Analyses Robustly Map Sequence Sites Under Balancing Selection in a Malaria Vaccine Candidate Antigen

Spencer D. Polley^{*,1} Watcharee Chokejindachai^{*,†} and David J. Conway^{*}

^{*}Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom and [†]Faculty of Tropical Medicine, Mahidol University, Rajthawe, Bangkok 10400, Thailand

Manuscript received April 11, 2003
Accepted for publication June 17, 2003

ABSTRACT

The *Plasmodium falciparum* apical membrane antigen 1 (AMA1) is a leading candidate for a malaria vaccine. Here, within-population analyses of alleles from 50 Thai *P. falciparum* isolates yield significant evidence for balancing selection on polymorphisms within the disulfide-bonded domains I and III of the surface accessible ectodomain of AMA1, a result very similar to that seen previously in a Nigerian population. Studying the frequency of nucleotide polymorphisms in both populations shows that the between-population component of variance (F_{ST}) is significantly lower in domains I and III compared to the intervening domain II and compared to 11 unlinked microsatellite loci. A nucleotide site-by-site analysis shows that sites with exceptionally high or low F_{ST} values cluster significantly into serial runs, with four runs of low values in domain I and one in domain III. These runs may map the sequences that are consistently under the strongest balancing selection from naturally acquired immune responses.

Eet al. 1991; NIELSEN and YANG 1998; YANG *et al.* 2000; JIGGINS *et al.* 2002; ZHANG *et al.* 2002), but they have limitations for some applications. First, phylogenetic relationships among alleles within a species may not be accurately derived if recombination occurs frequently (HOLMES *et al.* 1999; SCHIERUP and HEIN 2000). Second, the probability of mutational changes between particular codons (or overall summarized rates of synonymous and nonsynonymous changes among codons) may not be correctly interpreted, if there are unusual constraints on codon usage that are not completely known and accounted for (AKASHI 1995).

One species that has both a high recombination rate (CONWAY *et al.* 1999; SU *et al.* 1999; POLLEY and CONWAY 2001) and an extreme bias in codon usage (FORSDYKE 2002) is the human malaria parasite *Plasmodium falciparum*. There is a great need to study selection on genes of this species, as the causes of directional selection (particularly due to drug resistance) and frequency-dependent balancing selection (due to acquired immune responses) are important for understanding how

to control malaria in the future (CONWAY *et al.* 2000a; WOOTTON *et al.* 2002). Naturally acquired, nonsterile immunity to *P. falciparum* is seen in highly endemic populations of Africa and Asia (MARSH and SNOW 1997). As antibody and T cell responses involve memory, parasites with rare antigenic types can have a frequency-dependent selective advantage. Frequency-based methods could be applied to detect this type of selection (TAJIMA 1989b; FU and LI 1993; McDONALD 1994). This approach offers the potential to identify targets of naturally acquired immunity (CONWAY *et al.* 2000a; CONWAY and POLLEY 2002) and aid the design of a vaccine to induce immunity against each allelic component (CONWAY 1997).

Apical membrane antigen 1 (AMA1) is a leading candidate in the search for a vaccine against *P. falciparum*. The protein is located in the microneme organelles at the apical end of the merozoite, the stage of the parasite that invades erythrocytes (HEALER *et al.* 2002). It is encoded by a single-locus gene that is essential for parasite growth *in vitro* (TRIGLIA *et al.* 2000). Immunization of mice with AMA1 of *P. chabaudi* can protect against lethal challenge in a strain-specific manner (CREWETHER *et al.* 1996), and antibodies to *P. falciparum* AMA1 have been shown to inhibit erythrocyte invasion *in vitro* with a fairly high degree of strain specificity (HODDER *et al.* 2001; KENNEDY *et al.* 2002; KOCKEN *et al.* 2002). There is a strong predominance of nonsynonymous *vs.* synonymous polymorphisms ($d_N > d_S$) among *ama1* alleles in *P. falciparum* (HUGHES and HUGHES 1995; VERRA and HUGHES 2000), more so than for fixed differences with the closely related species *P. reichenowi* (KOCKEN *et al.* 2000; POLLEY and CONWAY 2001). Constraints on synon-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AJ494866–AJ494915 and ALIGN_000459.

¹Corresponding author: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom.
E-mail: spencer.polley@lshtm.ac.uk

ymous mutation in *P. falciparum* (FORSDYKE 2002) mean that an observation of $d_N > d_S$ is not interpretable simply as positive selection on amino acids, and even the validity of the more robust comparison with *P. reichenowi* in the McDonald-Kreitman test (KOCKEN *et al.* 2000; POLLEY and CONWAY 2001) might be violated if constraints are different in that species (AKASHI 1995).

