

Recurrent Gene Amplification and Soft Selective Sweeps during Evolution of Multidrug Resistance in Malaria Parasites

Shalini Nair,* Denae Nash,* Daniel Sudimack,* Anchalee Jaidee,† Marion Barends,†‡ Anne-Catrin Uhlemann,§ Sanjeev Krishna,§ François Nosten,†‡|| and Tim J. C. Anderson*

*Southwest Foundation for Biomedical Research, San Antonio, Texas; †Shoklo Malaria Research Unit, Mae Sot, Tak, Thailand; ‡Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; §Division of Cellular and Molecular Medicine, Centre for Infection, St. George's University of London, London, United Kingdom; and ||Centre for Tropical Medicine and Vaccinology, Churchill Hospital, Oxford, United Kingdom

When selection is strong and beneficial alleles have a single origin, local reductions in genetic diversity are expected. However, when beneficial alleles have multiple origins or were segregating in the population prior to a change in selection regime, the impact on genetic diversity may be less clear. We describe an example of such a “soft” selective sweep in the malaria parasite *Plasmodium falciparum* that involves adaptive genome rearrangements. Amplification in copy number of genome regions containing the *pfmdr1* gene on chromosome 5 confer resistance to mefloquine and spread rapidly in the 1990s. Using flanking microsatellite data and real-time polymerase chain reaction determination of copy number, we show that 5–15 independent amplification events have occurred in parasites on the Thailand/Burma border. The amplified genome regions (amplicons) range in size from 14.7 to 49 kb and contain 2–11 genes, with 2–4 copies arranged in tandem. To examine the impact of drug selection on flanking variation, we genotyped 48 microsatellites on chromosome 5 in 326 parasites from a single Thai location. Diversity was reduced in a 170- to 250-kb (10–15 cM) region of chromosomes containing multiple copies of *pfmdr1*, consistent with hitchhiking resulting from the rapid recent spread of selected chromosomes. However, diversity immediately flanking *pfmdr1* is reduced by only 42% on chromosomes bearing multiple amplicons relative to chromosomes carrying a single copy. We highlight 2 features of these results: 1) All amplicon break points occur in monomeric A/T tracts (9–45 bp). Given the abundance of these tracts in *P. falciparum*, we expect that duplications will occur frequently at multiple genomic locations and have been underestimated as drivers of phenotypic evolution in this pathogen. 2) The signature left by the spread of amplified genome segments is broad, but results in only limited reduction in diversity. If such “soft” sweeps are common in nature, statistical methods based on diversity reduction may be inefficient at detecting evidence for selection in genome-wide marker screens. This may be particularly likely when mutation rate is high, as appears to be the case for gene duplications, and in pathogen populations where effective population sizes are typically very large.

Introduction

In classical selection models, a single newly arisen mutation spreads through a population dragging flanking neutral polymorphisms in its wake (Smith and Haigh 1974). This purges genetic variation and creates islands of elevated linkage disequilibrium in the vicinity of the selected site. Hence searching for these characteristic features can potentially provide a powerful means of locating genome regions involved in adaptation (Pollinger et al. 2005; Voight et al. 2006). However, the classical model may provide on oversimplification when there is repeated evolution of adaptive alleles (Pennings and Hermisson 2006a, 2006b) or when selection acts on standing genetic variation (Hermisson and Pennings 2005; Teshima et al. 2006). In these cases, multiple genetic backgrounds hitchhike with adaptive alleles, so much of the flanking neutral variation may be preserved. These events have been termed “soft” selective sweeps (Hermisson and Pennings 2005; Pennings and Hermisson 2006a, 2006b). Classical models of selection also generally involve point mutations with low mutation rates. However, a growing body of literature suggests that genome rearrangements, such as tandem duplications, may play a critical role in adaptive change (Coghlan et al. 2005). Whole-genome duplications have occurred commonly dur-

ing evolution (Dehal and Boore 2005; Paterson et al. 2005), while both cross species (Cheng et al. 2005) and intraspecific studies (Sharp et al. 2005) reveal extensive variation in genome size and abundant examples of copy number polymorphism. Furthermore, copy number changes show important phenotypic effects in many cases. Notably, copy number of CCR3 influences HIV resistance in humans (Townson et al. 2002; Gonzalez et al. 2005), esterase copy number influences insecticide resistance in mosquitoes (Raymond et al. 2001), tandem duplication in bacterial genes are involved in adaptation to benzoates, heavy metals, pollutants, and drugs (Romero and Palacios 1997; Reams and Neidle 2003, 2004b), and duplications are repeatedly selected during experiment evolution studies with yeast (Dunham et al. 2002) or *Escherichia coli* (Riehle et al. 2001). Copy number changes differ from point mutations in 2 important respects. First, duplications occur at a higher rate than point mutations in both prokaryotes and eukaryotes (Romero and Palacios 1997; Inoue and Lupski 2002). Second, whereas single nucleotide polymorphisms (SNPs) have a very low reversion rate, reversion of duplications to single-copy state may occur commonly due to unequal recombination. Selective events involving beneficial gene duplications may therefore differ substantially from those associated with point mutations.

Here we describe a “soft” selective sweep involving copy number amplification. The multidrug resistance locus of *Plasmodium falciparum* (*Pfmdr1*) (chromosome 5) encodes an ATP-binding cassette transporter similar to the *mdr* genes that mediate multidrug resistance in mammalian cell lines and the yeast *Candida albicans* (Duraisingh and Cowman 2005). This locus confers a true multidrug

Key words: adaptive evolution, tandem duplication, amplicon break point, adaptation, hitchhiking, soft selective sweeps, artemisinin combination therapy, antimalarial drug.

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

Mol. Biol. Evol. 24(2):562–573. 2007

doi:10.1093/molbev/msl185

Advance Access publication November 23, 2006

resistance phenotype changing in vitro response to multiple structurally unrelated antimalarials such as chloroquine (CQ), arylaminoalcohol compounds (mefloquine and halofantrine), and the newly introduced artemisinin derivatives. Both point mutations and copy number changes in *pfmdr1* can alter drug sensitivity. Five nonsynonymous point mutations commonly occur (N86Y, Y184F, S1034C, N1042D, and D1246Y). Mutations at these codons influence CQ response (Reed et al. 2000; Sidhu et al. 2005) and are selected by CQ treatment (Djimde et al. 2001). In contrast, copy number change in wild-type alleles is associated with increased resistance to mefloquine (and related compounds) and the artemisinin derivatives (Wilson et al. 1993; Price et al. 2004, 2006). Furthermore, manipulation of copy number alters response to multiple antimalarial drugs, demonstrating a causal link between gene dosage and drug resistance (Sidhu et al. 2006). Copy number change is particularly worrying because it confers increased resistance to both components of combination therapies (mefloquine–artesunate and lumefantrine–artemether) that are being introduced globally in an attempt to reduce the rate of evolution of drug resistance in malaria parasites (Price et al. 2006). Copy number changes were first documented in 1989 (Foote et al. 1989; Wilson et al. 1989) and rapidly increased in frequency in the 1990s (Price et al. 2004). These early studies demonstrated that 1) increased copy number could be selected in the laboratory by treatment with mefloquine (Cowman et al. 1994), whereas decreases were selected by CQ (Barnes et al. 1992), 2) copy number change involves large tandem amplifications up to 100 kb involving multiple genes (Foote et al. 1989; Triglia et al. 1991), and 3) copy number change was found in multiple geographical regions including South America, Southeast Asia, Africa, and Papua New Guinea (Triglia et al. 1991). Two previous studies provide useful information on *pfmdr1* copy number evolution. Triglia et al. (1991) identified the chromosomal break point sequences involved in tandem amplifications in 1 South American parasite (B8) and then used polymerase chain reaction (PCR) to determine whether the same break points were found in parasites with amplified *pfmdr1* from other locations. They found that none of 16 parasites with multiple *pfmdr1* copies from Southeast Asia, Africa, or Papua New Guinea carried the same break point and concluded that multiple independent origins have occurred. However, we note that these data are also compatible with as few as 2 independent origins because break point positions were not characterized in these samples. Similarly, Chaiyaroj et al. (1999) examined macrorestriction patterns of chromosome 5 and found 2 distinctive patterns consistent with ~100-kb and ~20-kb amplicons in Thai isolates, once again suggesting as few as 2 origins in these samples.

