Published in final edited form as:

Exp Parasitol. 2007 March ; 115(3): 233–241.

***Plasmodium falciparum*: A novel method for analyzing haplotypes in mixed infections**

Laura K. Certain and Carol H. Sibley

University of Washington, Seattle, Department of Genome Sciences

Abstract

Studying the population genetics of *Plasmodium falciparum* is necessary for understanding the spread of drug resistance. However, these studies are hampered by the inability to determine haplotypes from patient samples that contain multiple parasite populations. Therefore, we have developed a method for separating for genetic analysis the individual strains in a mixed infection. We amplified a 6 kb region of chromosome 4, including the dihydrofolate reductase gene and upstream microsatellite markers. This PCR product was inserted by recombination into a gapped yeast shuttle plasmid containing both selectable and counter-selectable markers. Because each plasmid contains only one insert and each yeast colony contains only one plasmid, the individual strains are now separate. We analyzed mixtures of 3D7, K1, and Dd2 DNA and correctly identified a haplotype in each case.

Keywords

Plasmodium falciparum; Malaria; Microsatellites; Drug resistance; Dihydrofolate reductase

Introduction

In order to prevent the selection of resistance to new antimalarial drugs and the spread of resistance to those in current use, we must first understand how resistance develops and spreads. In particular, we must determine whether resistance emerges once and then spreads over a large geographic area, or whether resistance emerges many times, on many genetic backgrounds.

This question can be investigated by studying the population genetics of *Plasmodium falciparum*. By constructing haplotypes from linked genetic markers, researchers can infer the evolutionary history of a resistance allele. Such studies have indicated that resistance to both chloroquine and sulfadoxine-pyrimethamine has emerged in Southeast Asia and subsequently spread to Africa (Nair et al., 2003; Roper et al., 2004; Roper et al., 2003; Wootton et al., 2002). Surprisingly, only one origin of resistance to sulfadoxine-pyrimethamine was identified for Asia and parts of East Africa (Roper et al., 2004), though a recent study found evidence of multiple origins in Kenya (McCollum et al., 2006).

In areas of low malaria transmission most patients are infected with only one haploid parasite; that is, the infections are mainly clonal. However, in many areas where evolutionary studies would be of interest, a large proportion of patients carry multiple parasites. Current methods

Corresponding author: Carol H. Sibley, Box 357730, University of Washington, Seattle WA 98195, Ph: 206-685-9378, Fax: 206-685-7301, E-mail: sibley@u.washington.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

require that each genetic marker be typed individually. Therefore, if the initial DNA sample is mixed one cannot easily construct unambiguous haplotypes. One way to construct haplotypes from a mixed infection would be to adapt the patient sample to culture and then dilute the culture to create clonal cultures, but such an option is extremely labor-intensive. Alternatively, one could apply statistical methods to predict the most likely haplotypes, but most available methods were created with human (diploid) data in mind (Stephens et al., 2003; Stephens et al., 2001), and may not extend to mixed samples containing more than two strains of malaria. Recently, a statistical method was developed for mixed infections of malaria, but it was designed for use with pyrosequencing of bi-allelic SNPs, and may not work as well for multi-allelic microsatellite data. (Takala et al., 2006) For these reasons, most workers have simply excluded from analysis any sample that appears to be polyclonal. This approach precludes efficient use of samples from many of the most interesting endemic regions.

We present a method for constructing haplotypes from mixed samples of *P. falciparum* DNA. Using basic molecular biology - PCR, yeast transformation, sequencing - it is possible to isolate the linked markers in a mixed sample of *P. falciparum* DNA into separate yeast colonies, thus generating a valid haplotype for that sample. Though we present the method as it applies to the dihydrofolate reductase gene (*dhfr*) and its flanking nearby microsatellite genetic markers, the method is applicable to any gene and any type of genetic marker.

Methods

The overall method is described in Figure 1. In brief, the protocol uses PCR to amplify a 6 kb region containing *dhfr* and upstream microsatellite markers. These PCR products are co-transformed with a plasmid vector into yeast, and resulting colonies are streaked or pinned onto new plates. DNA extracted from the resulting colonies is then ready for analysis at the microsatellites and *dhfr*. The following paragraphs explain the protocol in detail.