A recent study of *ama1* single-nucleotide allele frequency distributions in a large sample of sequences from a Nigerian population indicates that polymorphisms are selectively maintained within domains I and III of the surface accessible ectodomain (POLLEY and CONWAY 2001). These domains were previously defined on the basis of the predicted secondary structure of AMA1 (HODDER *et al.* 1996), although their role in the function of AMA1 is currently unknown. It is important to know if these significant results can be generalized to other populations of *P. falciparum* and if the precision of the findings can be further increased. Asian populations of *P. falciparum* have been shown to be distinct from African populations on the basis of microsatellite allele frequencies (ANDERSON *et al.* 2000; CONWAY *et al.* 2001) and mitochondrial haplotype frequencies (CONWAY *et al.* 2000b; JOY *et al.* 2003). A large sample of *ama1* alleles from Thailand has been sequenced so that independent within-population and between-population analyses can be performed to evaluate signatures of balancing selection. Results show that domains I and III of *ama1* are under strong balancing selection in Asia, as well as in Africa, and identify sequential runs of polymorphic sites within these domains that may be of particular importance for the design of an AMA1-based vaccine.

MATERIALS AND METHODS

Sequencing of *ama1* from Thai *P. falciparum* isolates: Genomic DNA was extracted from blood collected from malaria patients participating in clinical studies at the Hospital for Tropical Diseases (Mahidol University, Bangkok) with the approval of the institutional review board. Almost all cases were in the Karen and Maun ethnic groups who had been infected in an endemic area of the Thai-Myanmar border. DNA extractions were performed using the QiAmp DNA blood mini kits (QIAGEN, Chatsworth, CA). A 1371-nucleotide region of *ama1* (codons 144–599, which include the ectodomain), located on chromosome 11, was amplified and sequenced as three overlapping PCR fragments from 50 clinical isolates using the exact methods described previously in a study of a Nigerian population (POLLEY and CONWAY 2001). These isolates were selected from a larger panel of 100 clinical isolates because each contained a single-clone *P. falciparum* infection, as determined by microsatellite typing, thus allowing direct sequencing without cloning into bacterial plasmids. Any regions sequenced in only one direction (including primer sequences) were removed from each of the three fragments before their amalgamation into a single contiguous sequence for each isolate. Raw sequence data were checked and alignments were performed using the Seqman II and MegAlign programs (DNASTAR, Madison, WI). To ensure they were not PCR artifacts, all singletons (single-nucleotide alleles that

occur in only one isolate) were verified by reamplification and resequencing from the genomic template.

Genotyping of microsatellite loci in Thai *P. falciparum* isolates: Alleles at 11 microsatellite loci were typed in the 50 Thai *P. falciparum* isolates using the seminested PCR method described previously for other samples (ANDERSON *et al.* 1999; CONWAY *et al.* 2001). The loci used (and their chromosomal positions) are as follows: *POLYA* (chr. 4), *TA42* (chr. 5), *TA81* (chr. 5), *TA1* (chr. 6), *TA87* (chr. 6), *TA109* (chr. 6), *ARA2* (chr. 11), *PfPK2* (chr. 12), *Pfg377* (chr. 12), *TA102* (chr. 12), and *TA60* (chr. 13).

Analysis of sequence diversity and linkage disequilibrium: Sequence diversity (π , average pairwise nucleotide diversity) was calculated for distinct domains of the *ama1* gene (domains I–III). Analysis of linkage disequilibrium was performed between nucleotide sites at which the frequency of the minority allele was >0.1 using D' (LEWONTIN 1964) and R^2 (HILL and ROBERTSON 1968) indices with Fisher's exact test of significance, calculated via the DNAsp3.53 program (ROZAS and ROZAS 1999). The recombination parameter C (equal to $4Nr$, where N is the effective population size and r is the underlying recombination rate) was estimated using an approximate-likelihood coalescent method via the *pairwise* program (part of the LDhat package downloadable from <http://www.stats.ox.ac.uk/mcvean/LDhat/LDhat.html>), together with the correlation coefficient of both D' and R^2 with distance. A formal test for the presence of recombination was performed by a permutation test: 1000 randomly sampled data sets (in which the order of the polymorphic *ama1* loci was randomly shuffled) were produced and the proportion of these data sets showing a more extreme coefficient of correlation for D' and R^2 with distance was recorded (MCVEAN *et al.* 2002). A separate estimate of the population recombination parameter C was calculated according to the method of HUDSON (1987) together with the minimum number of recombination events (HUDSON and KAPLAN 1985) using the DNAsp3.53 program.

Within-population tests of neutrality: Tajima's D test (TAJIMA 1989b), which compares θ (nucleotide diversity predicted from the number of segregating sites) and π (observed pairwise nucleotide diversity), was used to investigate whether polymorphic single-nucleotide alleles tended to occur at higher or lower frequencies than expected under neutral drift. Fu and Li's F test (FU and LI 1993) was used to compare the number of singleton nucleotides in the *ama1* sequences with the number predicted under neutrality given the average number of nucleotide differences between pairs of alleles and using the *P. reichenowi ama1* sequence as an outgroup (KOCKEN *et al.* 2000). Critical values for the above tests were calculated by coalescent simulations with 10,000 replicates. As recombination tends to make these tests conservative (TAJIMA 1989b; FU and LI 1993; WALL 1999), an additional series of coalescent simulations were run to account for the level of recombination (C) observed in the *ama1* sequences. All of these tests were performed with the DNAsp3.53 program.

