Geographical Distribution of Selected and Putatively Neutral SNPs in Southeast Asian Malaria Parasites

Tim J. C. Anderson, Shalini Nair,* Dan Sudimack,* Jeff T. Williams,* Mayfong Mayxay,†‡ Paul N. Newton,‡§ Jean-Paul Guthmann,‖ Frank M. Smithuis,¶ Tran Tinh Hien,# Ingrid V.F. van den Broek,** Nicholas J. White,§†† and François Nosten§†††‡‡*

*Southwest Foundation for Biomedical Research, San Antonio, Texas; †Faculty of Medical Sciences, National University of Laos, Vientiane, Lao PDR; ‡Wellcome Trust-Mahosot-Oxford Tropical Medicine Research Collaboration, Mahosot Hospital, Vientiane, Lao PDR; §Centre for Tropical Medicine and Vaccinology, Churchill Hospital, Oxford, United Kingdom; ‖Epicentre (Médecins Sans Frontières-France), Paris, France; ¶Artsen Zonder Grenzen, Médecins Sans Frontières-Holland, Yangon, Myanmar; #Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam; **Médecins Sans Frontières (UK), London, United Kingdom; ††Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; and ‡‡Shoklo Malaria Research Unit, Mae Sot, Tak, Thailand

Loci targeted by directional selection are expected to show elevated geographical population structure relative to neutral loci, and a flurry of recent papers have used this rationale to search for genome regions involved in adaptation. Studies of functional mutations that are known to be under selection are particularly useful for assessing the utility of this approach. Antimalarial drug treatment regimes vary considerably between countries in Southeast Asia selecting for local adaptation at parasite loci underlying resistance. We compared the population structure revealed by 10 nonsynonymous mutations (nonsynonymous single-nucleotide polymorphisms [nsSNPs]) in four loci that are known to be involved in antimalarial drug resistance, with patterns revealed by 10 synonymous mutations (synonymous single-nucleotide polymorphisms [sSNPs]) in housekeeping genes or genes of unknown function in 755 *Plasmodium falciparum* infections collected from 13 populations in six Southeast Asian countries. Allele frequencies at known nsSNPs underlying resistance varied markedly between locations ($F_{ST} = 0.18$–0.66), with the highest frequencies on the Thailand-Burma border and the lowest frequencies in neighboring Lao PDR. In contrast, we found weak but significant geographic structure ($F_{ST} = 0$–0.14) for 8 of 10 sSNPs. Importantly, all 10 nsSNPs showed significantly higher F_{ST} ($P < 8 \times 10^{-5}$) than simulated neutral expectations based on observed F_{ST} values in the putatively neutral sSNPs. This result was unaffected by the methods used to estimate allele frequencies or the number of populations used in the simulations. Given that dense single-nucleotide polymorphism (SNP) maps and rapid SNP assay methods are now available for *P. falciparum*, comparing genetic differentiation across the genome may provide a valuable aid to identifying parasite loci underlying local adaptation to drug treatment regimes or other selective forces. However, the high proportion of polymorphic sites that appear to be under balancing selection (or linked to selected sites) in the *P. falciparum* genome violates the central assumption that selected sites are rare, which complicates identification of outlier loci, and suggests that caution is needed when using this approach.

Introduction

Lewontin and Krakauer (1973) originally suggested that patterns of genetic population structure could be used to identify loci that are under selection. They reasoned that loci involved in local adaptation should show greater levels of differentiation between populations than neutrally evolving loci in which allele frequencies are determined by genetic drift alone. For many years, this appealingly simple idea was disregarded largely as a result of problems with the theoretical approximations of the distribution of F_{ST} values of Lewontin and Krakauer (Nei and Maruyama 1975; Robertson 1975*a*, 1975*b*). A number of groups have recently revived this idea, improved the statistical methodology, and used this approach to search for loci involved in adaptation in organisms including humans, fish, intertidal snails, and mice (Beaumont and Nichols 1996; Wilding, Butlin, and Grahame 2001; Akey et al. 2002; Storz and Nachman 2003; Campbell and Bernatchez 2004; Storz and Dubach 2004; Storz, Payseur, and Nachman 2004; Vasemagi, Nilsson, and Primmer 2005). Typically, population structure is assessed at a number of randomly chosen loci, and those showing "extreme" levels of differentiation are identified as putatively selected genes. Both empirical and model-based approaches have been used to identify markers showing extreme patterns of genetic structure suggestive of selection. Work by Akey et al. (2002) exemplifies the empirical approach (see also McDonald 1994). They examined empirical distributions of F_{ST} at 26,530 single-nucleotide polymorphisms (SNPs) in humans and found 156 genes showing higher than expected values and 18 genes showing lower than expected values. The model-based approach has been formulated by Bowcock et al. (1991) and Beaumont and Nichols (1996). These authors used coalescent simulations of metapopulations to generate distributions of F_{ST} values expected under neutrality and used these distributions to detect outlier loci and assess probabilities for deviations from neutrality.

One problem with F_{ST}-based approaches is that we cannot tell how many of the outlier loci are actually evolving neutrally (type I errors) and how many selected loci are not detected (type II errors). Typically, in such studies, large numbers of loci may be compared which means that neutral loci may be misclassified as selected by chance alone. This problem can be minimized by raising the cutoff for statistical significance but comes at the cost of increasing the type II error rate and so is not entirely satisfactory. Beaumont and Balding (2004) examined the efficiency of both Bayesian and coalescent-based approaches for detecting selection in a simulated metapopulation under a wide variety of demographic conditions. They found that both

Key words: *Plasmodium falciparum*, genetic structure, single-nucleotide polymorphism, local adaptation, drug resistance, *pfcrt*, *pfmdr*, *dhfr*, *dhps*.

E-mail: tanderso@darwin.sfbr.org.

Mol. Biol. Evol. 22(12):2362–2374. 2005
doi:10.1093/molbev/msi235
Advance Access publication August 10, 2005

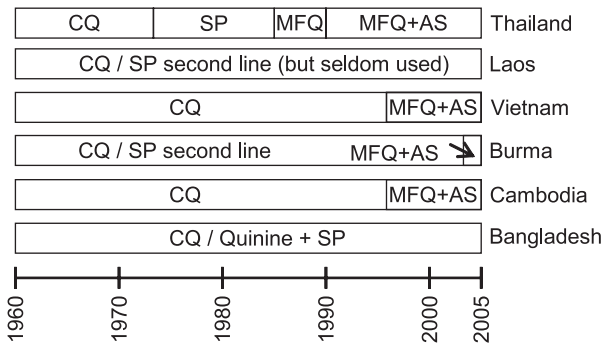

FIG. 1.—Antimalarial treatment policies in Southeast Asian countries. Each bar shows the changes in national treatment policies for the six countries studied, while the year is shown on the scale bar at the base. Abbreviations: SP, sulfadoxine-pyrimethamine; CQ, chloroquine; MFQ, mefloquine; and AS, artesunate. Laos is switching to lumifantrine-artemisinin (coartemether) combination therapy in 2005. While this chart shows national treatment policies, the degree to which this represents drug selection in practice varies between countries and locations within a country. For example, Burma switched to mefloquine-artesunate as a first-line treatment in 2002, but CQ and SP are still the predominant treatments used. Evaluation of country-specific selection on *Plasmodium falciparum* is further complicated by the fact that CQ is still used in many countries for treatment of *Plasmodium vivax* infections, which is likely to impose weak selection for CQ resistance in *P. falciparum*.

approaches detected selected loci efficiently using stringent statistical criteria with minimal type I or type II errors, but only when selection was strong (selection coefficients [s] $>$ 0.1). False negatives were common with weak selection ($s < 0.1$). As a general guideline, they suggest that selected loci may be detected when the selection coefficient is five times greater than the migration rate.

Studies of functional polymorphisms that are known to be under contrasting selection in different populations provide an alternative empirical approach for validating the utility of such between-population comparisons for detecting selected regions of the genome (Taylor, Shen, and Kreitman 1995; Kohn, Pelz, and Wayne 2003). Malaria parasites infecting humans provide an especially useful system in this respect because four different genes are known to be involved in resistance to antimalarial treatments. Selection has been intense as the parasites are host specific and the drugs are very widely used. Furthermore, all the drugs responsible for selection at these loci have been used in Southeast Asian countries over the past 50 years, although drug treatment history differs between countries (fig. 1). For example, mefloquine-artemisinin combination therapy is currently used in Thailand, while in neighboring Lao PDR (Laos) to the east, chloroquine (CQ) and sulfadoxine-pyrimethamine (SP) are still the first- and second-line treatments in most of the country. We have previously described considerable heterogeneity in allele frequencies at the dihydrofolate reductase (*dhfr*) locus on Chr 4 that underlies resistance to pyrimethamine in Southeast Asia (Nair et al. 2003), with some polymorphisms fixed for alternative mutations ($F_{ST} = 1$) in different countries. These geographical patterns were strongly suggestive of selection, but comparative data from other putatively neutral loci in the genome are required to verify this.