This study was designed to investigate the evolutionary dynamics of copy number amplification in a single geographical location. We have 2 reasons for doing this. First, recent studies have focused on the evolution of antimalarial drug resistance caused by point mutations and have shown surprisingly few origins of resistance and global selective sweeps that have purged genetic variation from large genomic regions (Wootton et al. 2002; Hastings 2004; Roper et al. 2004; Anderson and Roper 2005). These studies have suggested that genome-wide screens of marker variation

may provide a useful approach to locating additional regions of the genome that are under drug selection. We would like to compare these results with patterns of adaptive change in copy number, in which the rate of both forward and backward mutation may be very different from SNPs (Romero and Palacios 1997). Second, most studies of copy number investigate the historical remains of ancient copy number change by comparing divergence of gene families within genomes or closely related organisms (Newman et al. 2005; Sharp et al. 2005). In contrast, we are able to investigate the early stages of copy number evolution under current selection by antimalarial drugs. We use 2 complementary approaches (microsatellite genotyping of markers flanking *pfmdr1* and real-time PCR characterization of amplicon size, break points, and gene content) to investigate the evolutionary dynamics of *pfmdr1* amplification in a population sample from the Thailand–Burma border. The data generated show 1) that amplification has occurred between 5 and 15 times, generating amplicons from 15 to 49 kb in size, 2) that break points invariably occur in monomeric tracts of As or Ts, and 3) that rapid spread of chromosomes carrying amplified segments has resulted in genetic hitchhiking and “soft” selective sweeps for over 170 kb (10 cM) on chromosome 5.

Materials and Methods

Study Design

This study was designed to ask how many independent *pfmdr1* amplification events have occurred, to examine the size and break points of the chromosomal segments (amplicons) involved, and to measure the extent of genetic hitchhiking on selected chromosomes. The project involved 3 stages: 1) collection of haploid single-clone parasites from the blood of infected patients, 2) determination of *pfmdr1* copy number and measurement of amplicon size, and 3) microsatellite genotyping across chromosome 5 to identify haplotypes on which amplification has occurred and to determine the extent of genetic hitchhiking around *pfmdr1*.

Collection of Single-Clone Parasite Isolates

We collected 5 ml blood samples from *P. falciparum*-infected patients visiting the malaria clinic at Mawker-Thai on the Thailand–Burma border between December 2000 and September 2003. These patients had not visited the malaria clinic within the past 60 days and had no history of prior malaria treatment during this time. The clinic serves people on both sides of the border; 90% of patients travel from within a 10-km radius around the clinic (Nosten F, personal communication). Collection protocols were approved by the Ethical Committee of the Faculty of Tropical Medicine, Mahidol University, Bangkok and by the Institutional Review Board at the University of Texas Health Science Center at San Antonio. Parasite DNA was prepared by phenol/chloroform extraction of whole blood, following removal of buffy coats as described previously (Nair et al. 2002).

Microsatellite Genotyping

We used microsatellite markers for 3 purposes in this study. 1) Initially, we used 7 microsatellite markers to identify

infected blood samples containing multiple malaria clones. These were then removed from the study because parasite haplotypes cannot be determined when 2 or more clones are present. The methods and criteria used for removal of multiple clone infections were as described previously (Anderson, Haubold, et al. 2000; Anderson et al. 2005). 2) We genotyped all single-clone infections using microsatellite markers spaced at 50-kb (~ 3 cM) intervals across chromosome 5 (supplementary table S1, Supplementary Material online). These microsatellites were genotyped to examine the influence of genetic hitchhiking on genetic variation. 3) We placed 16 additional markers within an 18-kb region containing *pfmdr1* (between 953,644 and 971,251 bp) in order to determine variation immediately flanking this locus. These loci included 1 trinucleotide marker in the spacer region within the *pfmdr1* gene. For all analyses, we selected microsatellite markers that had >12 perfect repeats in the strain 3D7 genome sequence data in order to maximize our chances of detecting variation (Anderson, Su, et al. 2000).

Determination of *pfmdr1* Copy Number and SNP Polymorphism

We used a multiplex real-time PCR assay to measure copy number of *pfmdr1* relative to single-copy β -tubulin gene. We measured *pfmdr1* copy number relative to a standard calibrator parasite (the genome strain 3D7) that has a single copy of *pfmdr1* using the $\Delta\Delta C_T$ method. Our methods closely follow that described previously (Price et al. 2004) with 3 changes: 1) we ran samples in quadruplicate rather than in triplicate to improve precision and simplify liquid handling, 2) reactions were conducted in 10- μ l volumes on 384-well plates on an ABI 7900HT real-time PCR machine, and 3) we changed one of the primer sequences, which improved assay performance in our hands. To determine reproducibility of this assay, we measured copy number twice and compared results using contingency tables. We also genotyped 5 point mutations in these samples (N86Y, Y184F, S1034C, N1042D, and D1246Y) using primer extension as describe previously (Anderson et al. 2005).

Characterization of Amplicon Size and Break Point Junctions

To determine the extent of chromosomal regions that are duplicated or amplified on chromosome 5, we designed 25 real-time PCR assays at ~ 5 -kb intervals for 120 kb on either side of the *pfmdr1* on chromosome 5 (supplementary table S2, Supplementary Material online). We used singleplex assays with target sequence and β -tubulin amplified in separate wells. In each of these assays, we used the single-copy β -tubulin genes as our endogenous control and the genome strain 3D7, which has a single copy of *pfmdr1*, as the calibrator as described above. We measured amounts of β -tubulin product generated during PCR using Minor Groove Binding (MGB) β -tubulin probes labeled with VIC (Applied Biosystems, Foster City, CA), whereas we measured amounts of chromosome 5 genes using 6FAM labeled MGB probes. All assays were conducted in quadruplicate

on 384-well plates using an ABI 7900HT real-time PCR system. We estimated gene copy number (relative to 3D7) at varying distances from *pfmdr1* using the $\Delta\Delta C_T$ method.

The real-time assays allow approximate mapping of break points on chromosome 5. To locate the precise break points, we used a simple PCR strategy. Primers were designed facing away from each other close to the boundaries of the amplified segment (supplementary fig. S1, Supplementary Material online). These break point-specific oligos amplify a product in isolates with tandem duplications, but not in samples without duplications. Hence, we ran clone 3D7, which has a single copy of *pfmdr1*, as a negative control to ensure the specificity of all break point amplifications. We sequenced break point-specific PCR products in both directions and made alignments with the 3D7 genome sequence (Gardner et al. 2002) to identify the sequences in which chromosome breakage has occurred.

Statistical Methods

To examine the numbers of origins of copy number amplification we investigated genetic relationships between chromosome 5 segments containing *pfmdr1*. We measured the proportion of shared microsatellite alleles (*ps*) in all pairwise combinations of 16-locus haplotypes in the 18-kb region containing *pfmdr1*. We used the distance metric $-\log(ps)$ to construct UPGMA trees using PHYLIP. We used bootstrap resampling (implemented using Microsatellite Analyser [Dieringer and Schlotterer 2002]) to generate 1,000 resampled distance matrices and generated bootstrap confidence values for the nodes. The tree generated was used to examine the relationships between chromosome haplotypes bearing multiple-copy *pfmdr1*.

To examine the extent of hitchhiking on chromosome 5, we compared microsatellite diversity on chromosomes containing a single copy of *pfmdr1* with diversity on chromosomes bearing multiple copies. We quantified variation using expected heterozygosity (H_e) at each locus using the formula $H_e = [n/(n-1)] [1 - \sum p_i^2]$, where p is the frequency of i th allele and n is the number of chromosomes sampled. For haploid blood-stage malaria parasites, this statistic measures the probability of drawing 2 different alleles from a population. The sampling variance of H_e was calculated as $2(n-1)/n^3 [2(n-2) \sum p_i^3 - (\sum p_i^2)^2]$, a slight modification of the standard diploid variance (Nei 1987).

Results

Population Sample

We collected 5 ml blood samples from 733 *P. falciparum*-infected patients visiting the malaria clinic at Mawker-Thai on the Thailand–Burma border between December 2000 and September 2003. We genotyped each infected sample using 7 microsatellite markers to identify samples containing a single predominant clone. Thirty-five percent (265) of infections contained multiple clones and were excluded. Additional clones were excluded if they provided low quality in vitro drug resistance data (data not shown) or if insufficient amounts of DNA were available for molecular analyses. A total of 326 infections containing a predominant single clone were included in this analysis.