DNA Preparation and Long-range PCR

As described in Figure 1, the first step of the protocol is to amplify a fragment containing *dhfr* and upstream microsatellites (or any target region). The template for this PCR was *P. falciparum* DNA extracted from cultured strains of 3D7, K1, and Dd2 parasites. For 3D7 and K1, we used the QIAamp DNA Mini Kit from Qiagen (Valencia, CA) to extract the DNA. We followed the printed protocol, with the exception of using inversion rather than vortexing to mix, in order to minimize shearing of the DNA (Sakihama et al., 2001). Dd2 DNA was extracted from cultured parasites using 15% (w/v) saponin to lyse the red blood cells, followed by the DNeasy kit from Qiagen (Valencia, CA; John White, personal communication). DNA concentration was determined by a spectrophotometer.

Primers for this initial PCR amplification are in Supplementary Table 1. Either of the two pairs listed may be used, with no appreciable difference in output. The longer pair has a more extensive region of homology with the plasmid vector. We obtained primers from Invitrogen (Carlsbad, CA) and used two different polymerase enzymes, KOD XL Polymerase (Novagen, Madison, WI) and FailSafe (Epicentre, Madison, WI), both of which amplified the target fragment well. For KOD XL, each 20 μ l reaction contained 1x KOD XL buffer, 0.2 mM each dNTP, 0.3 μ M forward primer, 0.3 μ M reverse primer, 1 unit KOD XL Polymerase, and approximately 300 ng template DNA. For FailSafe, each 20 μ l reaction contained 1x Buffer A, 1 μ M forward primer, 1 μ M reverse primer, 1 unit FailSafe Enzyme Mix, and approximately 300 ng template DNA. Using an adequate amount of template DNA is essential. Cycling parameters for both enzymes were as follows: 94 °C for 2 minutes; 30 cycles of 94 °C for 30 seconds, 50 °C for 45 seconds, 60 °C for 6 minutes; 60 °C for 5 minutes. Products were confirmed using a 0.7% agarose gel.

Transformation

The plasmid used in the transformations was pRSU, derived from pRS424 (Christianson et al., 1992). pRS424 is a shuttle vector that can be used in either yeast or *E. coli* and contains the following markers: 2 micron sequence that allows autonomous replication in yeast, TRP1 marker for selection in yeast, the T1 ori and pMB1 ori, for replication in *E. coli*, and the β -lactamase gene to confer ampicillin resistance in *E. coli*. To create pRSU, we amplified the URA3 gene from yeast and inserted it into the multiple cloning site in pRS424 using the Sal I and Sma I sites. The resulting plasmid is 6911 basepairs, and is depicted in Supplementary Figure 1.

Prior to transformation into yeast, pRSU was cut with Bam H1 to stimulate gap repair of the double-strand break region. The 5' ends of the primers used in the longrange PCR were designed with homology to pRSU upstream of URA3 and downstream of the Bam H1 site. 5-Fluoro-orotate (FOA) is toxic to yeast cells that can metabolize uracil, so transformants that contain the intact plasmid cannot grow on plates containing FOA. With appropriate integration the PCR product should replace URA3, thus rendering the host cell able to grow without tryptophan and not susceptible to poisoning by FOA (Hua et al., 1997; Jansen et al., 2005; Oldenburg et al., 1997; Raymond et al., 2002).

For the transformations, any yeast strain that lacks the ability to produce its own tryptophan and uracil is appropriate as a host strain; we used strain BB14-3a (*MATa bar1 his6 leu2-3,112 trp1-289 ura3-52*). We used standard protocols for lithium acetate transformations in yeast (Gietz et al., 1992; Gietz et al., 1995; Schiestl et al., 1989), growing cultures at 30 °C in YEPD to an OD₆₆₀ of 0.4-1.3 ($0.5-3 \times 10^7$ cells/ml). Each transformation used 10^8 cells (Gietz et al., 1992), 50µg salmon testes DNA (Sigma, St. Louis, MO), 300-900 ng linear plasmid vector, and 1.5-3µg PCR product (3-6µl). For a negative control, we used the product from a no-template PCR in a transformation. We plated cells onto standard media that lacked tryptophan and contained FOA (1 g/L) and uracil (0.05 g/L).