In this paper, we analyze patterns of geographical variation at *dhfr* and three other loci that are known to be involved in resistance to antimalarial drugs in 13 parasite populations from Southeast Asia. The additional loci analyzed are (1) dihydropteroate synthase (*dhps*) Chr 8, which underlies resistance to sulfonamides such as sulfadoxine that are used widely in combination with pyrimethamine (Plowe et al. 1997); (2) the CQ resistance transporter (*pfcrt*) (Chr 7), a major gene determining CQ resistance (Fidock et al. 2000; Wootton et al. 2002); and (3) the multidrug resistance gene (*pfmdr*) (Chr 5), an ATP binding cassette (ABC) transporter that influences resistance to multiple drug classes including CQ, quinine (QN), mefloquine (MFQ), and artemisinin derivatives (Reed et al. 2000; Price et al. 2004). To evaluate the relative roles of selection and genetic drift in determining the geographical distributions of SNPs observed at resistance loci, we also examine differentiation in 10 synonymous single-nucleotide polymorphisms (sSNPs) in housekeeping genes or genes of unknown function. We used these data in three ways: (1) to evaluate the efficacy of F_{ST}-based methods for detecting selection, (2) to determine patterns of neutral genetic structure in Southeast Asian malaria parasites, and (3) to map allele frequencies at four key genes involved in antimalarial resistance in Southeast Asia.

Materials and Methods

Geographical Sampling

We collected finger prick blood samples (~50 μl of blood absorbed on Whatmann 3-mm filter paper and dried) from malaria-infected patients from 13 locations in six Southeast Asian countries (Myanmar, Thailand, Cambodia, Laos, Vietnam, and Bangladesh). Ten of these populations have been described previously (Nair et al. 2003). In addition to these, we obtained samples from an additional population in Laos and two locations in Bangladesh. These are all areas of relatively low seasonal transmission. The samples from Laos were collected from Muang Feuang in Ventiane Province in 2000 (Mayxay et al. 2004). The samples from Bangladesh were collected from patients visiting the Médecins Sans Frontières clinics in Dighinala and Panchari (Khagrachari District) in the Chittagong Hill Tracts (van den Broek et al. 2004). These two sites are ~25 km apart. Use of blood samples described in this paper was approved by the institutional review board at the University of Texas Health Science Center at San Antonio and by review boards in the countries where blood samples were collected.

DNA Preparation

We cut 3-mm discs from each blood spot using a sterile hole punch and prepared DNA using the Generation Card Capture Kit (Gentra Systems, Minneapolis, Minn.). We also prepared DNA concurrently from a disc of plain 3-mm Whatman paper. These preparations were then used as negative controls in all amplification experiments. This procedure provides efficient control for contamination during both DNA preparation and amplification. We prepared one negative DNA preparation for every 12 samples prepared.

Genotyping of Mutations in Drug Resistance Genes

We used primer extension to genotype five key mutations in both dihydrofolate reductase (*dhfr*) and

dihydropteroate synthase (*dhps*). These mutations are known to play important roles in the resistance against pyrimethamine and sulfadoxine, the two components of the combination drug SP (Fansidar) (Plowe et al. 1997). Details of the primer extension method are reported elsewhere (Nair et al. 2002). *Pfcrt* alleles conferring resistance to CQ contain multiple point mutations (in codons 72, 74, 75, 76, 220, 271, 326, 356, and 371) of which the *pfcrt-K76T* mutation is critical for resistance. Only codon 76 was genotyped here because the additional mutations were in complete linkage disequilibrium (LD) in a large sample of Thai parasites analyzed previously (Anderson et al. 2005). Both the K76T mutation in *pfcrt* and the N86Y mutation in *pfmdr* were genotyped by *Apo*I digestion fluorescent end-labeled polymerase chain reaction (PCR) products and scoring of product sizes on a capillary sequencer (Anderson et al. 2003). The methods are similar to those used for scoring of sSNP polymorphisms in housekeeping genes (below).

SNPs on Chr 2 and 3

For comparison with the nonsynonymous single-nucleotide polymorphisms (nsSNPs) in known drug resistance genes, we scored point mutations at synonymous point mutations (sSNPs) in housekeeping genes or genes of unknown function on Chr 2 and 3. We used SNPs on just two chromosomes because SNP locations on the other 12 chromosomes of *Plasmodium falciparum* are not yet available (J. Mu and X.-Z. Su, personal communication). However, because linkage disequilibrium decays rapidly with distance in the *P. falciparum* genome (Conway et al. 1999; Anderson 2004), we believe that this is unlikely to substantially bias our results. The Chr 3 SNPs were identified by Mu et al. (2002). They sequenced 204 genes on Chr 3 in a panel of five parasites from Africa (two clones), Southeast Asia, South America, and Papua New Guinea, identifying 62 synonymous mutations. We excluded sSNPs in known antigen genes, in genes containing excessive levels of polymorphism suggestive of selection, or those for which restriction digest assays could not be designed. This left a total of 19 sSNPs. The approximate locations of the Chr 2 SNPs were identified by Volkman et al. (2002) by hybridization with Affymetrix chips. We identified 29 housekeeping genes or genes of unknown function that showed differential hybridization with 25mer oligo probes. We then sequenced these in the four laboratory malaria strains (W2, D6, 7G8, and HB3) studied in the microarray experiments and four samples from Thailand to identify the exact position of the SNPs within the oligohybridization probes. We screened the sSNPs on both Chr 2 and 3 for polymorphism in Southeast Asia using a panel of 25 samples (five each from Bangladesh, Thailand, Laos, Vietnam, and Myanmar). This was done by restriction digestion of PCR products using appropriate restriction enzymes and running the digested products on agarose gels. In cases where no restriction sites were available, sites were manufactured by incorporating a mismatch of 3–4 bp from the 3′ end of an oligonucleotide. SNPs that were polymorphic in at least one population were included in the study. These SNPs were then genotyped in all 755 samples by restriction digestion of fluorescent-labeled PCR products. Primer sequences, amplification conditions, restriction enzymes, and positions of SNPs are described in table 1. We also provide details of additional SNPs on Chr 2 and 3 that were not polymorphic in the panel of 25 samples and were therefore not included in this study (Supplementary Tables S1 and S2, Supplementary Material online).

Measurement of Allele Frequencies

Accurate estimation of allele frequencies is critical for assessing population structure. Measurement of allele frequencies in microparasite populations is complicated by the presence of multiple clones within many infections (Hill and Babiker 1995; Rannala, Qiu, and Dykhuizen 2000). Counting all alleles identified within an infection results in overestimation of frequencies of rare alleles and underestimation of common alleles. To minimize bias, we used three different methods to estimate allele frequencies: (1) we excluded infections in which >1 allele was observed. Infections were deemed as containing multiple clones if peaks at the minor allele were $>10\%$ the height of those for the predominant allele on the ABI electropherograms. This approach is unbiased but results in considerable loss of data. (2) We picked the predominant allele present in each infection. Using this approach, a single genotype is obtained from each infection scored. Predominant alleles were identified by comparing the relative height of peaks on electropherograms. (3) We used a maximum likelihood procedure (Hill and Babiker 1995) to estimate both the allele frequency and the mean number of clones per infection. For each locus, we counted the numbers of infections containing allele 1, allele 2, or alleles 1 and 2. We assumed a positive Poisson distribution of parasite clones per host and used maximum likelihood to obtain the best-fit values for frequencies of alleles 1 and 2 and mean infection rate. A positive Poisson distribution was used because we sampled only *P. falciparum*–positive patients, so we had no hosts carrying zero parasite alleles.

Geographical Differentiation

We assessed genetic differentiation at each of the selected and putatively neutral loci using allele frequencies estimated using the three procedures described. We calculated F_{ST} in the total population sample using FSTAT (Goudet 2001), and confidence errors for each locus were derived by jackknifing over populations to determine if values differed significantly from zero. To identify which sampling locations contribute most to the differentiation observed, we also calculated F_{ST} between all pairwise combinations of locations.

We used two approaches to compare population differentiation in selected and putatively neutral SNPs. First, we simply compared F_{ST} values in the two categories of loci using nonparametric tests (Barbujani 1985). This approach indicates whether the two groups of loci differ but does not demonstrate that values for individual SNPs differ from neutral expectations. To do this, we used the simulation approach outlined in Beaumont and Nichols (1996). We plotted weighted F_{ST} values against heterozygosity for each locus examined. We then used the mean weighted F_{ST} values from our sample of putatively neutral sSNPs,

Table 1
Locations and Assay Methods for sSNPs on Chr 2 and 3

Chr	Position	Gene	Change	Description	Uncut Product (bp)	Restriction Enzymes	Label	Notes	Oligos
2	619397	PFB0685c	G-A	Acyl-CoA synthetase	220	*Apo*1	6FAM	*Made RE site	TTCATACACATACAATATATACG TATATGAGTTTCTTATCTCCATTA TTACAATTTTTTTATTGTGTaAAT
2	649509	PFB0715w	C-T	DNA-directed RNA polymerase II second largest subunit	150	*Bsm*I	HEX	*Made RE site	TGATAAATTCGCTAGTAGACA CTGCAACTTTTCCTGTTA CAAAAGGGTACTATTGGTATT
3	197185	PFC0180c	G-T	Hypothetical protein	138	*Xmn*I	NED		AAGAAAATGAAACTACAGCA CTTGTTGTACTACAGGAGGA GTATAAACGAAACAAACGAT
3	221432	PFC0215c	C-G	Hypothetical protein	176	*Mnl*I	6FAM		ATTTAAAGAGGAAAATGACC AAATTAATGCGTCTTCATAG TACTATTGGATTTTCAAGGA
3	309640	PF0295	T-G	40S ribosomal protein S12	104	*Bst*Y1	NED	*Made RE site	CCAATGTCCTAAATCTTTACT AGTTATAAAATCCATCGgATC TTTCAGCACATAAGGTAGTAA
3	360148	PFC0350c	G-C	Putative T-complex protein eta subunit	207	*Aci*I	NED		AATTATTGAATGAAGCAAAG AACTAATTTTGAATTTAAgGC TTAAATGATGGTATTGAACC
3	454628	PFC0440c	T-A	Putative helicase	152	*Dra*I	HEX		TGCAGAACAACATGATAATA ATTATTATGCATACCAGGAC TGAAGCATAGTACTAAAGGTG
3	526792	PFC0530w	G-A	Hypothetical protein	175	*Dra*I	HEX		CCCAAATACATTATTTTCAT GTAATGGACATGGGAAAG ACATAAAATAACATACACTAAAGG
3	679987	PFC0745c	G-A	Proteasome component C8	143	*Nsi*I	HEX		AACATATTCAGGATTTGATG TCATGTGCCAATATAATGTA TGGAGATGCAAGAAATATAA
3	888282	PFC0940c	G-A	Hypothetical protein	122	*Hin*FI	6FAM		GAGCATATGTTTTGATGATT GAGCATATGTTTTGATGATT GTTGGTTGTTTATGTCTTGT