Pfmdr1 Copy Number and SNPs

We used multiplex real-time PCR to measure *pfmdr1* copy number. Multiple copies of *pfmdr1* were found in 109 of 326 (33%) of infections. When we rounded real-time data to the nearest integer value, 215 (66%) carried a single copy, 81 (25%) carried 2 copies, 22 (7%) had 3 copies, and 7 (2%) had 4 copies. Measurement of copy number was highly repeatable. The correlation coefficient (r^2) between noninteger replicate copy number estimates was 0.82 (fig. 1). Differentiation between isolates carrying single and multiple copies was particularly robust, with only 3% of samples showing discrepant values between replicate measures in contingency tables. Two codons (S1034C and D1246Y) of the 5 nonsynonymous SNPs frequently found in *pfmdr1* were monomorphic for wild-type bases, whereas 3 codons (N86Y, Y184F, and N1042D) were variable. For these 3 codons, the *pfmdr1* SNP genotypes present were NYN (65%), NFN (26%), NFD (5%), and YYN (4%) (mutant codons shown underlined). Only the 2 common genotypes (NYN and NFN) were present in parasite isolates carrying multiple *pfmdr1* copies.

Microsatellite Markers Flanking *pfmdr1*

We genotyped a cluster of 16 loci (supplementary table S1, Supplementary Material online) that were situated within an 18-kb region containing *pfmdr1* in order to investigate whether copy number amplification has occurred on different genetic backgrounds. We constructed a UPGMA tree based on a simple allele-sharing statistic to examine the haplotypes associated with *pfmdr1* amplification (fig. 2A). This analysis was performed only on those samples that contained complete microsatellite data for all 16 loci and for which a single allele was observed at each locus ($n = 254$). We found 5 groups of 16-locus haplotypes associated with multiple-copy *pfmdr1*. Parasites in each of these groups shared identical alleles at $\geq 12/16$ loci flanking *pfmdr1*. These groups (designated clades a–e) contained 8, 62, 5, 2, and 8 parasites with multiple copies of *pfmdr1*, respectively. We found that copy number varied in parasites from a single clade. For instance, in clade B *pfmdr1* may be copied 2–4 times. Furthermore, parasites bearing single-copy *pfmdr1* often had identical microsatellite haplotypes as parasites bearing multiple copies. The SNP data from *pfmdr1* provide additional information for defining the amplicons. Alleles from clades A, B, and E contained *pfmdr1* SNP genotypes NYN, whereas alleles from clades C and D carried NFN at residues 86, 184, and 1042. We observed no evidence for variation at SNPs on different copies of *pfmdr1* within a parasite clone.

Amplicon Size and Break Point Junction Sequences

The analysis of flanking 16-locus microsatellites haplotypes revealed 5 different groups of haplotypes associated with copy number amplification (fig. 2A). To further characterize amplicon evolution, we identified the position of chromosomal break points in these samples. We initially chose 2 or 3 samples from each haplotype group for further characterization of amplicon size using a battery of 25 real-time assays (supplementary table S2, Supplementary Mate-

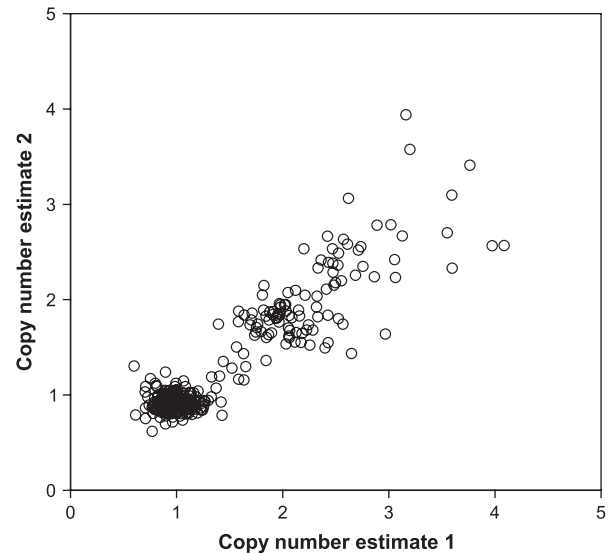


Fig. 1.—Reproducibility of real-time assays of *pfmdr1* copy number. Results from independent assays of *pfmdr1* copy number plotted against one another ($r^2 = 0.82$).

rial online) spanning 120 kb on either side of *pfmdr1* (fig. 2B). To map the precise chromosomal break points, we used break point–specific PCR primers (supplementary fig. S1, Supplementary Material online) to amplify and sequence across break point junctions. We aligned the junction sequences with the 3D7 genome sequence to identify the genome sequences involved. Break point–specific primers are listed in supplementary table S3 (Supplementary Material online), whereas alignments of break point sequences with 3D7 are shown in supplementary figure S2 (Supplementary Material online).

We tested all isolates bearing multiple copies of *pfmdr1* using the break point–specific PCR assays. Generation of a product of the expected size allowed us to determine the frequency of different amplicon types in the population sample. In some isolates none of the break point–specific assays developed initially amplified products from samples known to carry >1 copy of *pfmdr1*. In this case, additional real-time assays were conducted on these samples and break points identified as described. Using this iterative process, we successfully determined break points for 12 of 15 amplicons detected by real-time PCR (fig. 2C). These amplicons ranged in size from 14.7 to 49 kb and contained between 2 and 11 genes (table 1). The commonest amplicon of these was found on clade B (table 1 and fig. 2A). This measured 16.4 kb, contained 3 genes, and was found in 57 samples (table 1). Other amplicons were only found in a single isolate. We found distinctive amplicons on identical microsatellite backgrounds. For example, 5 different amplicons were found for parasites in clade A, whereas 3, 3, 1, and 3 amplicon types were found for parasites falling in clade B, C, D, and E, respectively (fig. 2A and C).

We also examined amplicon size and break point sequences in 2 laboratory isolates (B8 from South America and Dd2 from Southeast Asia). B8 provides an important control for our methods because break points were previously characterized in this parasite by Triglia et al.

(1991). The amplicon we mapped was 95 kb in size, and the amplicon junction sequences we identified were identical to those found previously (1991). The Dd2 amplicons were generated by laboratory selection of asexual parasite stages in continuous blood culture. The amplicon measured 81 kb and contained 14 genes.

All amplicon break points were found in monomeric tracts of A or T in intergenic regions (table 1; supplementary fig. S2, Supplementary Material online). The monomeric tracts involved ranged in size from 9 to 45 bp in the 3D7 genome sequence. We compared the distribution of tract sizes associated with break points with the distribution of monomeric tracts sampled from the complete 3D7 genome and for the region spanning 100 kb on either side of *pfmdr1* (fig. 3). Monomeric tracts of A- or T-containing break points ($n = 28$) were significantly longer than those from the whole genome ($n = 37,177$; Mann–Whitney U test, $Z = -4.80$, $P < 0.0001$) or from 100 kb on either side of *pfmdr1* on chromosome 5 ($n = 439$; Mann–Whitney U test, $Z = -4.72$, $P < 0.0001$). The same break point junctions were frequently found in different amplicons and on different microsatellite backgrounds. For example, 5 of the 15 distinct amplicons shared the same 5' junction site. These included amplicons 4 and 5 on clade A and 7, 12, and 13 on clades B, C, and D, respectively. Similarly, three 3' junction sites were shared between different amplicons (3, 4, and 7) on clades A and B, and 1 amplicon with identical 5' and 3' junction sites was found on 2 different microsatellite clades (clades A and B).

Variation across Chromosome 5

To measure the extent of hitchhiking resulting from spread of multiple-copy *pfmdr1*, we plotted H_e for 48 microsatellite markers on chromosome 5 in parasites with single or multiple copies of *pfmdr1* (fig. 4). These included the cluster of 16 markers used to describe the genetic backgrounds associated with amplification and are listed in supplementary table S1 (Supplementary Material online). These data reveal that variation is reduced for at least 131 kb on the 5' side of *pfmdr1* and for at least 36 kb to the 3' end. The total window of reduced variation spans between 170 and 250 kb (10–15 cM). Amplicon 7 (table 1) plays a disproportionate role in generating the observed reduction in variation. This amplicon accounts for 67% of samples with more than 1 copy of *pfmdr1*. Although a broad swathe of chromosome 5 is affected by hitchhiking, the reduction in diversity was limited. We find that H_e is reduced by just 42% (range, 21–63%) in the 16 loci surrounding *pfmdr1* on chromosomes carrying multiple copies of *pfmdr1* relative to those carrying a single copy.