Streaking or Pinning

After three days of growth at 30 °C, colonies were either streaked or pinned onto new plates and returned to 30 °C for 2-4 days. Pinning was done manually with a platinum wire. This step was necessary to remove residual, unincorporated PCR product from the surface of the cells and the plate, so that it did not interfere with the microsatellite analysis (Josh Veatch, personal communication).

DNA Extraction and Analysis

To extract the DNA from the yeast colonies, we suspended each colony in 40µl of 20 mM NaOH and incubated at 97 °C for 10 minutes. The resulting mixture was used immediately or stored at -20 °C for up to four months. For microsatellite analysis, we followed the protocol of Nair *et al.* (2003). Each 10µl reaction contained 2 mM MgCl₂, 0.2 mM dNTPs, 0.4µM forward (labeled) primer, 0.4µM reverse primer, 0.15 units Taq polymerase, and 1µl DNA. (Supplementary Table 1 lists primer sequences for amplifying two microsatellites within the 6 kb fragment, S780 and S784.) Cycling conditions were: 94 °C for 2 minutes; 25 cycles of 94 °C for 30 seconds, 45 °C for 30 seconds, 60 °C for 30 seconds; 60 °C for 2 minutes. (Nair et al., 2003) A nested PCR was not necessary. After PCR amplification, we diluted the products 20-fold prior to analysis on a MegaBACE 1000 DNA Analysis System (Amersham, Piscataway, NJ). Product length was determined by inspection using Genetic Profiler (Amersham, Piscataway, NJ). Product length at a given microsatellite locus for a single yeast colony was recorded as the allele at that locus for that colony. The combination of lengths (alleles) at the two microsatellite loci was the haplotype for that colony. The most common haplotype among all colonies from a particular *P. falciparum* sample was the haplotype for

that sample. For determining the *dhfr* genotype, we used the PCR/RFLP method of Duraisingh (Duraisingh et al., 1998). On the rare occasions when two alleles were detected at a single locus, those alleles were not recorded.

Results

We tested our method by examining two microsatellites upstream of *dhfr*, one immediately upstream (S784; chr.4 position 755,005-755,056; *dhfr* begins at 755,069) and the other 3.8 kb upstream (S780; chr.4, position 751,282-751,329). These correspond to loci d_104_5 and d_100_8 in Nair *et al.* (2003). Both are dinucleotide repeats, (AT)_n. We followed the protocol outlined above, using DNA extracted from cultured 3D7, K1, and Dd2 parasites. DNA from one, two, or all three strains of *P. falciparum* was added in equal amounts to the long-range PCR as the template. By starting with DNA from only one strain of *P. falciparum* (3D7, K1, or Dd2) in the initial long-range PCR, we simulated how this method would behave for clonal samples (i.e. samples from patients infected with only one strain of *P. falciparum*); by starting with DNA from multiple strains we simulated mixed infections.

Results at a Single Locus

At the S780 locus, we determined alleles (microsatellite lengths) for 96 3D7 colonies, 83 K1 colonies, and 97 Dd2 colonies (Table 1). The distribution of alleles for each sample at S780 is shown in Figure 2a-c. When we started with DNA from only one strain, the majority of the colonies contained an allele of the expected size. However, as is routinely seen in analysis directly from genomic DNA, we did see larger and smaller sizes at each microsatellite, presumably due to polymerase slippage (stutter) in the initial long-range PCR (Figure 3; see Discussion). Colonies derived from K1 gave the tightest distribution of lengths and those from Dd2 gave the broadest. To verify that this variation was not due to replication in yeast, we pinned 36 colonies a second time, extracted DNA from the new colonies, and compared the alleles (microsatellite lengths) to those observed in the original colonies. Thirteen of the colonies showed a clear peak in fluorescence at a specific length on the capillary gel readout at one or both loci for both pinnings (first and second); in all cases the allele from the new colony was identical to that from the parent colony. Therefore, the size variation is not due to instability of the microsatellite during propagation in yeast.

To determine whether instability was a problem for other types of genetic markers, we genotyped the SNPs within the *dhfr* coding region in 26 colonies (Duraisingh et al., 1998) (Table 2). The twelve colonies from samples that contained only one type of *P. falciparum* DNA in the initial long-range PCR all showed the correct *dhfr* genotype. For the 14 from mixed samples, all had a genotype that matched one of the component strains. Therefore, SNPs are more stable than microsatellites in this system.