NOTE.—All primers are listed 5′ → 3′. For each locus, the first and second primers are used in the primary reaction and the second and third primers are used in the secondary reaction. Primers listed third are end labeled with fluorescent tags. The primary reaction (25 μl) contained 2.5 μl of template, 2.5 μl of 10× buffer, 0.75 units of *Taq* polymerase (Takara, Otsu, Shiga, Japan), 3 mM Mg^{2+}, and 0.1 pM of each primer. The secondary reaction (10 μl) contained 1 μl of template, 1 μl of 10× buffer, 0.35 units of *Taq* polymerase (Takara), 2.5 mM Mg^{2+}, and 0.4 pM of each primer. Cycling conditions for both primary and secondary rounds were as follows: 2 min, 94°C; (30 s, 94°C; 30 s, 50°C; 30 s, 60°C) × 5 cycles, followed by (30 s, 94°C; 30 s, 45°C; 30 s, 60°C) × 25 cycles and an extension of 60°C for 2 min. Second round products were digested overnight using 0.8–1 unit of appropriate restriction enzyme. In some cases (marked by *), restriction enzyme sites were created by making a mismatch in 3–4 bp from the 3′ end of the unlabeled oligo in the second reaction. The digestion products were pooled together in the following combinations for capillary electrophoresis: (PFB0685c, PFB0715w, PFCO350); (PFC0180C, PFC0215C, PFC0745C); and (PFC0440C, PFC0530W, PFC0940C, PFC0295). We also designed assays for 11 additional sSNPs on Chr 3 and 6 additional sSNPs on Chr 2. These were not polymorphic in initial inspection of 25 Southeast Asian isolates (Supplementary Table S1, Supplementary Material online).

together with sample size and population number information to seed coalescent simulations of genetic differentiation within metapopulations. Initially, we used an island model with 100 islands and repeated the procedure 50,000 times to generate 99% confidence intervals for neutral differentiation and estimated *P* values for departure of selected loci from these expectations. This analysis was conducted using the FDIST2 program (http://www.rubic.rdg.ac.uk/~mab/software.html). This approach is fairly robust to variation in mutation rate between loci, to sample size, and whether populations are in equilibrium or nonequilibrium (Beaumont and Nichols 1996). However, reducing the number of populations used in the simulations can influence the distribution of F_{ST}, depressing it at high heterozygosities (Flint et al. 1999). We do not have a good idea of the appropriate number of islands in the Southeast Asian *Plasmodium* population. We therefore reran simulations using 15 populations to examine the robustness of our conclusions to variation in this parameter.

Isolation by Distance

We investigated the fit of the data from each sSNP to isolation-by-distance models using Mantel tests. For these analyses, we measured distances between locations from maps and used the natural log of distance. We used the $F_{ST}/(1 - F_{ST})$ as our measure of genetic similarity (Rousset 1997). The significance of Mantel correlations was assessed by random permutation of the matrices 10,000 times and recalculation of test statistics. Tests were significant if the observed test statistics were within the top 5% of the distribution of test statistics measured in the permuted data sets. These analyses were performed using permutation tests implemented using POPTOOLS (http://www.cse.csiro.au/poptools/).

Results

Allele-Frequency Estimation

We analyzed a total of 755 malaria infections for 20 point mutations (10 nsSNPs and 10 sSNPs, with sample sizes ranging from 21 [Meuang Feuang, Laos] to 107 [Pujgang, Bangladesh]). The three different procedures used to estimate allele frequencies gave extremely similar values—these estimates are listed for each population in the Supplementary Table S3 (Supplementary Material online). The maximum difference between estimators was found in Ratchaburi (Thailand) where the predominant allele measure was 13% higher than both the maximum likelihood and single-allele estimates for the *dhfr-51*. We measured correlations between the 260 allele-frequency estimates (20 loci × 13 populations) derived using the three methods. For each comparison, the slopes were indistinguishable from 1, indicating minimal bias, and the correlation coefficients were >0.995 (fig. 2). We used the predominant allele method for most analyses reported here because it utilizes data from all people sampled and makes few assumptions. However, for the analysis of *dhfr* and *dhps*, where alleles are defined by multiple SNPs, we used the single-allele method because this allows unambiguous construction of haplotypes.

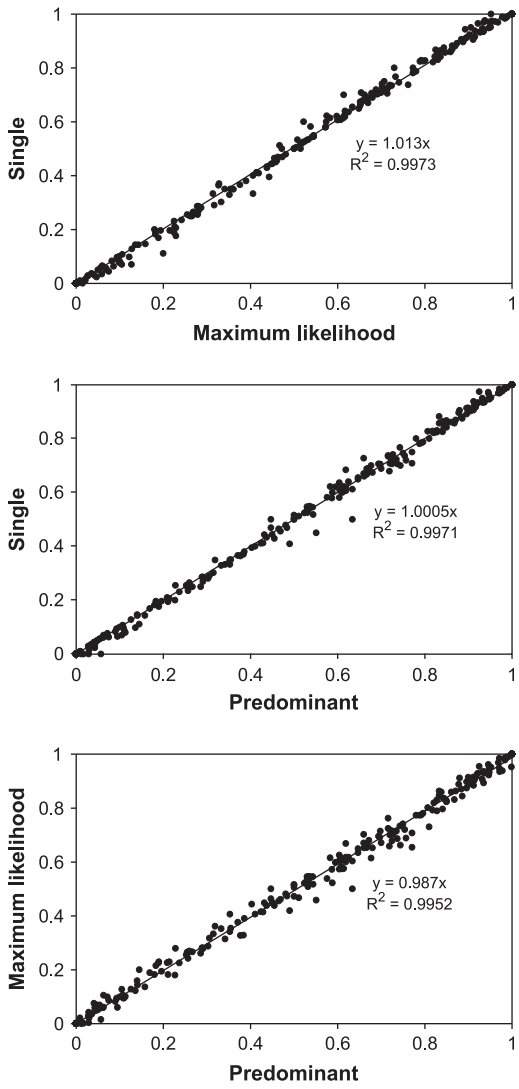

FIG. 2.—Measurement of allele frequencies in *Plasmodium falciparum* populations. Three different approaches were used to estimate allele frequencies (see text for details). The three graphs show comparisons between these methods. Allele frequencies estimated for all 20 SNPs for each of the 13 populations examined (260 points) are plotted on each graph. The lines are linear regressions and the equations and R^2 value are shown. All three measures (predominant, single, and maximum likelihood) show strong correlation ($R^2 \geq 0.995$) and slopes are close to 1.

nsSNPs in Resistance Genes

Figure 3 summarizes allele frequencies observed at *dhfr*, *dhps*, *pfcrt*, and *pfmdr*. (1) *pfcrt*: The K76T at *pfcrt* critical for CQ resistance is found at frequencies of 0.58–1.00 in the populations examined. Eight of 13 populations examined had >95% frequency of this allele, and all four populations sampled from the Thailand-Burma border were fixed for this allele. Only three populations in Burma, Laos, and Vietnam showed high frequencies (>30%) of the wild-type allele. (2) *pfmdr*: The N86Y mutation in *pfmdr* is at low frequency in Thailand, Laos, Cambodia, and Vietnam (<30%) but at high frequency in one population from Western Burma (73%) and Bangladesh (61%–79%). (3) *dhfr*: Data from 10 of the 13 populations have been presented previously (Nair et al. 2003). We found between