Discussion

Multiple Origins of Amplification and Soft Selective Sweeps

We observed multiple independent origins of gene amplification. Genotyping of microsatellite markers flanking *pfmdr1* show that amplification has occurred on 5 different haplotype groups, suggesting at least 5 independent origins. However, the real-time mapping of break point se-

quences and amplicon size revealed 15 distinct amplicons, consistent with up to 15 independent amplification events. Which of these estimates is nearest to the truth? Microsatellite data may underestimate numbers of origins, if duplications have occurred independently on similar genetic backgrounds. Similarly, the number of amplicons could provide an overestimate because distinctive amplicons could be generated by recombination. Perhaps the best estimate comes from counting the number of independent 5' or 3' break point sequences because this will not be influenced by recombination. We found 8 distinct 5' break points and 10 distinct 3' break points. Furthermore, these estimates are derived from just 1 location. Genotyping of microsatellites immediately flanking *pfmdr1* in infections collected from Mae Ia (situated 120 km away from Mawker-Thai) show that additional microsatellite backgrounds are associated with copy number amplification (Uhlemann A-C, Anderson TJC, unpublished observations). Hence, our estimates almost certainly underestimate numbers of origins across Thailand or Southeast Asia and suggest that copy number change occurs very frequently.

Selective events in which there have been multiple origins of the beneficial alleles have been termed “soft” sweeps (Hermisson and Pennings 2005; Pennings and Hermisson 2006a, 2006b). A number of researchers have analyzed expected patterns of variation in markers flanking selected sites in such soft sweeps (Innan and Kim 2004; Hermisson and Pennings 2005; Przeworski et al. 2005; Pennings and Hermisson 2006a, 2006b). Multiple origins of selected alleles are expected when mutation rates are high and populations are large. More precisely, multiple origins are expected when the population-level mutation parameter $2N_e\mu > 0.01$, where N_e is the effective population size and μ is the per generation rate of mutation to the favorable allele. Estimates of both μ and N_e are large in this case. We do not have direct estimates of the μ for copy number change for *P. falciparum*. However, in *Salmonella* mutation rates may be as high as 10^{-2} – 10^{-5} for copy number change in any region of the genome (Anderson and Roth 1977; Romero and Palacios 1997), whereas in *Acinetobacter* tandem amplifications underlying benzoate resistance arise at a frequency of $\sim 10^{-6}$ (Reams and Neidle 2003). Similarly, in eukaryotes, genome rearrangements occur at a frequency that is 1–2 orders of magnitude greater than point mutations (Inoue and Lupski 2002; Repping et al. 2006). N_e estimates from mtDNA sequence variation are $\sim 10^5$ for Asian *P. falciparum* populations (Joy et al. 2003). Hence, with gene duplication rate conservatively estimated at 10^{-5} – 10^{-6} and population based N_e of 10^5 estimates of the population mutation parameter $2N_e\mu$ equals 0.2–2. Hence our observation of multiple independent origins is consistent with theory, if we assume biologically plausible mutation rate estimates.

We use the theoretical framework developed by Pennings and Hermisson (2006a, 2006b) to further explore how our empirical data conform to expectations. The expected reduction in H_e in the direct neighborhood of the selected locus is given by $1/(1 + 2N_e\mu)$ (Pennings and Hermisson 2006b, Equation 8). The 42% reduction observed corresponds to a $2N_e\mu$ of ~ 1.4 . Given estimates for N_e (10^5) and μ (10^{-5} – 10^{-6}), this seems plausible. Using an estimate of 1.4 for $2N_e\mu$, the theory also predicts the

average number of independent origins of a beneficial allele (Pennings and Hermisson 2006a, Equation 12). In a sample of size 85 (only carriers of the amplified *pfmdr1*), this average number is 6.3. Again, this corresponds reasonably well with our empirical derived estimates of 5–15 origins. There are even predictions for the frequency distribution of beneficial alleles with different origins in a population sample (Pennings and Hermisson 2006a, Equation 13). For a given value of $2N_e\mu$, this is expected to follow the Ewens sampling distribution, with overrepresentation of some ancestral haplotypes in the sample. In our data, chromosomes carrying a single initial duplication event predominate, comprising 67% of all chromosomes carrying more than 1 copy of *pfmdr1*, consistent with this prediction (table 1). Although these empirical data show encouraging concordance with theoretical predictions that assume a random mating population of constant size, it should be stressed that an optimal model should include population structure and demography with serial bottlenecks. The properties of such a model have not yet been explored. Furthermore, it should be possible to directly measure *pfmdr1* duplication events in asexual cultures of *P. falciparum* in the laboratory to provide empirical measures of μ for this system.

These data provide a stark contrast with evolution of CQ and pyrimethamine resistance in malaria parasites. Point mutations in the CQ resistance transporter (*pfcr1*) result in resistance to CQ (Fidock et al. 2000), whereas mutations in dihydrofolate reductase (*dhfr*) result in progressively increasing levels of resistance to pyrimethamine (Plowe et al. 1998). In both cases, there has been a single origin of resistance alleles bearing multiple point mutations in Southeast Asia and these resistance alleles have subsequently spread to Africa (Wootton et al. 2002; Nair et al. 2003; Roper et al. 2004; Nash et al. 2005). The reduced numbers of origins in these genes most likely reflects the fact that SNP mutations have lower mutation rate than copy number. Furthermore, in both these genes, multiple point mutations are involved, which further reduces the frequency with which resistance alleles arise. For example, 3–4 mutations are required for high-level pyrimethamine resistance, whereas ~5 additional nonsynonymous mutations are found together with the critical K76T mutation on resistance alleles at the CQ resistance transporter (*pfcr1*) (Fidock et al. 2000). Independent origins of *pfcr1* mutations have occurred in different continents resulting in parallel selective sweeps in South America and Papua New Guinea (Wootton et al. 2002). However, independent origins of copy number amplification at *pfmdr1* occur far more frequently, so multiple independent origins are observed “within” single parasite populations, obscuring signatures of selection.

There has been much recent interest in using patterns of genome-wide variation to search for genes underlying phenotypic traits and to identify genome regions that have been under strong selection in both malaria parasites (Anderson 2004; Su and Wootton 2004) and other organisms (Schlotterer 2003; Nielsen 2005; Storz 2005; Voight et al. 2006). This rationale underlies projects currently underway at the Broad Institute (Volkman S, personal communication), the Wellcome Trust Sanger Institute (Newbold C, Berriman M, personal communication) and

National Institutes of Health (NIH) (Su X-Z, personal communication), which aim to generate a dense SNP hapmap by sequencing multiple parasite isolates. These results sound a note of caution for these genome-wide approaches. We found reduced genetic variation for between 170 and 250 kb (10–15 cM) around *pfmdr1* on chromosome 5 in parasites carrying multiple copies relative to those carrying a single copy. The regions affected included at least 131 kb on the 5' side and >36 kb to the 3' of *pfmdr1*. The most likely explanation for this is genetic hitchhiking resulting from rapid spread of selected chromosomes under strong mefloquine selection during the 1990s (Price et al. 2004) (supplementary fig. S3, Supplementary Material online). The size of the chromosomal regions affected is comparable to that recorded around the CQ resistance locus *pfcr1* (195–268 kb [11–16 cM]) on chromosome 7 or pyrimethamine resistance alleles at *dhfr* (98–137 kb [6–8 cM]) on chromosome 4 in the same study site (Nash et al. 2005). However, the reduction in heterozygosity is modest compared with the chromosome 4 and chromosome 7 selective events, in which variation was completely removed close to the selected alleles. We find that H_e is reduced by just 42% (range 21–63%) in the 16 loci surrounding *pfmdr1* on chromosomes carrying multiple copies of *pfmdr1* relative to those carrying a single copy. In fact, the reduction in diversity that we observed in this case is almost completely driven by the one very common amplicon associated with resistance (the 16.4-kb amplicon on haplotype clade B). The reduction in diversity is limited for 2 reasons. First, amplified *pfmdr1* is found on multiple genetic backgrounds because these alleles have evolved multiple times. Second, chromosomes carrying single and multiple copies of *pfmdr1* are often identical. This pattern is predicted for copy number changes where reversion is common. If multiple origins and parallel evolution are common in microbial populations, as we suspect, then genome-wide screens will lack power to detect signals of selection (Przeworski et al. 2005). In particular, methods that search for reduction in genetic diversity (“hitchhiking mapping” [Schlotterer 2003]) or skews in the frequency of rare alleles may be ineffective when there are multiple selected alleles because background variation is only weakly affected. Methods that exploit patterns of linkage disequilibrium are more promising (Sabeti et al. 2002; Toomajian et al. 2003; Hanchard et al. 2006). However, these methods typically assume a single selected allele and infer selection by comparison among haplotypes within a genomic region (Sabeti et al. 2002). This assumption is violated if multiple haplotypes have been selected.