The overall goal of the method is to allow analysis of samples that contain more than one parasite genotype. To test this aspect of the method, we mixed equal quantities of purified DNA from the different strains prior to the initial long-range PCR. Figure 2d-f shows the analysis of the S780 microsatellite for paired mixtures of the strains, and Figure 2g shows the result when DNA from all three strains was combined. Although the initial long-range PCR started with the same amount of DNA from each strain, far more colonies were derived from one strain in each case. With the three strains that we used, there was a strong bias in favor of colonies derived from the K1 template, and colonies derived from Dd2 were observed far less often than expected. Though alleles derived from each input strain are present, this extreme bias underscores the conclusion that one cannot use this approach to infer the relative proportions of strains in the input sample.

Results at Both Loci - Constructing Haplotypes

We determined haplotypes (lengths for both loci) for 83 3D7 colonies, 71 K1 colonies, and 67 Dd2 colonies. Because some colonies derived from each DNA sample had the true allele and some had stutter alleles, each sample produced a range of haplotypes. For example, the haplotypes found for colonies derived from K1 DNA are shown in Supplemental Figure 2. The most frequently observed haplotype was 196/178, the true haplotype. However, there were 13 other haplotypes found in at least one colony. They were various combinations of the correct allele at one locus paired with a stutter allele at the other locus, or combinations of stutter alleles at both loci. The first and second most common haplotypes for each sample are given in Table 3.

We constructed haplotypes for the 276 mixed-sample colonies that showed a clear peak in fluorescence at a specific length for both loci. In the mixtures, the most common haplotype observed among the colonies was always the correct haplotype for one of the strains. The second most common haplotypes were stutter versions of the most common haplotypes. For example, in the K1 and Dd2 mixture, the most common haplotype was 196/178 (K1), but the second most common was not 210/178 (Dd2). Rather, it was 194/178, a stutter version of K1. Therefore, we can clearly identify a haplotype from a mixed sample; however, minor haplotypes are not informative. Moreover, the dominant haplotype does not necessarily represent the predominant strain in the sample, because there is a bias in the detection of some microsatellite alleles over others.

To compare our results to the standard protocol used by Nair *et al.* (2003), we used nested PCR on DNA samples from the Malaria Research and Reference Reagent Resource Center (MR4) to determine alleles for 3D7, K1, Dd2, W2, and V1/S *P. falciparum* strains. We also examined mixtures of 3D7 and K1 DNA, 3D7 and Dd2 DNA, K1 and Dd2 DNA, and 3D7, K1, and Dd2 DNA. In all cases the alleles for the MR4 samples matched the allele sizes we found using our method (Supplementary Table 2). In the mixtures, Dd2 had a weaker fluorescent peak than 3D7 or K1 (data not shown), mirroring what we saw using yeast. Therefore, this skew seems not to derive from our method, but rather to be inherent in the DNA sequence. Also of note, though the V1/S haplotype matches that seen for the triple mutant in Nair *et al.* (2003), the Dd2 and W2 haplotypes do not. V1/S has four mutations in *dhfr*; Dd2 and W2 have three (Dd2 is derived from W2). All three strains came from Southeast Asia.

Efficiency of Method

Because all of the colonies that grow on the FOA selection plates have lost the URA3 gene, each should contain our 6 kb insert, but not all of them did. The fraction of colonies that contained the *P. falciparum* insert is shown in Table 1. We analyzed 3259 colonies at one microsatellite (S780) and 3279 colonies at a second microsatellite (S784); we analyzed 3024 colonies at both loci. However, of the 3259 analyzed at S780, only 687 had a clear peak in fluorescence at a specific length; the remainder were blank (or had a peak that was unclear or double). That is, only 687 of 3259 colonies that grew on the selective plates clearly contained the insert. The corresponding proportion at S784 was 807/3279. Of the 3024 colonies analyzed at both loci, 497 had a clear peak in fluorescence at a specific length for both loci. The average success rate across transformations was 26% for S780 and 30% for S784, with standard deviations of 19% and 22%, respectively. These numbers (26%, 30%) differ from those in Table 1 (21%, 25%), because we did not weight the average to reflect the number of colonies analyzed per transformation. The plates from the control transformations performed with the PCR reaction that contained no template DNA had some colonies; however, none of those analyzed (S780: n=118; S784: n=112) contained either microsatellite.