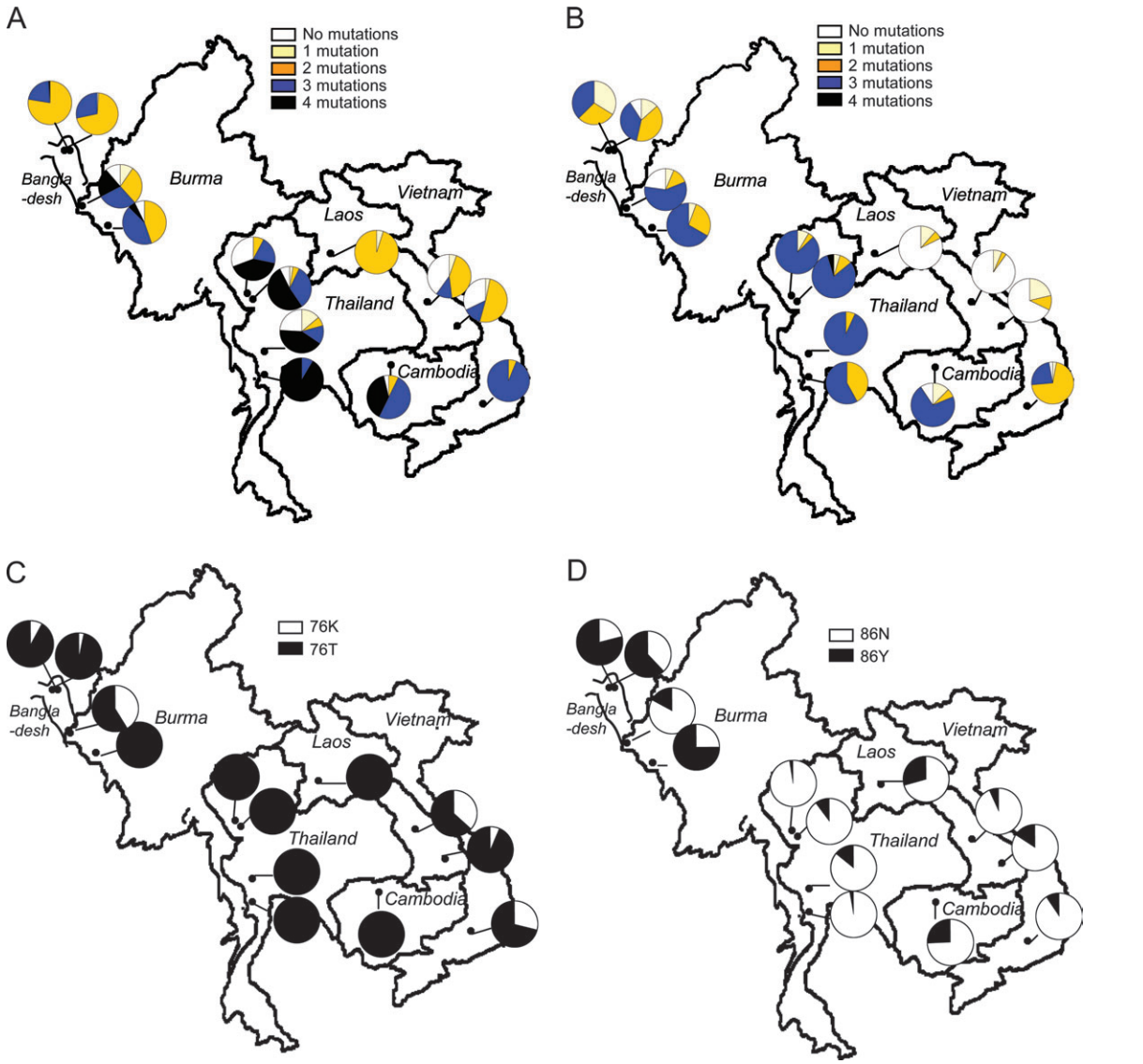

FIG. 3.—Maps showing allele frequencies at four drug resistance loci in Southeast Asian parasite populations. (*A*) *dhfr*, (*B*) *dhps*, (*C*) *pfcrt*, and (*D*) *pfmdr*. From north to south, the Thai locations are Shoklo, Maela, Ratchaburi, and Kanchanaburi; the Laos locations are Muang Feuang (central), Phalanxay (southwest), and Sekong (southeast); the Myanmar locations are Dabhine (upper) and Myothugyi (lower); and the Bangladeshi locations are Dighinala and Panchari. Only the Chittagong Hill Tracts region of Bangladesh is shown. Frequencies of alleles present shown are estimated from single-clone infections only. These and other estimates are provided in Supplementary Table 3 (Supplementary Material online), and statistics describing population differentiation are shown in table 2 and figure 4.

zero and four mutations in *dhfr* alleles. The highest level of resistance alleles was found in Thai populations, where most parasites contain two to four mutations and the I164L mutation conferring high-level resistance to pyrimethamine is common (67%–100%). (4) *dhps*: We found between zero and four mutations in *dhps* alleles. The highest level of resistance was found in Thailand, where most parasites carry three mutations and no wild-type parasites were found, while the lowest levels of resistance was found in neighboring Laos, where >66%–85% of parasites carried wild-type (sensitive) alleles.

sSNPs in Putatively Neutral Loci

Of the 62 sSNPs identified on Chr 3 by Mu et al. (2002), we screened 19 for variation in Southeast Asian samples. Of these, eight loci were variable (alternative alleles seen in at least one parasite) in the 25 Southeast Asian parasites screened. These were included in the full population survey. Of the 29 regions sequenced on Chr 2, in which 25mer Affymetrix probes suggested presence of SNPs (Volkman et al. 2002), 25 gave good sequence. Eighteen of 25 regions contained variable sites, including 14 sSNPs and 15 nsSNPs (Supplementary Table 3, Supplementary Material online). Of these, only two of eight sSNPs examined were polymorphic in the 25 Southeast samples screened. Six additional sSNPs on Chr 2 were not genotyped because restriction digest assays could not be designed or they were only variable in the 3D7 sequence. We genotyped all 10 sSNPs (eight on Chr 3 and two on Chr 2) in 755 infections. One of these loci (PFB0685c) had rare

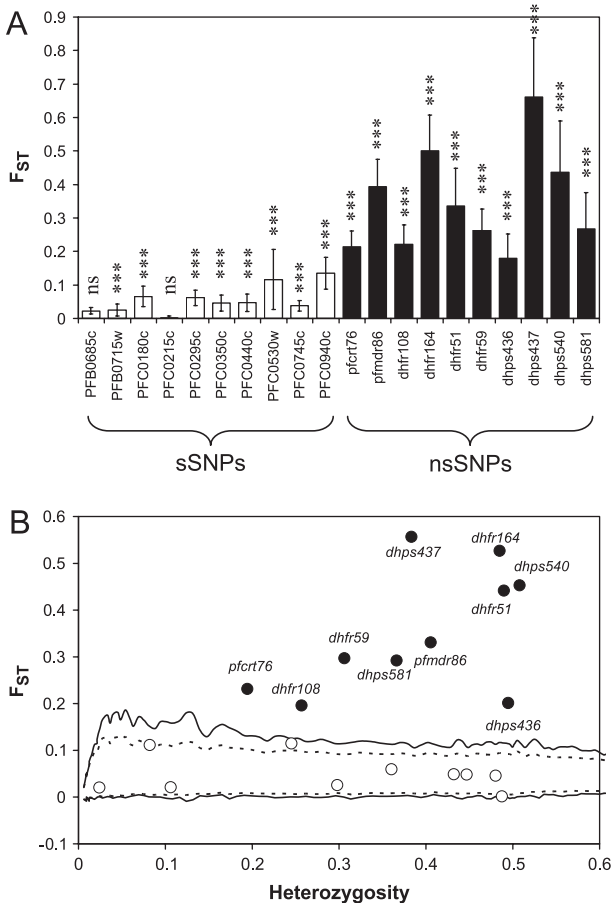

FIG. 4.—Differences in population structure between selected and putatively neutral SNPs. (*A*) Histogram showing F_{ST} (±SE). Black bars indicate SNPs in resistance genes and unfilled bars indicate SNPs in housekeeping genes or genes of unknown function. Asterisks indicate significant differences from random distributions as assessed by permutation tests. (*B*) Estimated F_{ST} values for 10 nsSNPs in four drug resistance genes (●) and 10 synonymous changes in housekeeping genes or genes of unknown function (○) plotted against heterozygosity. The dotted lines indicate 95% confidence intervals while the solid lines indicate 99% confidence intervals for neutral expectations. These confidence intervals are derived by simulation using the mean values in the 10 putatively neutral loci as starting values and parameters values (15 subpopulations containing 50 individuals) and methods described in the text. The *P* values for deviation from neutral expectations for the 10 drug resistance polymorphisms are shown in table 2.

minor alleles with maximal allele frequencies of 6% in the 13 populations. Eight of the 10 sSNPs showed weak but significant differentiation between populations (fig. 4) with average F_{ST} of 0.029 (range 0–0.14). Much of the differentiation is explained by populations from Bangladesh and Laos. Parasites from these countries are significantly differentiated from those collected from other countries in the region (fig. 5).

Comparison of Selected and Putatively Neutral SNPs

F_{ST} values for all nsSNPs in resistance genes were higher than those for the sSNPs examined (fig. 4). The differences between these two groups of SNPs are highly significant (Mann-Whitney test: $Z = -3.78$, $P < 0.0002$). The four mutations within both *dhfr* and *dhps* are not independent and may elevate the significance of the differences observed. To avoid this, we reanalyzed these data using the mean F_{ST} over all four polymorphic sites within both *dhfr* and *dhps*, but the result is still highly significant ($Z = -2.83$, $P < 0.0047$). To evaluate whether individual SNPs show patterns of genetic differentiation suggestive of selection, we used coalescent simulations to generate the distribution of expected values for neutral loci using FDIST2 (Beaumont and Nichols 1996). To seed these simulations, we used the mean weighted F_{ST} values for the 11 sSNPs typed and 15 or 100 populations. In both cases, all 10 nsSNPs within resistance genes show strong deviations from neutral expectations with *P* values ranging from 8×10^{-5} to $<10^{-6}$ (table 2). The *P* values for these deviations were derived by comparing observed results with the distributions derived from 50,000 coalescent simulations. To investigate the robustness of this result, we also ran simulations seeded with an F_{ST} value 2 standard deviations (SDs) above that estimated from the 10 sSNPs. Even in this case, we found that 9 of 10 (15 populations) and 8 of 10 (100 populations) SNPs still showed significant ($P < 0.05$) deviations from simulated expectations (table 2).