Dynamics of Amplicons within Populations

Our data suggest rapid turnover of amplicon types within parasite populations, with small streamlined amplicons replacing larger versions. The amplicons found in this study were small (15–49 kb) compared with those in laboratory isolates B8 and Dd2 and with those inferred from a previous study in which chromosome 5 from Thai parasites was digested and the fragments separated by pulsed field gel electrophoresis (Chaiyaroj et al. 1999). In their study, the predominant amplicon type was associated with

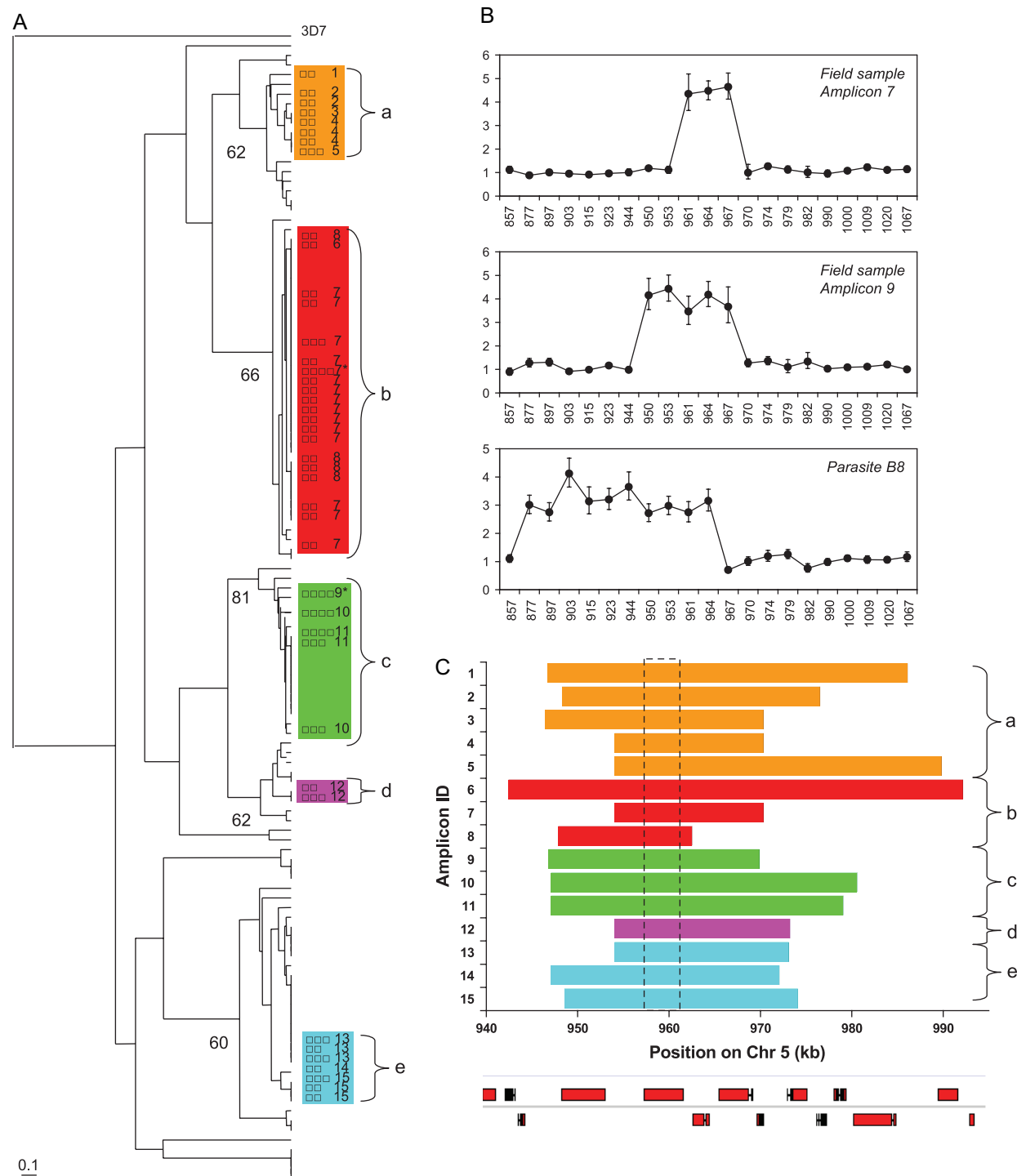


FIG. 2.—Genomic organization of amplified chromosomal regions. (A) UPGMA tree showing the relationship between 16-locus microsatellite haplotypes on chromosome 5. The tree is based on a simple distance metric ($-\log ps$), where ps is the proportion of shared alleles. Only 120 parasites are shown for clarity. These include all parasites with amplicons in clades a, b, d, and e, and a random sample of parasites bearing type b amplicons or with a single copy of *pfmdr1*. The unfilled squares at the branch tips indicate the number of copies of *pfmdr1* (Unmarked, 1; \square , 2; $\square\square$, 3; $\square\square\square$, 4). The numbers (1–15) are identifiers describing the different amplicon types shown in panel C of this figure and table 1. Bootstrap support values $>60\%$ are shown at the tree nodes. Brackets labeled a–e and color shading indicate the 5 related groups of haplotypes carrying more than 1 copy of *pfmdr1*. (B) Size and span of amplicons containing *pfmdr1*. Estimates of copy number derived from real-time PCR are plotted against chromosomal position for parasite stain B8 from South America and 2 field isolates (marked by asterisks in panel A) from the Thailand–Burma border. The error bars show 95% confidence intervals around our estimates of copy number in each position. The amplicon size and break point positions for B8 are consistent with a previous study (Triglia et al. 1991), demonstrating the utility of our methods. *Pfmdr1* is located between 958 and 962 kb. (C) Bar chart showing break points and size of amplicons. The positions of the break points were determined by sequencing. For 3 amplicons (10, 11, and 14), exact break point positions could not be determined: for these we show approximate sizes estimated from real-time PCR. The x axis shows the position of the start and end of amplicons relative to the 3D7 genome sequence, whereas the gene content of this region is shown at the base of the figure. The position of *pfmdr1* is marked by the dotted box.

Table 1
Amplicon Descriptions

ID	Clade	5' Break Point		3' Break Point		Amplicon Size (kb)	SNPs	Number of Samples
		Position ^a	Sequence	Position ^a	Sequence			
Field samples								
1	A	946617	T ₂₅	986030	T ₃₂	39.41	NYN	1
2	A	948204	T ₄ CT ₂₁ AT ₄ AT ₃ GT ₆	976440	T ₂₆	28.24	NYN	2
3	A	946351	T ₃₀	970240	T ₃₈	23.89	NYN	1
4	A	953950	T ₂₂	970240	T ₃₈	16.29	NYN	3
5	A	953950	T ₂₂	989750	T ₂₀	35.80	NYN	1
6	B	942350	T ₉ CT ₆ CTCT ₂₆	992081	T ₂₈	49.73	NYN	1
7	B	953950	T ₂₂	970240	T ₃₈	16.29	NYN	57
8	B	947783	T ₄ AT ₁₁	962437	T ₁₆ CCT ₂ CT ₂ CCT ₁₁ A ₆ A ₇	14.65	NYN	4
9	C	946690	A ₂₂	969790	A ₄ TA ₅ GA ₃₀	23.10	NFN	1
10	C	?	?	?	?	32.00	NFN	2
11	C	?	?	?	?	33.50	NFN	2
12	D	953950	T ₂₂	973150	T ₂₃	19.20	NFN	2
13	E	953950	T ₂₂	973004	T ₂₆	19.05	NYN	4
14	E	?	?	?	?	25.00	NYN	1
15	E	948458	A ₁₉	974010	A ₉	25.55	NYN	3
Laboratory isolates								
B8		868690	A ₃₃	964500	A ₃₆	95.81		0
DD2		888335	T ₂₆	970240	T ₃₈	81.91		0

NOTE.—The column entitled “Clade” indicates the genetic background of these amplicons, determined by analysis of microsatellite markers (fig. 2A). Because chromosomal breakage sites contain runs of A or T, the precise positions cannot be determined; the base positions listed are accurate to within ~10 bp. The sequence of the breakage sites is from 3D7. The number of A or T nucleotides are likely to vary in the field samples. Alignments of the chromosomal breakage sites and junction sequences are provided in the supplementary material (supplementary fig. S2, Supplementary Material online). Amino acids in residues 86, 184, and 1042 are described by 3-letter codes. The number of samples bearing each of the 15 amplicon types is shown. Amplicons found in 2 laboratory isolates are also shown for comparison. The B8 amplicon has previously been characterized by Triglia et al. (1991). “?” indicates that we were unable to locate chromosomal break points in these isolates.