Comparison of between-population divergence at different loci: Comparison of within- and between-population diversity in the *ama1* sequence data from the Thai population in this study and a Nigerian population (POLLEY and CONWAY 2001) was performed with the θ -estimator of Wright's fixation index (F_{ST}) of interpopulation variance in allele frequencies (WEIR and COCKERHAM 1984) using the FSTAT program (version 2.9.3.1; GOUDET 1995). For *ama1*, mean F_{ST} values were calculated for the whole region sequenced, as well as each domain separately, together with a nucleotide site-by-site analysis. Of the 65 single-nucleotide sites that were polymorphic in the combined data set, only the 48 sites at which the minority allele(s) had a total frequency >0.1 (in the combined data set) were included for this analysis. The distributions of F_{ST}

TABLE 1
Within-population analysis of *ama1* nucleotide polymorphisms in Thailand

	No. of polymorphic sites	No. of haplotypes	Average pairwise diversity (π)	Tajima's D value	Fu and Li's F value
Total sequence	51	27	0.014	2.12*	1.95*
Domain I	32	18	0.025	1.92*	1.80*
Domain II	6	9	0.006	0.09	0.58
Domain III	6	8	0.014	2.73**	2.00*

Analyses are shown for the whole region sequenced (codons 148–582), domain I (codons 149–302), domain II (codons 320–418), and domain III (codons 443–509). Seven polymorphic sites lie outside the three domains. Three alternative single-nucleotide alleles exist at site 200b; hence the total number of mutations for the whole region and domain I is 52 and 33, respectively. For Tajima's D test and Fu and Li's F test, a value significantly different from zero is marked by an asterisk (* $P < 0.05$; ** $P < 0.01$), as calculated by coalescent simulations with no recombination ($C = 0$). More highly significant results were obtained for domains I and III when recombination was accounted for ($C = 66$).

values for individual sites within each domain of *ama1* and the F_{ST} values of 11 microsatellite loci were compared using the nonparametric Wilcoxon's rank sum test using the SPSS 11.01 program (SPSS, Chicago). The Thai microsatellite data were those obtained in this study, while the Nigerian microsatellite data were those previously reported (CONWAY *et al.* 2001). The distribution of F_{ST} values among the *ama1* polymorphic sites (considered here as a simple spatial series of 48 sequential sites) was also analyzed with Moran's I index of global spatial autocorrelation (MORAN 1950). The I index between neighboring sites was calculated using the ROOKCASE visual basic add-in for Excel (SAWADA 1999) and statistically analyzed by comparison with the number of 1000 Monte Carlo randomly sampled data sets (in which the order of the 48 *ama1* sites was randomly rearranged) that gave a higher value of I .

RESULTS

ama1 sequence polymorphism in Thai *P. falciparum* isolates:

Double-stranded sequence was generated for a 1303-nucleotide portion of the *ama1* gene (codons 148–582) from each of 50 single-clone *P. falciparum* infections. Sequences are available individually (EMBL accession nos. AJ494866–AJ494915) or as an alignment (EMBL-Align database: ALIGN_000459). There were 51 polymorphic sites, 48 of which have been described in previously sequenced *ama1* alleles (KOCKEN *et al.* 2000; ESCALANTE *et al.* 2001; POLLEY and CONWAY 2001). The remaining three polymorphic sites each contain singleton variants (where the rare nucleotide variant was found in only one individual), and these were all confirmed by repeat amplifications from genomic DNA and resequencing. Pairwise diversity among alleles (π) was 0.014 for the whole region sequenced and 0.025, 0.006, and 0.014 for domains I, II, and III, respectively (Table 1).

Recombination and linkage disequilibrium: The effects of recombination on *ama1* alleles in the Thai population are shown graphically by the decline in levels of linkage disequilibrium (D' and R^2 indices) with increasing distance between pairs of nucleotide sites (Figure

1). For D' and R^2 the coefficient of correlation with distance was -0.275 and -0.261 , respectively (none of the 1000 randomly sampled data sets produced more extreme values). Using a maximum-likelihood approach (MCVEAN *et al.* 2002), we estimated the recombination parameter C to be 60 for the whole sequence. A minimum of 16 recombination events were required to ar-