The analysis above compares the extent of geographical differentiation revealed by the two groups of SNPs. We might also expect differences in the pattern of differentiation, with nsSNPs in resistance loci showing poorer fit to isolation by distance than putatively neutral sSNPs. To examine this, we conducted Mantel tests for each of the 20 SNPs and for nsSNPs and sSNPs combined (table 3). We observed significant Mantel tests for both categories of SNPs (sSNPs: correlation coefficient = 0.43, $P < 0.0001$; nsSNPs: correlation coefficient = 0.302, $P = 0.012$). Mantel tests for individual loci demonstrated that among the nsSNPs, only *pfmdr-86* and *dhps-437* were significantly influenced by distance, while for the sSNPs, four of 10 loci showed significant isolation by distance (table 3). A comparison of correlation coefficients for nsSNPs and sSNPs revealed no significant difference between the two groups (Mann-Whitney *U* test, $Z = -0.282$, not significant).

Discussion

Efficacy of F_{ST} for Detecting Selection

We found dramatic differences in geographical structure between nsSNPs in drug resistance genes and sSNPs in putatively neutral loci. While this result was expected, the strength and consistency of the differences observed were not. All 10 nsSNPs within the four resistance genes show higher F_{ST} than the 10 sSNPs, and simulations suggest that the F_{ST} values observed deviate significantly from neutral expectations with *P* values for all comparisons below 8×10^{-5}. Eight of 10 nsSNPs remain as outliers even when we use an F_{ST} value 2 SDs greater than that observed to seed the coalescent simulations. This suggests that any inaccuracy in our measure of F_{ST} due to sampling of putatively neutral sSNPs is unlikely to explain the patterns observed. These empirical results provide strong support for recent simulation studies that demonstrated efficient detection of loci under directional selection, with few false-negative

A

	BAN_DI	BAN_PA	BUR_DA	BUR_MY	CAM_AN	LAO_MF	LAO_PH	LAO_SE	THA_KA	THA_MA	THA_RA	THA_SH	VIE_KH
BAN_DI	0.000	-0.001	0.088	0.025	0.098	0.090	0.061	0.113	0.059	0.073	0.084	0.076	0.091
BAN_PA		0.000	0.094	0.032	0.106	0.081	0.061	0.114	0.068	0.092	0.088	0.092	0.103
BUR_DA			0.000	0.026	0.038	0.135	0.040	0.072	0.014	0.034	0.043	0.046	0.019
BUR_MY				0.000	0.031	0.075	0.024	0.052	0.019	-0.001	0.009	0.009	0.030
CAM_AN					0.000	0.078	0.036	0.020	0.008	0.035	-0.007	0.011	-0.002
LAO_MF						0.000	0.030	0.028	0.049	0.094	0.076	0.084	0.062
LAO_PH							0.000	0.005	0.004	0.038	0.027	0.039	0.001
LAO_SE								0.000	0.016	0.048	0.020	0.038	-0.001
THA_KA									0.000	0.033	0.026	0.014	-0.008
THA_MA										0.000	0.011	0.011	0.024
THA_RA											0.000	0.014	0.004
THA_SH												0.000	0.023
VIE_KH													0.000

B

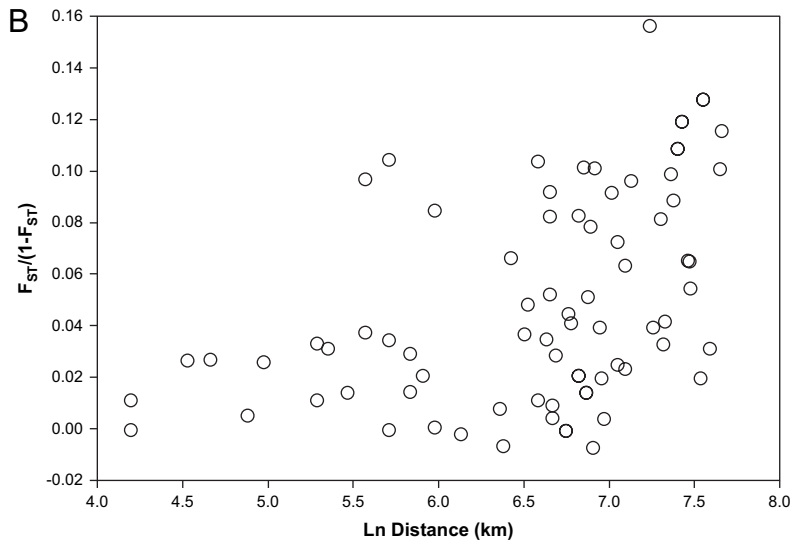

FIG. 5.—Patterns of population differentiation revealed by sSNPs in Southeast Asia. (*A*) Matrix of F_{ST} values showing pairwise comparisons between the 13 populations for 10 sSNPs. Shading indicates degree of significance: black shading and white text, $P < 0.01$; gray shading and black text, $P < 0.05$. Parasite population samples from Bangladesh and Laos are significantly differentiated from other countries in the region. (*B*) Relationship between differentiation and geographical distance. The Mantel correlations show significant evidence for isolation by distance (permutation test, $P < 0.0001$).

or false-positive results when selection coefficients were >0.1 (Beaumont and Balding 2004). We also observed that the combined sSNPs fitted a model of isolation by distance better than the combined nsSNPs. However, many individual loci from both groups showed poor fit to isolation-by-distance models, and differences between the nsSNPs and sSNPs were not significant. Hence, analyses of patterns of differentiation (isolation by distance) were less useful for discriminating selected loci than analyses of the extent of differentiation (measured using F_{ST}) in this system.

Two features of the drug resistance may contribute to the distinctive differences in the extent of population structure revealed by nsSNPs in resistance loci and putatively neutral sSNPs. First, selection is strong. Selection coefficients driving antimalarial drug resistance have been estimated at between 0.03 and 0.3, based on allele-frequency change over time (Nair et al. 2003; Roper et al. 2003; Anderson 2004). Selection coefficients are expected to be particularly high in many Southeast Asian countries, where most infections are symptomatic and a high proportion of infections is exposed to drug treatment (Luxemburger et al. 1996, 1997). However, while the selection coefficients driving differentiation in allele frequencies at drug resistance loci are strong, they are by no means atypical of selection coefficients measured in many natural populations (Endler 1986; Hoekstra et al. 2001). Second, drug resistance has evolved very recently in malaria parasites. The clarity of the differences in population structure observed in the two groups of SNPs may in part reflect the fact that drug resistance mutations are not in migration-drift equilibrium, and allele frequencies have not yet been homogenized by migration. These data provide strong empirical support for the idea that genome regions containing selected loci could be located by comparing F_{ST} values (Lewontin and Krakauer 1973).

In this study, we examined 13 parasite populations. Two of the drug resistance polymorphisms studied (*pfmdr-86* and *pfcrt-76*) show homogeneous distribution in the majority of populations examined and their high F_{ST} values result from distinctive allele frequencies in just a few of the populations. In the case of *pfcrt*, for example, resistance allele frequencies are >95% in eight of the populations examined, and wild-type alleles are found at frequencies >30% in only three parasite populations. It would be prohibitively expensive to genotype parasites from all 13 populations in a genome-wide F_{ST} screen.

Table 2
Significance of Deviations from Neutral Expectations in Drug Resistance Loci

			P Values			
			Mean F_{ST}		Mean F_{ST} + 2 SD	
Locus	H_e	F_{ST}	15pops	100pops	15pops	100pops
pfcrt	0.194	0.231	**3.1×10^{-5}**	**3×10^{-6}**	**0.044**	0.053
dhf108	0.257	0.196	**4.1×10^{-5}**	**7.9×10^{-5}**	0.091	0.106
dhf164	0.485	0.527	**$<10^{-6}$**	**$<10^{-6}$**	**$<10^{-6}$**	**$<10^{-6}$**
dhf51	0.490	0.442	**$<10^{-6}$**	**$<10^{-6}$**	**$<10^{-6}$**	**$<10^{-6}$**
dhf59	0.306	0.297	**$<10^{-6}$**	**2×10^{-6}**	**0.003**	**0.002**
dhp436	0.495	0.201	**10^{-5}**	**8×10^{-5}**	**0.038**	**0.042**
dhp437	0.383	0.557	**$<10^{-6}$**	**$<10^{-6}$**	**$<10^{-6}$**	**10^{-6}**
dhp540	0.508	0.453	**$<10^{-6}$**	**$<10^{-6}$**	**$<10^{-6}$**	**2×10^{-6}**
dhp581	0.366	0.292	**$<10^{-6}$**	**$<10^{-6}$**	**0.002**	**0.004**
pfmdr	0.406	0.331	**$<10^{-6}$**	**$<10^{-6}$**	**0.001**	**0.001**

NOTE.—The table shows the weighted mean heterozygosity and F_{ST} and *P* values for deviation from neutral expectations for each of the nsSNPs within drug resistance loci. The coalescent simulations used to derive these *P* values were based on average F_{ST} values observed for the 10 sSNPs, sample sizes of 50, and either 15 or 100 subpopulations within a metapopulation. We also simulated expectations using F_{ST} values 2 SD greater than that observed in our sample of sSNPs to investigate the robustness of these results. Significant ($P < 0.05$) results are shown in bold text.