Because analyzing microsatellite length on a capillary sequencer is expensive, we explored ways to reduce cost by screening for colonies that did have an insert before analyzing the microsatellites. One simple way to screen is to attempt to amplify *dhfr* from the colonies. For one set of transformations (38 colonies), we both did the normal microsatellite analysis (using a capillary sequencer) and determined the presence of *dhfr* by attempting to amplify it with PCR and then visualizing on an agarose gel. Of the 20 colonies for which *dhfr* amplified (band visible on the gel), 19 also had a clear peak in fluorescence at the microsatellite loci (Supplementary Table 3). However, many colonies that gave no visible band on the gel *also* had a clear peak in fluorescence, presumably because the microsatellite analysis is more sensitive. To address this lack of sensitivity, one could do a nested PCR for amplifying *dhfr*, or amplify with fluorescent primers and visualize with an appropriate UV camera system. Alternatively, one could avoid the capillary sequencer altogether and determine microsatellite length by using a high-percentage agarose gel (Mwangi et al., 2006).

Translation to the Field

This method will only be useful if it works for samples collected in the field. The most problematic step is likely to be the initial long-range PCR, because a successful amplification of such a long fragment requires a large amount of template DNA. To test the applicability of this protocol to field samples, we attempted to amplify the 6 kb fragment from DNA extracted from three different field sample types: 2 ml of whole blood from patients in Mae Sot, Thailand; 25µl cultured parasites (50% hematocrit, ~9% parasitemia) spotted onto filter paper; ~100µl blood from patients in Sri Lanka spotted onto filter paper (Hapuarachchi et al., 2006). Using the same PCR protocol as described in the methods, we successfully amplified the 6 kb fragment from 4/4 whole blood samples, 4/4 mock filter paper samples, and 1/4 Sri Lankan filter paper samples. We did not make any attempt to optimize the protocol for field samples, to use a nested PCR, or to amplify a fragment shorter than 6 kb. However, using a nested protocol would increase the inherent noise in the system due to polymerase slippage, thus reducing efficiency further.

Discussion

Prior haplotype analyses have been limited to monoclonal samples. Since transmission is high in many areas of Africa and Asia, polyclonal samples are the rule rather than the exception. These are also areas where the opportunity for recombination between different genotypes is highest, and it is of great interest to determine the relatedness of parasites from these areas, as well. With this in mind, we have designed a method that will allow haplotype analysis of multiclonal samples.

We have demonstrated that our method can determine a valid haplotype for mixed samples of *P. falciparum* DNA. Though this method can determine only one haplotype per sample, it marks an improvement over current methods, which exclude mixed infections from analysis. With our method, one can use all samples from areas of intense transmission, rather than discarding most of them. However, because some alleles are amplified preferentially over others, it will be problematic to get accurate prevalence data for a particular haplotype. Rather, one can simply determine presence or absence of haplotypes.

The microsatellites in *P. falciparum* are almost all dinucleotide repeats. As a result, when one types a microsatellite in the capillary sequencer, in addition to the main peak (at the expected allele length) there are flanking peaks at lengths two bases off from the true length (Figure 3a). This stutter is routinely observed due to polymerase slippage during the PCR amplification of the microsatellites. Because the polymerase in the initial long-range PCR also slips, these small size differences are generated and then are stably maintained when the DNA is incorporated into the yeast plasmid. As a consequence, a few of the yeast colonies produced from a single

strain of *P. falciparum* are products of stutter and show allele lengths slightly different from the parent strain. Figure 3b shows that the distribution of alleles among the yeast colonies reflects the shape of the stutter peaks observed in the microsatellite analysis of DNA from an individual colony. This effect does not preclude accurate identification of a correct haplotype, but multiple colonies need to be analyzed. SNPs or trinucleotide repeats would show less instability, and consequently would require analysis of fewer colonies. None were present in the genomic region we examined, but future studies using this method could use trinucleotide repeats or SNPs.