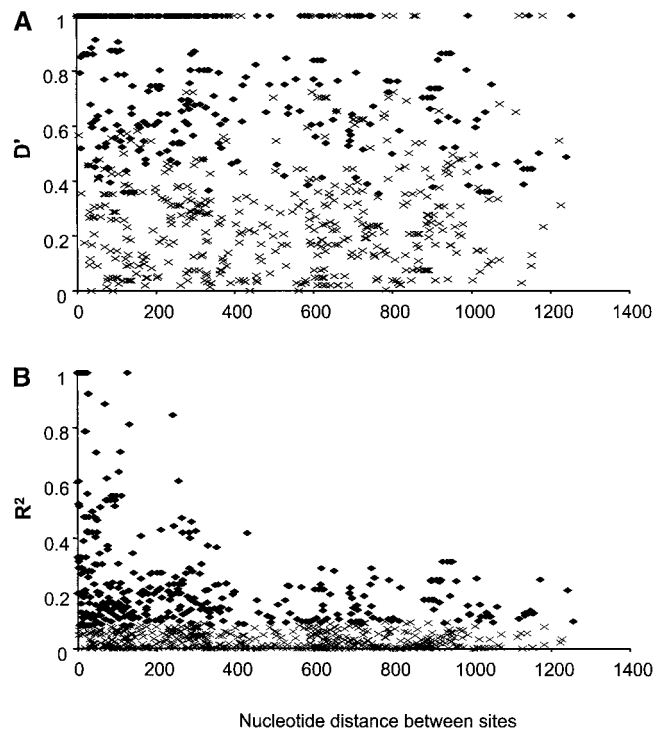


FIGURE 1.—Linkage disequilibrium across *ama1* in the Thai population. The D' and R^2 indices (A and B, respectively) for pairs of sites are plotted against the distance between them. Pairs of sites that show statistically significant linkage disequilibrium are shown by solid diamonds (crosses mark those showing nonsignificant linkage disequilibrium). Only sites containing a minority single-nucleotide allele with a frequency >0.1 were included in the analysis.

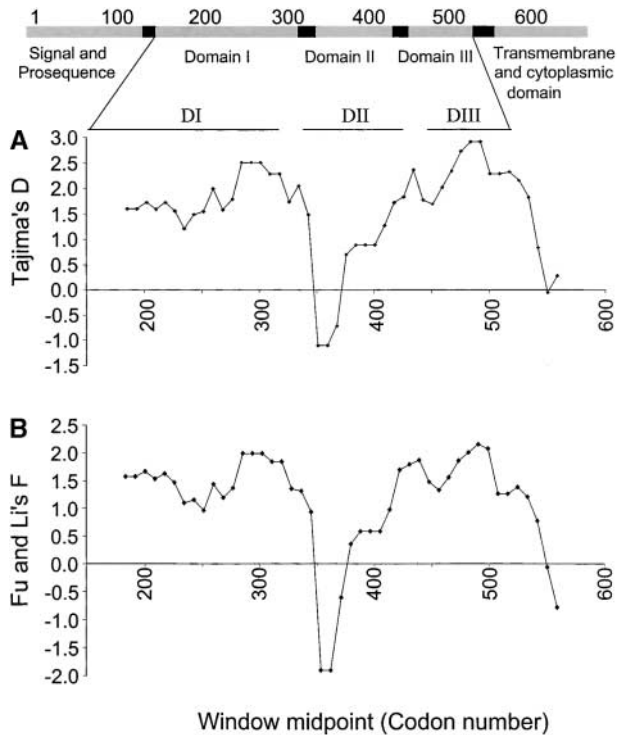


FIGURE 2.—Sliding-window plot of Tajima's D index (A) and Fu and Li's F index (B) based on *ama1* nucleotide polymorphisms in the Thai population. A window size of 200 nucleotides was used with a step size of 25 nucleotides. The three domains of AMA-1 are marked DI, DII, and DIII, respectively.

range the 51 polymorphic sites into the 27 haplotypes seen (HUDSON and KAPLAN 1985). Using the method of HUDSON (1987), we calculated C to be 66 for the whole sequence and 0.051 per adjacent site.

Within-population tests of neutrality: A highly positive Tajima's D value of 2.12 for the entire *ama1* region sequenced shows that single-nucleotide alleles in the Thai population are present at intermediate frequencies that cannot be accounted for by neutral evolution in a constant-size panmictic population (D value is significantly greater than zero, $P < 0.05$).

This departure from neutrality indicates that balancing selection is maintaining single-nucleotide alleles in the population. A sliding-window plot of D shows highly positive values for domains I and III and low values for domain II (Figure 2). There was significant evidence for balancing selection in domains I and III (with D values of 1.92 and 2.73, respectively), but the lower diversity in domain II produced no such trend (with a D value of 0.09; Table 1). When recombination was considered in the calculation of critical D values, even the higher level estimated above ($C = 66$) did not result in a significant departure from neutrality with domain II sequences, but for each of domains I and III the departure was highly significant ($P < 0.01$). Fu and Li's F test shows very similar results, revealing a high

proportion of polymorphisms in domains I and III that are likely to be more ancient than would be expected under neutrality (Table 1; Figure 2).