^a bp position on chr 5 of *P. falciparum* genome sequence

a *Bgl*I fragment of ~100 kb, suggesting amplicons of this size were present. These samples were collected in the late 1990s from patients acquiring malaria in Thailand. Although the sampling locations are not directly comparable (Chaiyaroj et al. sampled parasites from patients in a Bangkok hospital from the Thailand–Burma border, as well as other regions in Thailand, whereas our samples come from a single clinic in Tak province), the size of amplicons are very different and suggest that a reduction in amplicons size has occurred over the past 10 years. Large amplicons carrying multiple genes may be strongly selected against if increased gene dosage carries fitness costs or if costs of replicating additional DNA are substantial. In this case, we might expect to see small amplicons replacing large amplicons over time. The most common amplicon type found in this study measured 16.4 kb and spanned just 3 genes. This was the second smallest of the 15 amplicons found. Furthermore, the microsatellite data suggest that this amplicon has spread extremely rapidly in this population because all copies of this amplicon are identical or very similar for microsatellite markers flanking *pfmdr1* (fig. 2A). We suggest that longitudinal sampling of amplicons from a single location and laboratory competition experiments to inves-

tigate fitness costs would be fruitful for further investigation of copy number evolution.

The microsatellite data suggest that the dynamics of drug-resistance genes determined by tandem duplications may be very different from those determined by point mutations. We find that chromosomes bearing multiple copies of *pfmdr1* often have identical (or closely related) haplotypes to those showing no copy number amplification (fig. 2A). The simplest explanation for this pattern is that there is loss as well as gain of amplicons, resulting in frequent reversion to single-copy forms. Such loss of repeats in tandem arrays can readily occur through unequal crossing over during meiosis. These data suggest that equilibrium frequencies of amplification within populations will be determined by a balance of forces. New origins and selection due to drug treatment will maintain or increase copy number. On the other hand, amplification will be lost during meiosis as a consequence of unequal recombination, whereas fitness costs of amplification may select against parasites bearing amplified *pfmdr1* in the absence of drug pressure. CQ selection may also reduce frequencies of amplified *pfmdr1* alleles because amplification is negatively associated with drug response and results in deamplification and reversion

←

The genes in this region are PFE1130w hypothetical protein (HP), PFE1135w HP, PFE1140c G10 protein, PFE1145w HP, PFE1150w *pfmdr1*, PFE1155c mitochondrial processing peptidase alpha subunit, PFE1160w HP, PFE1165c HP, PFE1170c HP, PFE1173c outer arm dynein lc3, PFE1175w HP, PFE1180c HP, PFE1185w transporter, and PFE1190c HP. Amplicons from Mawker-Thai range from 16 to 49 kb in size, whereas amplicons from B8 (South America) and Dd2 (Southeast Asia) are 95 and 81 kb, respectively (data not shown). The different colored bars indicate whether these chromosomal regions are closely related from our analysis of flanking microsatellites (orange, clade a; red, clade b; green, clade c; purple, clade d; and blue, clade e).

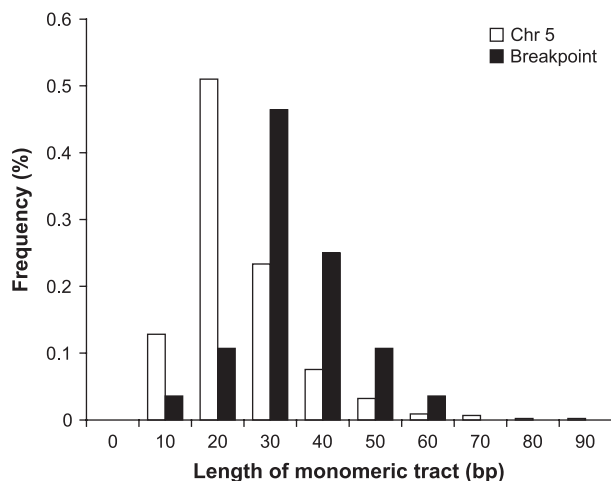


FIG. 3.—Break points occur preferentially in long monomeric A/T tracts. Frequency distribution of monomeric A or T tract sizes ($n = 439$) in the 100-kb region on either side of *pfmdr1* (black bars) compared with frequency distribution of tracts found in amplicon break points ($n = 28$) (open bars). The comparison is almost identical when results from the whole genome are plotted.

to single copy in laboratory selection experiments (Barnes et al. 1992; Peel et al. 1994). We note that CQ is still the first-line drug for treatment of *Plasmodium vivax* on the Thailand–Burma border and is widely used for treatment of *P. falciparum* in Burma, so exposure of coinfecting *P. falciparum* malaria to this drug occurs frequently.

Mechanism of Duplication and Amplification

The sequences of chromosome break points provide information about probable mechanisms of gene amplifica-

tion. Gene amplification in both prokaryotes and eukaryotes generally involves 2 steps: duplication followed by amplification. The initial gene duplication step is rate limiting and can occur by a number of processes. Homologous recombination generally involves large regions of homology and involves RecA in prokaryotes or RAD52 in eukaryotes. However, “illegitimate” recombination can also occur without involvement of these proteins, using only short regions of homology (5–35 bp). The break point sequences we found in *P. falciparum* invariably contained monomeric tracts of A or T nucleotides (9–45 bp). Hence, the initial duplication step most probably results from “illegitimate” recombination involving regions of short sequence homology. The second stage, amplification, can then occur by homologous (RAD52 dependent) recombination using the extensive regions of homology generated by the initial duplication event. The initial duplications could occur during mitotic division in the blood stream or during meiosis in the mosquito midgut. Two lines of argument suggest duplication during blood-stage growth: 1) We found tandem duplications of an 81-kb amplicon in the parasites Dd2 and in W2^{mef} from which it was derived (Oduola et al. 1988). In this case, duplications were selected by drug selection during asexual propagation in the laboratory. The similarity between laboratory and field amplification events is consistent with initial duplication occurring during mitosis in blood-stage forms. 2) Duplication is also more likely during the blood stage because there may be 10^{12} parasites within a single infected patient.

The monomeric A/T tracts found in break point junctions were significantly longer than the genome-wide average (or the average within 100 kb from *pfmdr1*), suggesting that large tracts are preferentially involved in amplification (fig. 3). This feature of the data is also consistent with

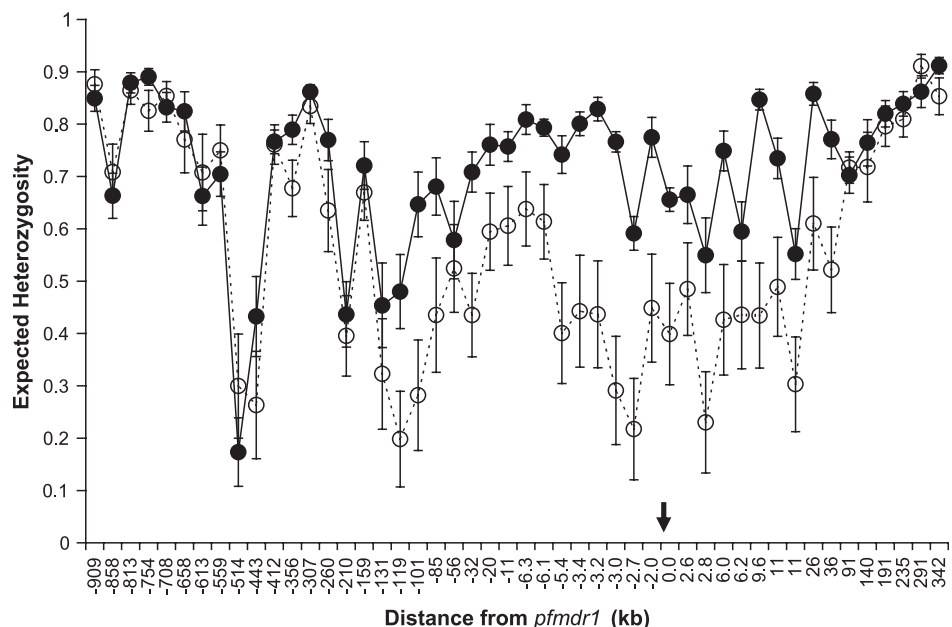


FIG. 4.—Reduced diversity around chromosomes containing multiple copies of *pfmdr1*. The distance of each marker from *pfmdr1* is shown on the x axis, which is not shown to scale. Unfilled circles, H_e plotted for chromosomes with amplified *pfmdr1*. Filled circles, chromosomes with a single copy of *pfmdr1*. The error bars show 95% confidence intervals around H_e . There is significant reduction in diversity for 131–159 kb to the 5' end and 36–91 kb to the 3' end of *pfmdr1*.

illegitimate recombination, which is strongly dependent on the size and sequence similarity of short stretches of sequence homology (Albertini et al. 1983; Whoriskey et al. 1987). The preferential involvement of long monomeric tracts in break points also suggests that parasites may have varying duplication rates depending on the length and purity of mononucleotide tracts surrounding particular genes. This would provide a simple mechanistic explanation for differences in duplication rates between isolates.