For a genome-wide comparison of F_{ST}, examination of smaller numbers of populations for which we suspect differences in selection regimes may be more practical. Comparisons between neighboring countries of Laos and Thailand may be particularly informative because these neighboring countries show the most marked differences in allele frequencies at loci encoding resistance. In Southern Laos, allele frequencies in drug resistance genes are consistently lower than in Thailand for all four loci, suggesting strong local adaptation in these neighboring countries. We also have independent evidence that drug selection is weaker in Laos than in Thailand because hitchhiking events associated with resistance to both pyrimethamine and CQ affect larger genome regions in Thailand than in Laos (Nash et al. 2005).

Some caveats are important to mention. F_{ST}-based approaches for detecting selection rely strongly on the assumption that selected sites (or SNPs in LD with selected sites) are rare relative to neutrally evolving sites and therefore show up as outliers. This may be broadly true in many free-living organisms. However, it is questionable whether this is the case for *P. falciparum*. Recent studies suggest that a high proportion of polymorphisms are under balancing selection in this organism as a result of immune selection. An oligo microarray study on Chr 2 of *Plasmodium* showed that 69% of all polymorphisms detected were in membrane-associated genes in telomeric and subtelomeric regions and that polymorphism is rather rare elsewhere in the genome (Volkman et al. 2002). The highly polymorphic genes included *Var* genes involved in antigenic variation (Bull et al. 1998; Newbold 1999), vaccine targets such as pfs230 (Bustamante et al. 2000; Williamson 2003), and merozoite antigens (M. K. Hughes and A. L. Hughes 1995) for which there is independent evidence of positive selection from sequence comparisons. Furthermore, many other polymorphisms were found in unknown genes that may also be surface expressed and exposed to immune selection (Volkman et al. 2002). Random sampling of point mutations from the *P. falciparum* genome using a technique such as amplified fragment length polymorphism (AFLP), as used in recent studies of free-living organisms (Wilding, Butlin, and Grahame 2001; Campbell and Bernatchez 2004), would tend to sample subtelomeric SNPs that are predominantly from antigen genes under balancing selection. This would result in underestimates of F_{ST} for randomly sampled SNPs presumed to represent neutrality and a consequent overestimation of the numbers of outlier SNPs that are identified as being under selection. We expect that this problem will apply to other parasites and pathogens in which many proteins are under balancing or frequency-dependent selection due to immune pressure.

The opposite bias—overestimation of F_{ST} for randomly sampled SNPs presumed to represent neutrality—could also generate bias in studies of free-living organisms if a substantial proportion of loci are locally adapted. In this case, this would lead to underestimation of the numbers of loci inferred to be under positive selection. For example, a large survey of human SNP variation showed higher F_{ST} for nsSNPs than sSNPs, suggesting that many amino acid changes may be locally adapted (Hinds et al. 2005).

Table 3
Fit to an Isolation-by-Distance Model for Selected and Putatively Neutral SNPs

Selected SNPs			Putatively Neutral SNPs		
Locus	Mantel	*P*	Locus	Mantel	*P*
pfcrt-76	−0.063	0.758	PFB0685c	−0.026	0.670
pfmdr-86	0.312	**0.009**	PFB0715w	0.137	0.081
dhfr-51	0.069	0.228	PFC0180c	0.124	0.105
dhfr-59	−0.016	0.520	PFC0215c	−0.046	0.636
dhfr-108	0.097	0.145	PFC0295c	0.411	**0.003**
dhfr-164	0.079	0.243	PFC0350c	0.229	**0.022**
dhps-436	−0.006	0.465	PFC0440c	0.286	**0.007**
dhps-437	0.174	**0.039**	PFC0530w	0.067	0.210
dhps-540	0.175	0.055	PFC0745c	0.097	0.163
dhps-581	0.148	0.079	PFC0940c	0.211	**0.039**
All nsSNPs	0.302	**0.012**	All sSNPs	0.436	**<0.0001**

NOTE.—Pearson's correlation coefficients and *P* values for Mantel tests comparing matrices of genetic and geographical distance for both individual loci and from nsSNPs and sSNPs. Significance of test statistics was assessed by permutation and values with $P < 0.05$ are shown in bold.

There are two alternative approaches to random sampling of SNPs that could be used for organisms such as *P. falciparum* in which balancing selection is expected to be common. First, we could measure neutral variation by deliberately choosing synonymous SNPs within internally expressed housekeeping genes to try to minimize the chance of picking SNPs under balancing selection. These data could be used as a basis for simulating neutral expectations. We have used this approach here but found relatively few sSNPs that satisfied our criteria for putatively neutral loci or were polymorphic in the samples examined. For example, of the 62 sSNPs identified on Chr 3 by Mu et al. (2002), only 19 were in housekeeping genes or genes of unknown function and only 8 were polymorphic in Southeast Asia. This is unlikely to be a problem in future studies. Genome-wide SNP maps are currently being constructed using oligo hybridization or direct sequencing, so many putatively neutral SNPs will be available for future comparisons of this kind (E. Winzeler and J. Mu, personal communication). Second, we can use positional information and search for peaks and troughs in F_{ST} throughout the genome. This may be particularly useful when used in conjunction with other measures of selection. Previous studies have demonstrated the value of "within-population" analyses of diversity and LD for detecting selective sweeps associated with genes under strong positive selection in *P. falciparum* (Wootton et al. 2002; Nair et al. 2003; Roper et al. 2003; Anderson 2004; Nash et al. 2005). As predicted by theory, genome regions flanking resistance mutations show reduced diversity and increased linkage disequilibrium relative to regions flanking wild-type alleles, as well as skewed allele frequencies with an excess of deficit of alleles at intermediate frequency (Anderson 2004). Elevated F_{ST} provides an additional parameter that can be measured using data from multiple populations. Given the problem with multiple statistical tests in genome-wide association studies, use of both within- and between-population measures of selection will strengthen arguments for prioritizing particular genome regions for further investigation.

Our results are similar to those of Abdel-Muhsin et al. (2003), who examined allele frequencies at *pfcrt*, *pfmdr*, *dhfr* and three microsatellite loci in parasites collected from Khartoum (Sudan) and two villages 420 miles away. They found greater geographical differentiation at SNPs within loci encoding resistance than in microsatellite loci. Interestingly, they also found greater variation in frequency between wet (high transmission) and dry (low transmission) seasons. Unfortunately, interpretation of these results is complicated by the fact that the markers compared (SNPs and microsatellites) have different mutation patterns. Microsatellite loci evolving under stepwise mutation are expected to show reduced F_{ST} relative to SNPs as a result of homoplasy when heterozygosity is high (Flint et al. 1999). Hence, it is difficult to distinguish the impact of selection on population structure from differences due to mutation processes. Similarly, Conway et al. (2001) describe high levels of geographical differentiation of nsSNPs within a male gamete surface protein (pfs48/45) relative to that seen in microsatellite markers. However, while these results are suggestive of positive selection, these authors cautioned that differences in the mutational processes make interpretation problematic. We note that mitochondrial DNA (mtDNA) SNPs show similarly high F_{ST} values ($F_{ST} = 0.65$ in intercontinental comparisons) to those seen for pfs48/45 (Conway et al. 2000*b*; Joy et al. 2003). This suggests that high levels of differentiation are not atypical when comparing global SNP distributions in *P. falciparum* and that homoplasy reduces F_{ST} for microsatellite loci.

F_{ST} approaches have also been used to investigate antigen-encoding genes and gene regions under balancing or frequency-dependent immune selection. These studies have focused on SNPs showing lower than expected F_{ST} rather than elevated F_{ST} (Conway et al. 2000*a*; Conway and Polley 2002). For example, examination of SNPs across the apical membrane antigen-1 gene, a leading vaccine candidate, in Nigeria and Thailand, revealed runs of SNPs showing low F_{ST} values in domains I and III (Polley, Chokejindachai, and Conway 2003). In comparison, F_{ST} was higher in domain II and in 11 unlinked microsatellite loci elsewhere in the genome. We note that comparisons of SNPs with microsatellites used in these studies may result in overly conservative tests for balancing selection because microsatellites tend to show reduced F_{ST} relative to SNPs (Flint et al. 1999). Putatively neutral SNPs, as used here, are expected to provide more reliable comparators for genes under balancing or frequency-dependent selection.

Neutral Population Structure

We observed weak but significant population structure for 8 of 10 loci examined. We have previously examined population structure on a global scale using microsatellite markers (Anderson et al. 2000). This study revealed a spectrum of population structures with strong population structure in South America ($F_{ST} = \sim 0.4$) and minimal differentiation between African populations. The putatively neutral sSNPs examined in this study reveal low ($F_{ST} =$ 0–0.14) but significant genetic structure for 8 of 10 loci and a good fit to isolation by distance ($P < 0.0001$). The level of population structure is intermediate between that observed in South America and Africa, consistent with our previous data (Anderson et al. 2000). The weak but significant population structure observed in the sSNPs cautions against combining parasite isolates collected from different regions of Southeast Asia for use in association analyses or candidate gene studies because population structure can generate spurious associations (Pritchard et al. 2000; Pritchard and Donnelly 2001). Contrasting patterns of isolation by distance have recently been documented in malaria parasite populations from Brazil (no evidence for isolation by distance) (Machado et al. 2004) and Indonesia (strong isolation by distance) (Anthony et al. 2005). The Southeast Asian populations examined here show weaker isolation by distance than seen in the Indonesian study.