Between-population comparison of single-nucleotide allele frequencies: An analysis of *ama1* single-nucleotide allele frequencies in this Thai population, together with comparable data from a previously studied Nigerian population (POLLEY and CONWAY 2001), shows that interpopulation divergence accounts for only 4.1% of total nucleotide diversity (mean $F_{ST} = 0.041$). A domain-by-domain analysis shows that F_{ST} values are lowest (*i.e.*, single-nucleotide allele frequencies are most similar) in domain I (mean $F_{ST} = 0.033$) and domain III (mean $F_{ST} = 0.032$), whereas domain II has a much higher mean F_{ST} value (0.113). As a comparison, the allele frequencies of 11 putatively neutral, unlinked microsatellite loci were studied in both populations. These had F_{ST} values (mean $F_{ST} = 0.106$) similar to those of domain II of *ama1*, but much higher than those of domains I and III (Figure 3). Wilcoxon's rank sum test was applied to test for differences in the distribution of F_{ST} values for the microsatellite loci and the individual polymorphic sites within each domain of *ama1*. This nonparametric analysis considers each site as an independent variable, although the F_{ST} values of individual sites within *ama1* may not be truly independent variables due to the physical linkage of the sites. The difference between the microsatellite F_{ST} values and the *ama1* nucleotide F_{ST} values was significant in the case of domain I ($P = 0.006$) and of borderline significance in the case of domain III ($P = 0.056$), but no significant difference was seen with domain II ($P = 0.479$).

The analysis of individual polymorphic sites in *ama1* reveals a great heterogeneity in the F_{ST} values among sites along the gene (Figure 3). However, a significant spatial autocorrelation between neighboring sites is seen when analyzed with Moran's I index (Moran's $I = 0.245$; $P = 0.036$). This is due to clusters of sites with high F_{ST} values (*e.g.*, polymorphic sites in codons 330–395) and others with very low values (*e.g.*, polymorphic sites 162c–172b, 196a–197c, 206a–225b, 283b–308a, and 493b–512b). Despite this clustering, however, sites in adjacent codons can also have very different F_{ST} values (0 and 0.191 for sites 242a and 243a, respectively).

DISCUSSION

The within-population analyses here reveal a significantly nonneutral pattern of *P. falciparum ama1* single-nucleotide allele frequencies in Thailand. The highly positive values of Tajima's D and Fu and Li's F are most likely a result of balancing selection maintaining rare single-nucleotide alleles within domains I and III (no evidence is seen for balancing selection on domain II). This closely concurs with results seen previously in a Nigerian population (POLLEY and CONWAY 2001), confirming the particular effects on domains I and III and

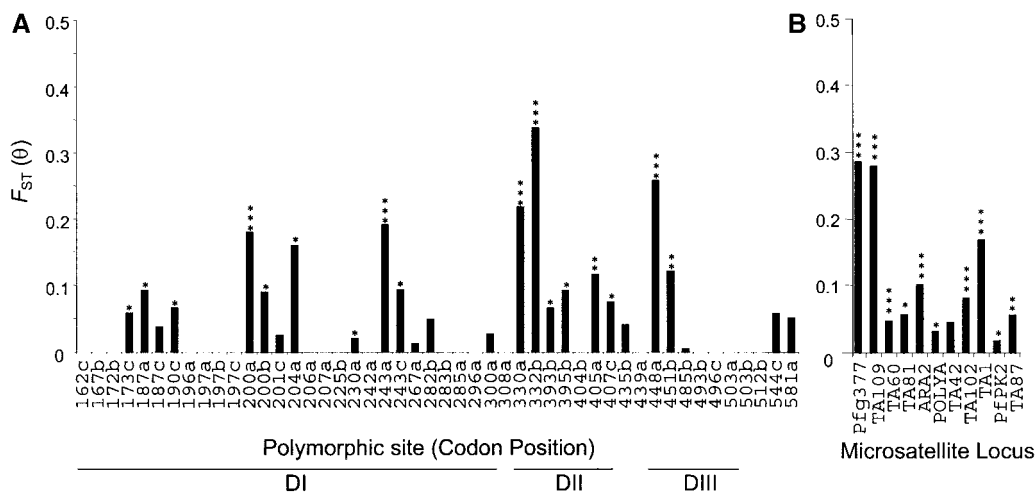


FIGURE 3.—A nucleotide site-by-site analysis of inter-population divergence between Thai and Nigerian populations (F_{ST}) in the *amaI* gene (A) and 11 unlinked microsatellites (B). Sites with an F_{ST} score significantly greater than zero are designated by asterisks (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$). The F_{ST} values of *amaI* sites [labeled as per HODDER *et al.* (1996) with a, b, and c representing the first, second, and third codon positions, respectively] were calculated from sequence data of 50 Thai alleles