The same break point sequences are found in multiple amplicons, suggesting preferential use of particular break point sequences in duplication events or recombination between different amplicons. Parallel evolution of identical break points has also been observed in experimental systems. Reams and Neidle (2004a) characterized break point sequences in 72 of 104 amplicons generated by an in vitro selection procedure in *Acinetobacter*. They found tandem amplifications of between 12 and 290 kb, with the most commonly used break point sequence evolving independently in 6 isolates. Similarly, Whoriskey et al. (1987) found identical break point junctions in 6 of 30 independent amplicons in *E. coli*. In 3 cases, we were unable to amplify across unique junction sequences despite considerable effort. We are currently using alternative methods to investigate whether these amplicons have been translocated to alternative locations in the genome.

Implications for Malaria Control and Biology

Our results have both negative and positive implications for management of antimalarial resistance. 1) In many countries, artemisinin combination therapy (ACT) is being introduced in an attempt to stall the spread of resistance (White 1999). However, because copy number amplification has multiple independent origins under monotherapy, we expect that ACT will prove less effective in preventing drug-resistance mutations involving copy number change than for point mutations. Suppose, ACT reduces the rate of emergence of drug resistant mutants tenfold relative to monotherapy. If resistant mutants emerge under monotherapy within 5 years, then use of ACT may extend the useful life of this drug to 50 years. On the other hand, if resistance mutations arise within 6 months under monotherapy, then ACT will only extend the life of a drug to 5 years. 2) A striking feature of the data is that prevalence of multiple clone *pfmdr1* is quite low (33%) even in the face of sustained drug pressure over 15–20 years. This is different from the situation for *pfert* and *dhfr*, in which resistance alleles rapidly spread to fixation. The conventional explanation for the low prevalence of amplification is that addition of artemisinin derivatives to mefloquine monotherapy in 1994 has stalled the spread of resistance allele (Nosten et al. 2000). Frequent loss of copy number amplification suggests a viable alternative explanation. Copy number amplification may be self-limiting because reversion to single-copy state occurs continuously by unequal recombination. As a consequence, the frequency of amplification may be a balance between spread due to drug selection and loss during meiotic cell division or due to adverse effects on fitness. Although it will be hard to prevent new origins of copy number change, we predict that *pfmdr1* amplification is un-

likely to spread to high frequencies and therefore may not severely impact treatment programs.

These data also have more fundamental implications for our understanding of malaria parasite genome evolution. Triglia et al. (1991) first suggested that monomeric tracts of A or T might be involved in duplication events based on sequencing of B8 break points. Our data strongly support this hypothesis. We found that break points invariably contain monomeric A/T tracts in all amplicons examined and that long A/T tracts are preferentially used. There are 37,177 monomeric tracts of A or T >10 bp in the 23-Mb genome sequence, giving an average of 1 monomeric tract every ~600 bp. We therefore expect that copy number changes will occur frequently across the genome and have been underestimated as drivers of phenotypic evolution in *P. falciparum*. Furthermore, because copy number changes have a high mutation rate and are reversible (Romero and Palacios 1997), they may allow rapid adaptation to selection within infections. Limited data from comparative genomic hybridization (CGH) studies support the assertion that copy number changes and rearrangements are widespread. For example, Kidgell et al. (2006) describe multiple genome regions that are overrepresented in CGH experiments. These include genes involved in folate synthesis, surface antigens, and cyclophilins in addition to chromosome 5 region containing *pfmdr1*. Similarly, comparison of just 3 isolates (HB3, Dd2, and 106/1) revealed at least 6 genome regions containing copy number variation (Ferdig M, Patel J, unpublished data). We note that copy number changes have previously been detected during drug selection experiments (Inselburg et al. 1987), that chromosome size differs between isolates (Sinnis and Wellems 1988), and that duplications and translocations have been observed during analysis of a genetic cross (Hinterberg et al. 1994). Systematic mapping and phenotypic characterization of copy number changes would usefully complement current efforts to generate genome-wide SNP maps for *P. falciparum* and should be a research priority.

Supplementary Material

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

Acknowledgments

Supported by NIH RO1 AI48071 to Tim JC Anderson. This investigation was conducted in facilities constructed with support from Research Facilities Improvement Program Grant Number C06 RR013556 from the National Center for Research Resources, NIH. The Shoklo Malaria Research Unit is part of the Wellcome Trust–Mahidol University–Oxford Tropical Medicine Research Program supported by the Wellcome Trust of Great Britain. F.N. is a Wellcome Trust Senior Clinical Fellow.

Literature Cited

Albertini AM, Hofer M, Calos MP, Tlsty TD, Miller JH. 1983. Analysis of spontaneous deletions and gene amplification in