The low F_{ST} results suggest extensive gene flow between Southeast Asia parasite populations. Gene flow of parasites may result from movement of both human and mosquito hosts. Direct measures of human traffic reveal that ~20 million people per year travel across international borders between China, Burma, Laos, and Vietnam (Hu et al. 1998). Actual numbers are probably much higher,

if we include people moving illegally between countries. We have no direct measures of mosquito migration, but indirect estimates can be assessed using population genetic approaches. For example, mtDNA and microsatellite analyses of *Anopheles dirus* species complex suggest little structure within each of the vector species A and D within Thailand (Walton et al. 2001). However, while these results suggest high levels of movement between populations, they could also be explained by rapid population expansion (Walton et al. 2000).

Distribution of Drug Resistance Alleles and Patterns of Drug Selection

Heterogeneity in allele frequencies among locations may result from both selection for resistance alleles in the presence of drug selection and/or selection against resistance alleles in the absence of drug selection. The highest frequencies of resistance alleles in *dhfr*, *dhps*, and *pfcrt* are seen in populations on the Thailand-Burma border. This is consistent with the fact that malaria is restricted to a few border regions of Thailand, where treatment regimes are carefully optimized and treatment coverage is good. In comparison, in neighboring countries such as Laos and Burma, health infrastructure is weaker, malaria is more widespread, and treatment regimes are frequently suboptimal utilizing CQ and SP which are relatively ineffective. Neither CQ nor SP has been used for treating *P. falciparum* in Thailand for 25 years, although CQ is still used widely for treatment of *Plasmodium vivax* infections and trimethoprim-sulfamethoxazole (an antifol-sulfa combination with activity similar to SP) is still widely used as an antibiotic throughout the region. The high frequency of mutations in *pfcrt*, *dhfr*, and *dhps* is surprising because resistance alleles at two of these loci (*pfcrt* and *dhfr*) carry fitness costs. In the case of *pfcrt*, evidence for fitness costs comes from three sources. First, removal of CQ pressure results in decline in resistance allele frequencies in Malawi (Kublin et al. 2003), Hainan Island (Wang et al. 2005), and Vietnam (T. T. Hien, unpublished data). Second, parasites carrying back mutations in residue 76 have been found in nature (Fidock et al. 2000). Finally, laboratory selection of mutations in residue 76 results in parasites with reduced growth rates (R. Cooper, personal communication). Similarly, for *dhfr*, enzyme manufactured from mutant alleles shows weaker binding to substrate than wild-type enzyme (Sirawaraporn et al. 1997), and yeast transfected with resistant alleles show lower growth rates than parasites transfected with wild-type alleles (Cortese and Plowe 1998).

Selection pressures acting on *pfmdr* are thought to be particularly complex, with both point mutations (Reed et al. 2000; Djimde et al. 2001) and copy number changes playing important roles (Price et al. 2004). Furthermore, while CQ selects for mutant alleles and low copy number, MFQ, QN, and artemisinin derivatives act in the opposite direction, selecting for wild-type alleles and copy number amplification. We genotyped the N86Y mutation in *pfmdr*. This mutation plays an important role in CQ drug resistance in African countries (Hayton and Su 2004), although its role in Southeast Asia is less well supported. We see low levels of the 86Y in all locations, except for Bangladesh and one Burmese population. The differentiation that we see at N86Y may not reflect direct selection on this SNP. Instead, differentiation may result from LD between this SNP and copy number amplification. Recent data suggest that *pfmdr* copy number plays the predominant role in determining resistance to MFQ in Southeast Asia (Price et al. 2004). Consistent with this, real-time assays show that 34% of samples from the Thailand-Burma border carry >1 copy of *pfmdr*, while in neighboring Laos all 80 infections examined carried a single copy (Anderson et al., unpublished data).

The geographical distribution of resistance alleles has important practical implications for resistance management. Mapping of resistance alleles provides a powerful approach to monitoring spread of resistance and allows rational choices to be made about treatment policy. The efficacy of this approach to resistance mapping is dependent on the spatial scale over which allele frequencies vary. In this study, we examined 13 populations including four populations from Thailand, three from Laos, and two from both Burma and Bangladesh. We found homogeneous allele frequencies within countries for alleles at loci (*dhfr* and *dhps*) underlying antifolate resistance, with most variation observed in between-country comparisons. These data suggest that low-density sampling of parasite populations may provide sufficient information for developing rational antimalarial policy. On the other hand, *pfcrt* polymorphism is more heterogeneous within countries and is less encouraging. The K76T has a patchy distribution, with locations situated <200 miles apart showing $>30\%$ difference in allele frequency. In this case, more intensive sampling may be required to evaluate countrywide resistance levels.

Supplementary Material

Supplementary Tables S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Jianbing Mu provided details of Chr 3 SNP locations, while Elizabeth Winzeler provided unpublished details from her group hybridization study of Chr 2 SNPs. The following people assisted in sample collection: Lucy Phaipun, Alan Brockman, and the staff of the Shoklo Malaria Research Unit; Samlane Phompida, Rattanaxay Phetsouvanh, Maniphone Khanthavone, and the staff of the Centre for Malariology, Parasitology, and Entomology, Vientiane, Laos; Wim Van Damme, Louis Goubert, Nuong Sophy, and Louis Goubert (Médecins Sans Frontières, Cambodia); staff at Médecins Sans Frontières-Holland (Myanmar); and in Vietnam, the Wellcome Unit, Hospital for Tropical Diseases, Ho Chi Min City. This investigation was conducted in a facility constructed with support from Research Facilities Improvement Program Grant Number C06 RR-013556 from the National Center for Research Resources at National Institutes of Health (NIH) and supported by NIH grant AI48071 (T.J.C.A.). The SMRU is part of the Wellcome-Trust Mahidol University-Oxford Tropical Medicine Research Program supported by the Wellcome Trust of Great Britain. F.N., P.N., M.M., and N.W. are supported by the Wellcome Trust.

Literature Cited

Abdel-Muhsin, A. A., M. J. Mackinnon, P. Awadalla, E. Ali, S. Suleiman, S. Ahmed, D. Walliker, and H. A. Babiker. 2003. Local differentiation in Plasmodium falciparum drug resistance genes in Sudan. Parasitology **126**:391–400.

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. 2004. Mapping drug resistance genes in Plasmodium falciparum by genome-wide association. Curr. Drug Targets Infect. Disord. **4**:65–78.

Anderson, T. J., B. Haubold, J. T. Williams et al. (16 co-authors). 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol. **17**:1467–1482.

Anderson, T. J., S. Nair, C. Jacobzone, A. Zavai, and S. Balkan. 2003. Molecular assessment of drug resistance in Plasmodium falciparum from Bahr El Gazal province, Sudan. Trop. Med. Int. Health **8**:1068–1073.

Anderson, T. J., S. Nair, W. G. Qiu, S. Singlam, A. Brockman, L. Paiphun, and F. Nosten. 2005. Are transporter genes other than the chloroquine resistance locus (*pfcrt*) and multi drug resistance gene (*pfmdr*) associated with antimalarial drug resistance? Antimicrob. Agents Chemother. **49**:2180–2188.

Anthony, T. G., D. J. Conway, J. Cox-Singh, A. Matusop, S. Ratnam, S. Shamsul, and B. Singh. 2005. Fragmented population structure of Plasmodium falciparum in a region of declining endemicity. J. Infect. Dis. **191**:1558–1564.

Barbujani, G. 1985. A two-step test for the heterogeneity of Fst values at different loci. Hum. Hered. **35**:292–295.

Beaumont, M., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B Biol. Sci. **263**:1619–1626.

Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. **13**:969–980.

Bowcock, A. M., J. R. Kidd, J. L. Mountain, J. M. Hebert, L. Carotenuto, K. K. Kidd, and L. L. Cavalli-Sforza. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc. Natl. Acad. Sci. USA **88**:839–843.

Bull, P. C., B. S. Lowe, M. Kortok, C. S. Molyneux, C. I. Newbold, and K. Marsh. 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat. Med. **4**:358–360.

Bustamante, P. J., D. C. Woodruff, J. Oh, D. B. Keister, O. Muratova, and K. C. Williamson. 2000. Differential ability of specific regions of Plasmodium falciparum sexual-stage antigen, Pfs230, to induce malaria transmission-blocking immunity. Parasite Immunol. **22**:373–380.

Campbell, D., and L. Bernatchez. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Mol. Biol. Evol. **21**:945–956.

Conway, D. J., D. R. Cavanagh, K. Tanabe et al. (12 co-authors). 2000*a*. 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 et al. (12 co-authors). 2000*b*. Origin of Plasmodium falciparum malaria is traced by mitochondrial DNA. Mol. Biochem. Parasitol. **111**:163–171.

Conway, D. J., R. L. Machado, B. Singh, P. Dessert, Z. S. Mikes, M. M. Povoa, A. M. Oduola, and C. Roper. 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.

Conway, D. J., and S. D. Polley. 2002. Measuring immune selection. Parasitology **125**(Suppl.):S3–S16.

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

Cortese, J. F., and C. V. Plowe. 1998. Antifolate resistance due to new and known *Plasmodium falciparum* dihydrofolate reductase mutations expressed in yeast. Mol. Biochem. Parasitol. **94**:205–214.