It is important to emphasize that none of the methods that depend on standard PCR amplification can determine the relative proportions of different alleles in a sample. There is a strong bias toward better amplification of particular alleles, usually the smallest of the size classes, and this was amply demonstrated in our experiments. Regardless of whether we analyzed the samples in the yeast system or analyzed them directly using the standard protocol, we found that the smaller K1 allele always amplified better than Dd2, even if we began with equal amounts of each. Because we simply used a spectrophotometer to determine DNA concentration, it is possible that this skew reflects slight differences in the amount of input DNA. However, the dominance of K1 appeared across many DNA preparations and in experiments using DNA from MR4. Therefore, we believe this skew is not simply a reflection of the proportion of the two haplotypes in the starting templates.

A concern during both the long-range PCR (Tanabe et al., 2002) and the homologous recombination in yeast is that there could be crossover between two different strains of DNA, resulting in a plasmid that contained DNA from both strains. From the data in Table 2 one could postulate that some crossover occurred; 6/14 samples had one or two microsatellite alleles that did not match the *dhfr* coding region genotype. In general, the switch from one strain to another occurred between the two microsatellites (the longest stretch between loci). However, in most cases one cannot distinguish between crossover and stutter. For example, one of the 3D7/K1 samples shows the K1 allele at all loci except S784. This haplotype is due either to double-crossover or to stutter at the S784 locus. Though the latter explanation is far more likely, one cannot exclude crossover as the cause. However, crossover is a relatively rare event; therefore, by recording only the most common haplotype for each sample we can be confident of avoiding false haplotypes.

One drawback to this method is its low efficiency; many of the colonies did not contain the target insert. A restriction digest of plasmid DNA extracted from these false-positive colonies indicated that these colonies had lost the URA3 gene without gaining the target insert (the long-range PCR product). Because the no-template-DNA control PCR product produced some (empty) colonies, we suspected that the yeast cells were using primer-dimers for gap repair rather than the 6 kb insert. Therefore, we redesigned the vector so that it had homology only to regions of *P. falciparum* DNA contained at the ends of the PCR product, not within the primers themselves. However, using this plasmid (pBDGR) instead of our original plasmid (pRSU) produced ten-fold fewer colonies for each transformation (~14 vs. ~140 per transformation), and the yield of true positive colonies was similar to or slightly less than with pRSU. We tried numerous other variations on the protocol to increase efficiency (see Supplemental Materials), but none led to an improvement over the protocol described in this paper.

Another challenge to this protocol, particularly in applying it to samples collected in the field, is the large amount of DNA required for the initial long-range PCR. Though we were able to amplify the 6 kb fragment from blood spotted onto filter paper, the PCR was more successful for DNA extracted from whole blood. One could certainly attempt to optimize the PCR protocol for blood spotted onto filter paper, but it is likely that there would always be some filter paper

samples that would fail to amplify. However, for many studies a fragment shorter than 6 kb would be sufficient. For example, the microsatellites within 100 bp upstream and 500 bp downstream of the *dhfr* coding region distinguish the triple mutant allele of *dhfr* that is thought to originate in Asia from other alleles that appear to have arisen in Africa (McCollum et al., 2006; Roper et al., 2004; Roper et al., 2003). Amplification from a filter paper sample of a fragment that would include both of these markers would be far easier, and would still be informative for many questions of interest. Thus, filter paper samples are likely to be amenable to this kind of analysis in many situations.

In summary, we have developed a method that allows determination of a valid haplotype from mixed samples of *P. falciparum* DNA. Though we tested the method using microsatellites and *dhfr*, it could be used for any type of genetic marker, and any target region of a genome. In fact, it would perform much better for SNPs or trinucleotide repeats, because they are more stable in PCR than dinucleotide microsatellites, and therefore false “stutter” alleles would be no problem or far less so. An obvious question is whether or not the method will work on field samples. We believe that it will, as we have successfully amplified the 6 kb fragment from dried blood samples taken in the field, and that is the step most prone to failure. We hope that the method will allow study of the population genetics of *P. falciparum* in regions that have hitherto been neglected due to the high proportion of mixed infections.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank: M.K. Raghuraman and Sandra Pennington for their advice on yeast strains and transformations; John White and Kasey Rivas for providing *P. falciparum* DNA and parasites; Joelle Thomas and Ann Riddle for assistance with the protocol; Conner Sandefur for making pBDGR; Lori Schoenfeld for the protocol for extracting DNA from yeast colonies; and Tim Anderson and Shalini Nair for valuable discussions and instruction on microsatellites in *P. falciparum*.