- the lac region of *Escherichia coli*. Cold Spring Harbor Symp Quant Biol. 47 (Pt 2):841–850.
- Anderson TJ. 2004. Mapping drug resistance genes in *Plasmodium falciparum* by genome-wide association. Curr Drug Targets Infect Disord. 4:65–78.
- Anderson TJ, Haubold B, Williams JT, et al. 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol Biol Evol. 17:1467–1482.
- Anderson TJ, Nair S, Qiu WG, Singlam S, Brockman A, Paiphun L, Nosten F. 2005. Are transporter genes other than the chloroquine resistance locus (*pfcr1*) and multi drug resistance gene (*pfmdr1*) associated with antimalarial drug resistance? Antimicrob Agents Chemother. 272:1153–1161.
- Anderson TJ, Roper C. 2005. The origins and spread of antimalarial drug resistance: lessons for policy makers. Acta Trop. 94:269–280.
- Anderson RP, Roth JR. 1977. Tandem genetic duplications in phage and bacteria. Annu Rev Microbiol. 31:473–505.
- Anderson TJ, Su XZ, Roddam A, Day KP. 2000. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. Mol Ecol. 9:1599–1608.
- Barnes DA, Foote SJ, Galatis D, Kemp DJ, Cowman AF. 1992. Selection for high-level chloroquine resistance results in deamplification of the *pfmdr1* gene and increased sensitivity to mefloquine in *Plasmodium falciparum*. EMBO J. 11:3067–3075.
- Chaiyaroj SC, Buranakiti A, Angkasekwinai P, Looressuwan S, Cowman AF. 1999. Analysis of mefloquine resistance and amplification of *pfmdr1* in multidrug-resistant *Plasmodium falciparum* isolates from Thailand. Am J Trop Med Hyg. 61:780–783.
- Cheng Z, Ventura M, She X, et al. (12 co-authors). 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature. 437:88–93.
- Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet. 21:673–682.
- Cowman AF, Galatis D, Thompson JK. 1994. Selection for mefloquine resistance in *Plasmodium falciparum* is linked to amplification of the *pfmdr1* gene and cross-resistance to halofantrine and quinine. Proc Natl Acad Sci USA. 91:1143–1147.
- Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3:e314.
- Dieringer D, Schlotterer C. 2002. Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol. 3:167–169.
- Dijmde A, Doumbo OK, Cortese JF, et al. (13 co-authors). 2001. A molecular marker for chloroquine-resistant *falciparum* malaria. N Engl J Med. 344:257–263.
- Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 99:16144–16149.
- Duraisingh MT, Cowman AF. 2005. Contribution of the *pfmdr1* gene to antimalarial drug-resistance. Acta Trop. 94:181–190.
- Fidock DA, Nomura T, Talley AK, 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 Cell. 6:861–871.
- Foote SJ, Thompson JK, Cowman AF, Kemp DJ. 1989. Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of *P. falciparum*. Cell. 57:921–930.
- Gardner MJ, Hall N, Fung E, et al. (45 co-authors). 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 419:498–511.
- Gonzalez E, Kulkarni H, Bolivar H, et al. (22 co-authors). 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 307:1434–1440.
- Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, Jallow M, Kimber M, McVean G, Mott R, Kwiatkowski DP. 2006. Screening for recently selected alleles by analysis of human haplotype similarity. Am J Hum Genet. 78:153–159.
- Hastings IM. 2004. The origins of antimalarial drug resistance. Trends Parasitol. 20:512–518.
- Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 169:2335–2352.
- Hinterberg K, Mattei D, Welles TE, Scherf A. 1994. Interchromosomal exchange of a large subtelomeric segment in a *Plasmodium falciparum* cross. EMBO J. 13:4174–4180.
- Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci USA. 101:10667–10672.
- Inoue K, Lupski JR. 2002. Molecular mechanisms for genomic disorders. Annu Rev Genomics Hum Genet. 3:199–242.
- Inselburg J, Bzik DJ, Horii T. 1987. Pyrimethamine resistant *Plasmodium falciparum*: overproduction of dihydrofolate reductase by a gene duplication. Mol Biochem Parasitol. 26:121–134.
- Joy DA, Feng X, Mu J, et al. (12 co-authors). 2003. Early origin and recent expansion of *Plasmodium falciparum*. Science. 300:318–321.
- Kidgell C, Volkman SK, Daily J, et al. (14 co-authors). 2006. A systematic map of genetic variation in *Plasmodium falciparum*. PLoS Pathog. 2:e57.
- Nair S, Brockman A, Paiphun L, Nosten F, Anderson TJ. 2002. Rapid genotyping of loci involved in antifolate drug resistance in *Plasmodium falciparum* by primer extension. Int J Parasitol. 32:852–858.
- Nair S, Williams JT, Brockman A, 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, Nair S, Mayxay M, Newton P, Guthmann JP, Nosten F, Anderson TJ. 2005. Selection strength and hitchhiking around two antimalarial resistance genes. Proc Biol Sci. 272:1153–1161.
- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, McGrath SD, Rocchi M, Eichler EE. 2005. A genome-wide survey of structural variation between human and chimpanzee. Genome Res. 15:1344–1356.
- Nielsen R. 2005. Molecular signatures of natural selection. Annu Rev Genet. 39:218.
- Nosten F, van Vugt M, Price R, Luxemburger C, Thway KL, Brockman A, McGready R, ter Kuile F, Looareesuwan S, White NJ. 2000. Effects of artesunate-mefloquine combination on incidence of *Plasmodium falciparum* malaria and mefloquine resistance in western Thailand: a prospective study. Lancet. 356:297–302.
- Oduola AM, Milhous WK, Weatherly NF, Bowdre JH, Desjardins RE. 1988. *Plasmodium falciparum*: induction of resistance to mefloquine in cloned strains by continuous drug exposure in vitro. Exp Parasitol. 67:354–360.
- Paterson AH, Bowers JE, Van de PY, Vandepoele K. 2005. Ancient duplication of cereal genomes. New Phytol. 165:658–661.
- Peel SA, Bright P, Yount B, Handy J, Baric RS. 1994. A strong association between mefloquine and halofantrine resistance and amplification, overexpression, and mutation in the P-glycoprotein gene homolog (*pfmdr1*) of *Plasmodium falciparum* in vitro. Am J Trop Med Hyg. 51:648–658.

- Pennings PS, Hermisson J. 2006a. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol.* 23:1076–1084.
- Pennings PS, Hermisson J. 2006b. Soft sweeps III—the signature of positive selection from recurrent mutation. *PLOS Genet.* in press
- Plowe CV, Kublin JG, Doumbo OK. 1998. *P. falciparum* dihydrofolate reductase and dihydropteroate synthase mutations: epidemiology and role in clinical resistance to antifolates. *Drug Resist Updat.* 1:389–396.
- Pollinger JP, Bustamante CD, Flidel-Alon A, Schmutz S, Gray MM, Wayne RK. 2005. Selective sweep mapping of genes with large phenotypic effects. *Genome Res.* 15:1809–1819.
- Price RN, Uhlemann AC, Brockman A, et al. (12 co-authors). 2004. Mefloquine resistance in *Plasmodium falciparum* and increased pfmdr1 gene copy number. *Lancet.* 364:438–447.
- Price RN, Uhlemann A-C, van Vugt M, et al. (12 co-authors). 2006. Molecular and pharmacological determinants of the therapeutic response to artemether-lumefantrine in multi-drug resistant *falciparum* malaria. *Clin Infect Dis.* 42:1570–1577.
- Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evol Int J Org Evol.* 59:2312–2323.
- Raymond M, Berticat C, Weill M, Pasteur N, Chevillon C. 2001. Insecticide resistance in the mosquito *Culex pipiens*: what have we learned about adaptation? *Genetica.* 112–113: 287–296.
- Reams AB, Neidle EL. 2003. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol Microbiol.* 47:1291–1304.
- Reams AB, Neidle EL. 2004a. Gene amplification involves site-specific short homology-independent illegitimate recombination in *Acinetobacter* sp. strain ADP1. *J Mol Biol.* 338: 643–656.
- Reams AB, Neidle EL. 2004b. Selection for gene clustering by tandem duplication. *Annu Rev Microbiol.* 58:119–142.
- Reed MB, Saliba KJ, Caruana SR, Kirk K, Cowman AF. 2000. Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature.* 403:906–909.
- Repping S, van Daalen SK, Brown LG, et al. (11 co-authors). 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet.* 38:463–467.
- Riehle MM, Bennett AF, Long AD. 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc Natl Acad Sci USA.* 98:525–530.
- Romero D, Palacios R. 1997. Gene amplification and genomic plasticity in prokaryotes. *Annu Rev Genet.* 31:91–111.
- Roper C, Pearce R, Nair S, Sharp B, Nosten F, Anderson TJ. 2004. Intercontinental spread of pyrimethamine resistant malaria. *Science.* 305:1124.
- Sabeti PC, Reich DE, Higgins JM, et al. (17 co-authors). 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature.* 419:832–837.
- Schlotterer C. 2003. Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends Genet.* 19:32–38.
- Sharp AJ, Locke DP, McGrath SD, et al. (14 co-authors). 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet.* 77:78–88.
- Sidhu AB, Uhlemann AC, Valderramos SG, Valderramos JC, Krishna S, Fidock DA. 2006. Decreasing pfmdr1 copy number in *Plasmodium falciparum* malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J Infect Dis.* 194:528–535.
- Sidhu AB, Valderramos SG, Fidock DA. 2005. pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in *Plasmodium falciparum*. *Mol Microbiol.* 57:913–926.
- Sinnis P, Wellems TE. 1988. Long-range restriction maps of *Plasmodium falciparum* chromosomes: crossingover and size variation among geographically distant isolates. *Genomics.* 3:287–295.
- Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.* 23:23–35.
- Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol.* 14:671–688.
- Su XZ, Wootton JC. 2004. Genetic mapping in the human malaria parasite *Plasmodium falciparum*. *Mol Microbiol.* 53:1573–1582.
- Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* 16:702–712.
- Toomajian C, Ajioka RS, Jorde LB, Kushner JP, Kreitman M. 2003. A method for detecting recent selection in the human genome from allele age estimates. *Genetics.* 165:287–297.
- Townson JR, Barcellos LF, Nibbs RJ. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol.* 32:3016–3026.
- Triglia T, Foote SJ, Kemp DJ, Cowman AF. 1991. Amplification of the multidrug resistance gene pfmdr1 in *Plasmodium falciparum* has arisen as multiple independent events. *Mol Cell Biol.* 11:5244–5250.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- White NJ. 1999. Delaying antimalarial drug resistance with combination chemotherapy. *Parassitologia.* 41:301–308.
- Whoriskey SK, Nghiem VH, Leong PM, Masson JM, Miller JH. 1987. Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. *Genes Dev.* 1:227–237.
- Wilson CM, Serrano AE, Wasley A, Bogenschutz MP, Shankar AH, Wirth DF. 1989. Amplification of a gene related to mammalian *mdr* genes in drug-resistant *Plasmodium falciparum*. *Science.* 244:1184–1186.
- Wilson CM, Volkman SK, Thaithong S, Martin RK, Kyle DE, Milhous WK, Wirth DF. 1993. Amplification of pfmdr1 associated with mefloquine and halofantrine resistance in *Plasmodium falciparum* from Thailand. *Mol Biochem Parasitol.* 57:151–160.
- Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, Magill AJ, Su XZ. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature.* 418: 320–323.

Sarah Tishkoff, Associate Editor

Accepted November 20, 2006