Djimde, A., O. K. Doumbo, J. F. Cortese et al. (13 co-authors). 2001. A molecular marker for chloroquine-resistant falciparum malaria. N. Engl. J. Med. **344**:257–263.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, N.J.

Fidock, D. A., T. Nomura, A. K. Talley et al. (14 co-authors). 2000. Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol. Cells **6**:861–871.

Flint, J., J. Bond, D. C. Rees et al. (10 co-authors). 1999. Minisatellite mutational processes reduce F(st) estimates. Hum. Genet. **105**:567–576.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices. Version 2.9.1. Available from http://www.unil.ch/izea/softwares/fstat.html.

Hayton, K., and X. Z. Su. 2004. Genetic and biochemical aspects of drug resistance in malaria parasites. Curr. Drug Targets Infect. Disord. **4**:1–10.

Hill, W. G., and H. A. Babiker. 1995. Estimation of numbers of malaria clones in blood samples. Proc. R. Soc. Lond. B Biol. Sci. **262**:249–257.

Hinds, D. A., L. L. Stuve, G. B. Nilsen, E. Halperin, E. Eskin, D. G. Ballinger, K. A. Frazer, and D. R. Cox. 2005. Whole-genome patterns of common DNA variation in three human populations. Science **307**:1072–1079.

Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. Proc. Natl. Acad. Sci. USA **98**:9157–9160.

Hu, H., P. Singhasivanon, N. P. Salazar et al. (10 co-authors). 1998. Factors influencing malaria endemicity in Yunnan Province, PR China (analysis of spatial pattern by GIS). Geographical Information System. Southeast Asian J. Trop. Med. Public Health **29**:191–200.

Hughes, M. K., and A. L. Hughes. 1995. Natural selection on *Plasmodium* surface proteins. Mol. Biochem. Parasitol. **71**:99–113.

Joy, D. A., X. Feng, J. Mu et al. (12 co-authors). 2003. Early origin and recent expansion of Plasmodium falciparum. Science **300**:318–321.

Kohn, M. H., H. J. Pelz, and R. K. Wayne. 2003. Locus-specific genetic differentiation at Rw among warfarin-resistant rat (Rattus norvegicus) populations. Genetics **164**:1055–1070.

Kublin, J. G., J. F. Cortese, E. M. Njunju et al. (10 co-authors). 2003. Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. J. Infect. Dis. **187**:1870–1875.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics **74**:175–195.

Luxemburger, C., F. Ricci, F. Nosten, D. Raimond, S. Bathet, and N. J. White. 1997. The epidemiology of severe malaria in an area of low transmission in Thailand. Trans. R. Soc. Trop. Med. Hyg. **91**:256–262.

Luxemburger, C., K. L. Thwai, N. J. White, H. K. Webster, D. E. Kyle, L. Maelankirri, T. Chongsuphajaisiddhi, and F. Nosten.

1996. The epidemiology of malaria in a Karen population on the western border of Thailand. Trans. R. Soc. Trop. Med. Hyg. **90**:105–111.

Machado, R. L., M. M. Povoa, V. S. Calvosa, M. U. Ferreira, A. R. Rossit, E. J. dos Santos, and D. J. Conway. 2004. Genetic structure of Plasmodium falciparum populations in the Brazilian Amazon region. J. Infect. Dis. **190**:1547–1555.

Mayxay, M., M. Khanthavong, N. Lindegardh et al. (12 co-authors). 2004. Randomized comparison of chloroquine plus sulfadoxine-pyrimethamine versus artesunate plus mefloquine versus artemether-lumefantrine in the treatment of uncomplicated falciparum malaria in the Lao People's Democratic Republic. Clin. Infect. Dis. **39**:1139–1147.

McDonald, J. H. 1994. Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms. Pp. 88–100 *in* B. Golding, ed. Non-neutral evolution: theories and molecular. Chapman and Hall, New York.

Mu, J., J. Duan, K. D. Makova, D. A. Joy, C. Q. Huynh, O. H. Branch, W. H. Li, and X. Z. Su. 2002. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum.* Nature **418**:323–326.

Nair, S., A. Brockman, L. Paiphun, F. Nosten, and T. J. Anderson. 2002. Rapid genotyping of loci involved in antifolate drug resistance in *Plasmodium falciparum* by primer extension. Int. J. Parasitol. **32**:852–858.

Nair, S., J. T. Williams, A. Brockman et al. (12 co-authors). 2003. A selective sweep driven by pyrimethamine treatment in SE Asian malaria parasites. Mol. Biol. Evol. **20**:1526–1536.

Nash, D., S. Nair, M. Mayxay, P. Newton, J. P. Guthmann, F. Nosten, and T. J. Anderson. 2005. Selection strength and hitchhiking around two antimalarial resistance genes. Proc. R. Soc. Lond. B Biol. Sci. **272**:1153–1161.

Nei, M., and T. Maruyama. 1975. Letters to the editors: Lewontin-Krakauer test for neutral genes. Genetics **80**:395.

Newbold, C. I. 1999. Antigenic variation in Plasmodium falciparum: mechanisms and consequences. Curr. Opin. Microbiol. **2**:420–425.

Plowe, C. V., J. F. Cortese, A. Djimde et al. (12 co-authors). 1997. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. J. Infect. Dis. **176**:1590–1596.

Polley, S. D., W. Chokejindachai, and D. J. Conway. 2003. Allele frequency-based analyses robustly map sequence sites under balancing selection in a malaria vaccine candidate antigen. Genetics **165**:555–561.

Price, R. N., A. C. Uhlemann, A. Brockman et al. (12 co-authors). 2004. Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. Lancet **364**:438–447.

Pritchard, J. K., and P. Donnelly. 2001. Case-control studies of association in structured or admixed populations. Theor. Popul. Biol. **60**:227–237.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. Am. J. Hum. Genet. **67**:170–181.

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.

Reed, M. B., K. J. Saliba, S. R. Caruana, K. Kirk, and A. F. Cowman. 2000. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. Nature **403**:906–909.

Robertson, A. 1975*a*. Gene frequency distributions as a test of selective neutrality. Genetics **81**:775–785.

———. 1975*b*. Letters to the editors: remarks on the Lewontin-Krakauer test. Genetics **80**:396.

Roper, C., R. Pearce, B. Bredenkamp, J. Gumede, C. Drakeley, F. Mosha, D. Chandramohan, and B. Sharp. 2003. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. Lancet **361**:1174–1181.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics **145**:1219–1228.

Sirawaraporn, W., T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, and D. V. Santi. 1997. Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. Proc. Natl. Acad. Sci. USA **94**:1124–1129.

Storz, J. F., and J. M. Dubach. 2004. Natural selection drives altitudinal divergence at the albumin locus in deer mice, Peromyscus maniculatus. Evolution Int. J. Org. Evolution **58**:1342–1352.

Storz, J. F., and M. W. Nachman. 2003. Natural selection on protein polymorphism in the rodent genus Peromyscus: evidence from interlocus contrasts. Evolution Int. J. Org. Evolution **57**:2628–2635.

Storz, J. F., B. A. Payseur, and M. W. Nachman. 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol. Biol. Evol. **21**:1800–1811.

Taylor, M. F., Y. Shen, and M. E. Kreitman. 1995. A population genetic test of selection at the molecular level. Science **270**:1497–1499.

van den Broek, I., W. S. van der, L. Talukder, S. Chakma, A. Brockman, S. Nair, and T. C. Anderson. 2004. Drug resistance in Plasmodium falciparum from the Chittagong Hill Tracts, Bangladesh. Trop. Med. Int. Health **9**:680–687.

Vasemagi, A., J. Nilsson, and C. R. Primmer. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in atlantic salmon (Salmo salar L.). Mol. Biol. Evol. **22**:1067–1076.

Volkman, S. K., D. L. Hartl, D. F. Wirth et al. (11 co-authors). 2002. Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum.* Science **298**:216–218.

Walton, C., J. M. Handley, F. H. Collins, V. Baimai, R. E. Harbach, V. Deesin, and R. K. Butlin. 2001. Genetic population structure and introgression in Anopheles dirus mosquitoes in South-east Asia. Mol. Ecol. **10**:569–580.

Walton, C., J. M. Handley, W. Tun-Lin, F. H. Collins, R. E. Harbach, V. Baimai, and R. K. Butlin. 2000. Population structure and population history of Anopheles dirus mosquitoes in southeast Asia. Mol. Biol. Evol. **17**:962–974.

Wang, X., J. Mu, G. Li, P. Chen, X. Guo, L. Fu, L. Chen, X. Su, and T. E. Wellems. 2005. Decreased prevalence of the *Plasmodium falciparum* chloroquine resistance transporter 76T marker associated with cessation of chloroquine use against P. falciparum malaria in Hainan, People's Republic of China. Am. J. Trop. Med. Hyg. **72**:410–414.

Wilding, C., R. Butlin, and J. Grahame. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. J. Evol. Biol. **14**:611–619.

Williamson, K. C. 2003. Pfs230: from malaria transmission-blocking vaccine candidate toward function. Parasite Immunol. **25**:351–359.

Wootton, J. C., X. Feng, M. T. Ferdig, R. A. Cooper, J. Mu, D. I. Baruch, A. J. Magill, and X. Z. Su. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum.* Nature **418**:320–323.

Edward Holmes, Associate Editor

Accepted July 25, 2005