Abbreviations

DNA, Deoxyribonucleic acid; PCR, Polymerase chain reaction; FOA, 5-fluoro-orotate; DHFR, Dihydrofolate reductase.

BIBLIOGRAPHY

- Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* 1992;110:119–22. [PubMed: 1544568]
- Duraisingh MT, Curtis J, Warhurst DC. Plasmodium falciparum: detection of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes by PCR and restriction digestion. *Experimental Parasitology* 1998;89:1–8. [PubMed: 9603482]
- Gietz D, St Jean A, Woods RA, Schiestl RH. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Research* 1992;20:1425. [PubMed: 1561104]
- Gietz RD, Schiestl RH, Willems AR, Woods RA. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* 1995;11:355–60. [PubMed: 7785336]
- Hapuarachchi HC, Dayanath MY, Bandara KB, Abeyesundara S, Abeyewickreme W, de Silva NR, Hunt SY, Sibley CH. Point mutations in the dihydrofolate reductase and dihydropteroate synthase genes of Plasmodium falciparum and resistance to sulfadoxine-pyrimethamine in Sri Lanka. *American Journal of Tropical Medicine and Hygiene* 2006;74:198–204. [PubMed: 16474070]
- Hua SB, Qiu M, Chan E, Zhu L, Luo Y. Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. *Plasmid* 1997;38:91–6. [PubMed: 9339466]

- Jansen G, Wu C, Schade B, Thomas DY, Whiteway M. Drag&Drop cloning in yeast. *Gene* 2005;344:43–51. [PubMed: 15656971]
- McCollum AM, Poe AC, Hamel M, Huber C, Zhou Z, Shi YP, Ouma P, Vulule J, Bloland P, Slutsker L, Barnwell JW, Udhayakumar V, Escalante AA. Antifolate Resistance in *Plasmodium falciparum*: Multiple Origins and Identification of Novel dhfr Alleles. *Journal of Infectious Diseases* 2006;194:189–97. [PubMed: 16779725]
- Mwangi JM, Omar SA, Ranford-Cartwright LC. Comparison of microsatellite and antigen-coding loci for differentiating recrudescing *Plasmodium falciparum* infections from reinfections in Kenya. *International Journal of Parasitology* 2006;36:329–36. [PubMed: 16442537]
- Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, Guthmann JP, Smithuis FM, Hien TT, White NJ, Nosten F, Anderson TJ. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Molecular Biology Evolution* 2003;20:1526–36.
- Oldenburg KR, Vo KT, Michaelis S, Paddon C. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Research* 1997;25:451–2. [PubMed: 9016579]
- Raymond CK, Sims EH, Olson MV. Linker-mediated recombinational subcloning of large DNA fragments using yeast. *Genome Research* 2002;12:190–7. [PubMed: 11779844]
- Roper C, Pearce R, Nair S, Sharp B, Nosten F, Anderson T. Intercontinental spread of pyrimethamine-resistant malaria. *Science* 2004;305:1124. [PubMed: 15326348]
- Roper C, Pearce R, Bredenkamp B, Gumede J, Drakeley C, Mosha F, Chandramohan D, Sharp B. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet* 2003;361:1174–81. [PubMed: 12686039]
- Sakihama N, Mitamura T, Kaneko A, Horii T, Tanabe K. Long PCR amplification of *Plasmodium falciparum* DNA extracted from filter paper blots. *Experimental Parasitology* 2001;97:50–4. [PubMed: 11207114]
- Schiessler RH, Gietz RD. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Current Genetics* 1989;16:339–46. [PubMed: 2692852]
- Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 2003;73:1162–9. [PubMed: 14574645]
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 2001;68:978–89. [PubMed: 11254454]
- Takala SL, Smith DL, Stine OC, Coulibaly D, Thera MA, Doumbo OK, Plowe CV. A high-throughput method for quantifying alleles and haplotypes of the malaria vaccine candidate *Plasmodium falciparum* merozoite surface protein-1 19 kDa. *Malaria Journal* 2006;5:31. [PubMed: 16626494]
- Tanabe K, Sakihama N, Farnert A, Rooth I, Bjorkman A, Walliker D, Ranford-Cartwright L. In vitro recombination during PCR of *Plasmodium falciparum* DNA: a potential pitfall in molecular population genetic analysis. *Molecular and Biochemical Parasitology* 2002;122:211–6. [PubMed: 12106875]
- Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, Magill AJ, Su XZ. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 2002;418:320–3. [PubMed: 12124623]

1. Use PCR to amplify a 6 kb fragment containing *dhfr* and flanking microsatellites.

2. Co-transform the PCR product and a plasmid vector into yeast.

3. Pin colonies onto a new plate.

4. Extract DNA from each colony for microsatellite analysis.

Figure 1.