(obtained in this study) and 51 Nigerian alleles (POLLEY and CONWAY 2001). Only sites with a minority single-nucleotide allele frequency >0.1 in the combined data set were included in the analysis. Microsatellite allele frequency data for the Thai population were obtained in this study (supplementary table at <http://www.genetics.org/supplemental/>), while microsatellite data for the Nigerian population are as published by CONWAY *et al.* (2001). The calculation of F_{ST} using the θ index here results in slightly negative values for some loci, which are not biologically meaningful and so were set to zero. Mean F_{ST} values for the *amaI* ectodomain as a whole, domain I (DI), DII, and DIII are 0.041, 0.033, 0.113, and 0.032, respectively. The mean F_{ST} value for the 11 microsatellite loci shown in B is 0.106.

indicating that balancing selection is occurring in these different populations. The highly positive values of these indices are unlikely to be due to confounding population effects, as populations of *P. falciparum* are subject to expansions (JOY *et al.* 2003) and will therefore tend to have negative values under neutrality (TAJIMA 1989a) and the endemic populations were each chosen to avoid substructure (TAJIMA 1989b). Neither would codon bias be likely to lead to the positive values seen here, given that purifying selection leads to negative values of these indices (TAJIMA 1989b; FU and LI 1993).

Differences between the Thai and Nigerian *P. falciparum* populations are evident, however, in the amount of recombination seen in *amaI*. The estimated recombination parameters of the Thai population are lower than those of the Nigerian population, with values of C calculated at 66 and 207, respectively, using one method (HUDSON 1987) and 60 and >100 , respectively, by another (MCVEAN *et al.* 2002). At 35 of the polymorphic single-nucleotide sites in *amaI*, the minority alleles are present at frequencies >0.1 in both populations, and these sites can be analyzed in 595 different pairwise combinations for each data set. A significantly higher proportion of these showed statistically significant linkage disequilibrium in the Thai population (278 combinations) compared to the Nigerian population (154 combinations; $P < 0.001$). The difference in recombination rates of the two populations concurs with studies on unlinked microsatellite loci, which indicate that Asian *P. falciparum* populations generally have a lower recombination rate than African populations (ANDERSON *et al.* 2000). This is probably because the endemicity is lower, restricting the proportion of mixed clone infections and

subsequent cross-fertilization and recombination in the mosquito stage of the life cycle (WALLIKER 2000).

Analysis of single-nucleotide allele frequencies between the populations shows that domains I and III of *amaI* have very low mean F_{ST} values, indicating that virtually all polymorphism is seen within each population. The F_{ST} values of these *amaI* domains are significantly lower than the F_{ST} values of 11 unlinked microsatellite loci. In the absence of selection the high allelic diversity at microsatellite loci will generally result in lower F_{ST} values than those for single nucleotide polymorphisms (SNPs; HEDRICK 1999), a point that underscores the significance of this result. Thus polymorphisms in domains I and III of *amaI* are apparently under balancing selection, in concordance with the within-population analyses above.

Analysis of the F_{ST} values on a nucleotide site-by-site basis reveals that they are nonrandomly distributed among the polymorphic sites, and runs of sites with very low values occur within both domains I and III. The resulting fine-resolution map of values among sites may allow the identification of key sites under immune selection, although due to the effects of close physical linkage it is likely that not all the sites within a run are under selection. The variability in the F_{ST} values also emphasizes the importance of full sequence data for this kind of analysis. An analysis employing a limited subset of polymorphic sites sampled across the gene could easily miss informative signatures of selection.

A previous study of *amaI* polymorphism in a Nigerian population concluded with a prediction that there are protective human immune responses to domains I and III (POLLEY and CONWAY 2001). There is now evidence

that antibodies to polymorphic epitopes in AMA1 can indeed inhibit invasion of *P. falciparum* *in vitro* (HODDER *et al.* 2001; KENNEDY *et al.* 2002; KOCKEN *et al.* 2002), and examination of sequence differences between cultured parasite lines suggests that polymorphic sites within domain I in particular are targets of inhibitory antibodies (HODDER *et al.* 2001). It may be suggested that polymorphic changes in domain III have been selected due to their ability to compensate for some loss of function induced by changes within domain I (or vice versa). However, no obvious patterns of linkage disequilibrium between sites in domains I and III exist, and there is no significant correlation between the overall charge of the polymorphic residues within these two domains among the different alleles (Spearman's rank correlation = 0.326; $P = 0.097$). It would appear, therefore, that both domains are independently under balancing selection.

This study provides new data and analyses of a Thai population, together with a novel between-population analysis, which strongly support and extend the evidence for balancing selection on particular sites in domains I and III of the *ama1* gene. Laboratory and field studies of immune responses to AMA1 provide a means of assessing the relative importance of these sequences in acquired immunity. Such studies are intended to allow the rational identification of AMA1 sequences to be represented in a multiallelic vaccine, which this study predicts would be relevant for use in different endemic populations. Thus allele frequency-based analyses offer a powerful and robust way of identifying sequences under selection in vaccine candidate antigen genes of malaria parasites and may be suitable for other endemic pathogens subject to recombination (STOTHARD *et al.* 1998; FUDYK *et al.* 1999; GUBBELS *et al.* 2000; RANNALA *et al.* 2000; SUAREZ *et al.* 2000). Such analyses can also be applied to screen for signatures of selection on a genomic scale (AKEY *et al.* 2002).

We thank colleagues from the Faculty of Tropical Medicine in Mahidol University, Bangkok, for sample collection and Neal Alexander for helpful discussions. This work was supported by the UK Medical Research Council (grant G9803180) and the World Health Organization Special Program for Research and Training in Tropical Diseases (WHO/TDR Research Training grant to W.C.).

LITERATURE CITED

- AKASHI, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.
- AKEY, J. M., G. ZHANG, K. ZHANG, L. JIN and M. D. SHRIVER, 2002 Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**: 1805–1814.
- ANDERSON, T. J. C., X.-Z. SU, M. BOCKAIRE, M. LAGOG and K. P. DAY, 1999 Twelve microsatellite markers for characterisation of *Plasmodium falciparum* from finger prick blood samples. *Parasitology* **119**: 113–125.
- ANDERSON, T. J. C., B. HAUBOLD, J. T. WILLIAMS, J. G. ESTRADA-FRANCO, L. RICHARDSON *et al.*, 2000 Microsatellites reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**: 1467–1482.
- CONWAY, D. J., 1997 Natural selection on polymorphic malaria antigens and the search for a vaccine. *Parasitol. Today* **13**: 26–29.
- CONWAY, D. J., and S. D. POLLEY, 2002 Measuring immune selection. *Parasitology* **125**: S3–S16.
- CONWAY, D. J., C. ROPER, A. M. J. 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.*, 2000a A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat. Med.* **6**: 689–692.
- CONWAY, D. J., C. FANELLO, J. M. LLOYD, B. M. A.-S. AL-JOUBORI, A. H. BALOCH *et al.*, 2000b Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* **111**: 163–171.
- CONWAY, D. J., R. L. D. MACHADO, B. SINGH, P. DESSERT, Z. S. MIKES *et al.*, 2001 Extreme geographical fixation of variation in the *Plasmodium falciparum* gamete surface protein gene Pfs48/45 compared with microsatellite loci. *Mol. Biochem. Parasitol.* **115**: 145–156.
- CREWETHER, P. E., M. L. S. M. MATTHEW, R. H. FLEGG and R. F. ANDERS, 1996 Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. *Infect. Immun.* **64**: 3310–3317.
- ESCALANTE, A. A., H. M. GREBERT, S. C. CHAIYAROJ, M. MAGRIS, S. BISWAS *et al.*, 2001 Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. X. Asembo Bay Cohort Project. *Mol. Biochem. Parasitol.* **113**: 279–287.
- FITCH, W. M., J. M. E. LEITER, X. LI and P. PALESE, 1991 Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA* **88**: 4270–4274.
- FORSDYKE, D. R., 2002 Selective pressures that decrease synonymous mutations in *Plasmodium falciparum*. *Trends Parasitol.* **18**: 411–417.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- FUDYK, T. C., I. W. MACLEAN, J. N. SIMONSEN, E. N. NJAGI, J. KIMANI *et al.*, 1999 Genetic diversity and mosaicism at the por locus of *Neisseria gonorrhoeae*. *J. Bacteriol.* **181**: 5591–5599.
- GOUDET, J., 1995 FSTAT (Version 1.2): a computer program to calculate F statistics. *J. Hered.* **86**: 485–486.
- GUBBELS, M.-J., F. KATZER, G. HIDE, F. JONGEJAN and B. R. SHIELS, 2000 Generation of a mosaic pattern of diversity in the major merozoite-piroplasm surface antigen of *Theileria annulata*. *Mol. Biochem. Parasitol.* **110**: 23–32.
- HEALER, J., S. CRAWFORD, S. RALPH, G. MCFADDEN and A. F. COWMAN, 2002 Independent translocation of two micronemal proteins in developing *Plasmodium falciparum* merozoites. *Infect. Immun.* **70**: 5751–5758.
- HEDRICK, P. W., 1999 Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.
- HILL, W. G., and A. ROBERTSON, 1968 Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**: 226–231.
- HODDER, A. N., P. E. CREWETHER, M. L. MATTHEW, G. E. REID, R. L. MORITZ *et al.*, 1996 The disulfide bond structure of Plasmodium apical membrane antigen-1. *J. Biol. Chem.* **271**: 29446–29452.
- HODDER, A. N., P. E. CREWETHER and R. F. ANDERS, 2001 Specificity of the protective antibody response to apical membrane antigen 1. *Infect. Immun.* **69**: 3286–3294.
- HOLMES, E. C., R. URWIN and M. C. J. MAIDEN, 1999 The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.* **16**: 741–749.
- 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, M. K., and A. L. HUGHES, 1995 Natural selection on Plasmodium surface proteins. *Mol. Biochem. Parasitol.* **71**: 99–113.
- JIGGINS, F. M., G. D. HURST and Z. YANG, 2002 Host-symbiont con-

- flicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Mol. Biol. Evol.* **19**: 1341–1349.
- JOY, D. A., X. FENG, J. MU, T. FURUYA, K. CHOTIVANICH *et al.*, 2003 Early origin and recent expansion of *Plasmodium falciparum*. *Science* **300**: 318–321.
- KENNEDY, M. C., J. WANG, Y. ZHANG, A. P. MILES, F. CHITSAZ *et al.*, 2002 In vitro studies with recombinant *Plasmodium falciparum* apical membrane antigen 1 (AMA1): production and activity of an AMA1 vaccine and generation of a multiallelic response. *Infect. Immun.* **70**: 6948–6960.
- KOCKEN, C. H., D. L. NARUM, A. MASSOUBODJI, B. AYIVI, M. A. DUBBELD *et al.*, 2000 Molecular characterisation of *Plasmodium reichenowi* apical membrane antigen-1 (AMA-1), comparison with *P. falciparum* AMA-1, and antibody-mediated inhibition of red cell invasion. *Mol. Biochem. Parasitol.* **109**: 147–156.
- KOCKEN, C. H. M., C. WITHERS-MARTINEZ, M. A. DUBBELD, A. VAN DER WEL, F. HACKETT *et al.*, 2002 High-level expression of the malaria blood-stage vaccine candidate *Plasmodium falciparum* apical membrane antigen 1 and induction of antibodies that inhibit erythrocyte invasion. *Infect. Immun.* **70**: 4471–4476.
- LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* **49**: 49–67.
- MARSH, K., and R. W. SNOW, 1997 Host-parasite interaction and morbidity in malaria endemic areas. *Philos. Trans. R. Soc. Lond. B* **352**: 1385–1394.
- MCDONALD, J. H., 1994 Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms, pp. 88–100 in *Non-Neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman & Hall, New York.
- MCVEAN, G., P. AWADALLA and P. FEARNHEAD, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.
- MORAN, P., 1950 Notes on continuous stochastic phenomena. *Biometrika* **37**: 17–23.
- NIELSEN, R., and Z. YANG, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929–936.
- POLLEY, S. D., and D. J. CONWAY, 2001 Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 (ama1) gene. *Genetics* **158**: 1505–1512.
- RANNALA, B., W. G. QIU and D. E. DYKHUIZEN, 2000 Methods for estimating gene frequencies and detecting selection in bacterial populations. *Genetics* **155**: 499–508.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SAWADA, M., 1999 ROOKCASE: an Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. *Bull. Ecol. Soc. Am.* **80**: 231–235.
- SCHIERUP, M. H., and J. HEIN, 2000 Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**: 879–891.
- STOTHARD, D. R., G. BOGUSLAWSKI and R. B. JONES, 1998 Phylogenetic analysis of the *Chlamydia trachomatis* major outer membrane protein and examination of potential pathogenic determinants. *Infect. Immun.* **66**: 3618–3625.
- SU, X.-Z., M. T. FERDIG, Y. HUANG, C. Q. HUYNH, A. LIU *et al.*, 1999 A genetic map and recombination parameters of the human malaria parasite *P. falciparum*. *Science* **286**: 1351–1353.
- SUAREZ, C. E., M. FLORIN-CHRISTENSEN, S. A. HINES, G. H. PALMER, W. C. BROWN *et al.*, 2000 Characterization of allelic variation in the *Babesia bovis* merozoite surface antigen 1 (MSA-1) locus and identification of a cross-reactive inhibition-sensitive MSA-1 epitope. *Infect. Immun.* **68**: 6865–6870.
- TAJIMA, F., 1989a The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- TAJIMA, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TRIGLIA, T., J. HEALER, S. R. CARUANA, A. N. HODDER, R. F. ANDERS *et al.*, 2000 Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. *Mol. Microbiol.* **38**: 706–718.
- VERRA, F., and A. L. HUGHES, 2000 Evidence for ancient balanced polymorphism at the apical membrane antigen-1 (AMA-1) locus of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **105**: 149–153.
- WALL, J. D., 1999 Recombination and the power of statistical tests of neutrality. *Genet. Res.* **74**: 65–79.
- WALLIKER, D., 2000 Malaria, pp. 93–114 in *Molecular Epidemiology of Infectious Diseases*, edited by R. C. A. THOMPSON. Arnold, London.
- WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- 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.
- YANG, Z., R. NIELSEN, N. GOLDMAN and A.-M. K. PEDERSEN, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.
- ZHANG, J., Y.-P. ZHANG and H. F. ROSENBERG, 2002 Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* **30**: 411–415.

Communicating editor: D. CHARLESWORTH