A schematic of the method. The first step is to amplify a 6 kb fragment containing *dhfr* and upstream microsatellite markers (shapes). In step two, the PCR products are co-transformed with a cut plasmid vector into yeast and plated on selective media containing 5-fluoro-orotate (FOA). In step three, colonies are pinned onto a new plate to remove residual, unincorporated PCR product. The last step is to extract DNA from the colonies and analyze it at the genetic loci. The different colors (black, gray, speckled in the non-color version) indicate the different genotypes/haplotypes of *P. falciparum* present in the sample. The white colonies are those that grow on the selective plates but do not contain the insert (i.e. false positives).

Figure 2. Distribution of alleles (lengths) at locus S780 in yeast colonies containing the 6 kb fragment amplified from genomic DNA of the indicated strains (a-g). Arrows indicate the size of true alleles.

Figure 3.

a. Sample capillary gel readout of analysis at locus S784 of DNA purified from a single yeast colony containing the 6 kb fragment amplified from K1 DNA; the large grey rectangle indicates the allele called by Genetic Profiler. b. Distribution of alleles (lengths) at locus S784 in all yeast colonies containing the 6 kb fragment amplified from genomic K1 DNA.

Table 1

Proportion of colonies containing the insert, as indicated by a clear peak in fluorescence at a specific length on the capillary gel readout for each microsatellite locus.

Template	Colonies analyzed		Colonies with a clear peak in fluorescence (%)		Analyzed at both loci	Clear peak at both loci (%)
	S780	S784	S780	S784		
3D7	329	350	96	(29)	314	83
K1	237	240	83	(35)	233	71
Dd2	617	606	97	(16)	529	67
3D7/K1	524	520	164	(31)	499	130
3D7/Dd2	482	496	51	(11)	441	37
K1/Dd2	454	458	63	(14)	434	37
3D7/K1/Dd2	616	609	133	(22)	574	72
Total	3259	3279	687	(21)	3024	497
H2O	118	112	0	(0)	0	(0)

Table 2

Haplotypes and DHFR genotypes for a subset of colonies.

Template DNA	S780	Alleles (Length in bp)	S784	51	DHFR Genotype (Codons)	108
K1	196		180	N	R	N
K1	196		176	N	R	N
K1	196		178	N	R	N
K1	196		178	N	R	N
3D7	202		174	N	C	S
3D7	196		174	N	C	S
3D7	200		174	N	C	S
3D7	202		174	N	C	S
Dd2	210		178	I	R	N
Dd2	210		178	I	R	N
Dd2	210		176	I	R	N
Dd2	208		178	I	R	N
3D7K1	202		174	N	C	S
3D7K1	202		174	N	C	S
3D7K1	196		178	N	C	S
3D7K1	196		174	N	R	N
3D7K1	202		178	N	R	N
3D7K1	202		178	N	R	N
3D7K1	196		180	N	R	N
3D7Dd2	196		172	N	C	S
3D7Dd2	202		174	N	C	S
3D7Dd2	202		172	N	C	S
3D7Dd2	204		174	N	C	S
K1Dd2	210		176	I	R	N
K1Dd2	210		178	N	R	N
3D7K1Dd2	210		180	I	R	N

Table 3

First and second most common haplotypes for each sample.

Sample	S780/S784	Most common haplotype Frequency	%	S780/S784	Second most common haplotype Frequency	%
3D7	202/174	27/83	33%	200/174	15/83	18%
K1	196/178	37/71	52%	192/178;194/178	6/71	8%
Dd2	210/178	19/67	28%	208/178	11/67	16%
3D7/K1	196/178	35/130	27%	196/174	11/130	8%
3D7/Dd2	202/174	15/37	41%	202/178;204/174	3/37	8%
K1/Dd2	196/178	17/37	46%	194/178	5/37	14%
3D7/K1/Dd2	196/178	27/72	38%	196/176	11/72	15%