Published in final edited form as:

Infect Genet Evol. 2006 September ; 6(5): 417–424.

Genetic diversity in the Block 2 region of the merozoite surface protein 1 (MSP-1) of *Plasmodium falciparum*: Additional complexity and selection and convergence in fragment size polymorphism

S.L. Takala^{a,b,*}, A.A. Escalante^c, O.H. Branch^{a,g}, S. Kariuki^d, S. Biswas^e, S.C. Chaiyaroj^f, and A.A. Lal^a

^aDivision of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, GA, USA ^bCenter for Vaccine Development, University of Maryland School of Medicine, Health Sciences Facility 1, Room 480, 685 W. Baltimore St., Baltimore, MD 21201, USA ^cSchool of Life Sciences, Arizona State University, AZ, USA ^dVector Biology and Control Research Center, Kenya Medical Research Institute, Kisumu, Kenya ^eMalaria Research Center, Indian Council of Medical Research, New Delhi, India ^fDepartment of Microbiology, Faculty of Science, Mahidol University, Bangkok, Thailand ^gDepartment of Geographic Medicine, University of Alabama at Birmingham, AL, USA

Abstract

Fragment size in the Block 2 repetitive region of merozoite surface protein 1 (MSP1) has commonly been used as a molecular marker in studies of malaria transmission dynamics and host immunity in *Plasmodium falciparum* malaria. In this study, we further explore the genetic variation in MSP-1 Block 2 underlying potential problems faced while studying the immune responses elicited by this vaccine target and while using it as a molecular marker in epidemiologic investigations. We describe the distribution of a new Block 2 recombinant allele family in samples collected from western Kenya and other malarious regions of the world and provide evidence that this allele family is found worldwide and that all MR alleles most likely originated from a single recombination event. We test whether the number of tandem repeats (i.e. fragment size) can be considered neutral in an area of high transmission in western Kenya. In addition, we investigate the validity of the assumption that Block 2 alleles of the same size and allele family are identical by examining MSP1 Block 2 amino acid sequences obtained from full-length MSP-1 clones generated from infected Kenyan children and find that this assumption does not hold. We conclude that the worldwide presence of a new allele family, the effect of positive natural selection, and the lack of conserved amino acid motifs within alleles of the same size suggest a higher level of complexity that may hamper our ability to elucidate allele family specific immune responses elicited by this vaccine target and its overall use as genetic marker in other types of epidemiologic investigations.

Keywords

Merozoite protein; Genetic diversity; *Plasmodium*; Molecular epidemiology

* Corresponding author at: Center for Vaccine Development, University of Maryland School of Medicine, Health Sciences Facility 1, Room 480, 685 W. Baltimore St., Baltimore, MD 21201, USA. Tel.: +1 410 706 5328; fax: +1 410 706 1204. E-mail address: stakala@medicine.umaryland.edu (S.L. Takala).

1. Introduction

Malaria has resurged in many tropical countries after it was thought to be on its way to eradication during the 1960s. Multiple factors are contributing to the global worsening of the malaria situation, among them, insecticide resistance in the mosquito, drug resistance in the parasite, and lack of funding and commitment in endemic countries to invest appropriate resources needed for prevention and control. At present, malaria is a major public health problem resulting in approximately 300 million clinical cases and over one million deaths per year worldwide (WHO, 2002). The most severe form of the disease is caused by *Plasmodium falciparum*, which is accountable for most malaria morbidity and almost all malaria mortality. A successful malaria control and prevention program will require the coordinated use of several strategies, including potential malaria vaccines. However, extensive genetic diversity in *Plasmodium* presents a challenge to the effectiveness of such strategies, as well as to studies aimed at understanding the development of natural immunity and complexity of infection as it relates to protection and pathogenesis.

The merozoite surface protein 1 (MSP1) is a leading vaccine candidate antigen. It is the most abundant surface protein on the blood stage of *P. falciparum*, and it is thought to play a role in erythrocyte invasion (Holder et al., 1992). The gene has been subdivided into several blocks with different degrees of genetic polymorphism (Tanabe et al., 1987; Miller et al., 1993). The Block 2 region includes repetitive motifs of three amino acids. Until recently, three allele-families had been identified in Block 2: K1, Mad20, and RO33. Alleles in K1 and Mad20 contain antigenically unique, tripeptide repeats, with extensive diversity in the number of repeats (Miller et al., 1993). RO33 lacks the tripeptide repeats observed in the other two families; however, outside Block 2, this allele is similar to the MAD-20 type (Hughes, 1992).

Fragment size in the three Block 2 allele families has commonly been used as a molecular marker in studies of malaria transmission dynamics and host immunity in *P. falciparum* malaria (Ariey et al., 1999; Da Silva et al., 1999; Farnert et al., 1999; Konate et al., 1999; Branch et al., 2001). However, we have previously shown a fourth allele family in Block 2, called MR, which results from a recombination event between alleles of the Mad20 and RO33 families (Takala et al., 2002). We postulated that this allele family arose from a single recombination event; however, samples from only one geographic region were used in the original study.

Protective immune responses have been observed against the motifs present in the major allele families of Block 2 (Polley et al., 2003; Cavanagh et al., 2004), and while the evidence suggests that the allele families are maintained by selection, it is not clear how selection operates against the number of tandem repeats (Sakihama et al., 2004).

Finally, MSP1 Block 2 fragment size analysis often makes the assumption that two alleles of the same size from the same allele family are identical by descent; that is, the two alleles have identical sizes because they share a recent common ancestor. Given that the K1 and Mad20 alleles contain different numbers of antigenically unique, tripeptide repeats, this assumption may not be valid.

The purpose of this study is to further explore the genetic variation in MSP-1 Block 2 underlying potential problems faced while studying the immune responses elicited by this vaccine target and while using it as a molecular marker in epidemiologic investigations. First, we describe the distribution of a new Block 2 recombinant allele family in samples collected from western Kenya and other malarious regions of the world. Second, we test whether the number of tandem repeats (i.e. fragment size) can be considered neutral in an area of high transmission in western Kenya. Finally, we investigate the validity of the assumption that Block 2 alleles of the same size and allele family are identical by examining Block 2 amino acid sequences obtained from full-length MSP-1 clones generated from infected Kenyan children.

2. Materials and methods

2.1. Sampling and DNA extraction

The samples used in this study came from participants in the Asembo Bay Cohort Project, a prospective, longitudinal study that took place in an area of high malaria transmission in western Kenya (Bloland et al., 1999; Branch et al., 2001). We extracted DNA from 362 parasitized blood samples taken from 45 different infants between 1992 and 1995. DNA was extracted from approximately 300 μ l infected pRBC's using the Pure-Gene extraction method (Gentra Systems). We also looked for the MR allele family in 31 samples collected from individuals with clinical malaria from each of three different locations in Thailand, Venezuela, and India. Thailand samples were collected during a hospital-based study in Bangkok, isolates from India were collected by the MRC in Delhi, and Venezuelan isolates were collected in Bolivar and Amazonas states.

2.2. Nested PCR to detect Block 2 alleles

We used a nested PCR method, as described previously (Takala et al., 2002), where the first PCR was done with the external 5' and 3' primers, which amplify Block 2 plus some of the flanking regions. The second PCR was done with allele specific primers, the 5' primer being specific for Mad20 (5'-GCTGTTACAACCTAGTACACC-3') and the 3' primer being specific for RO33 (5'-AGGATTTGCAGCACCTGGAGATCT-3'). For the PCR amplification mixture, 5 μ l of each of the appropriate 5' and 3' primers (40 ng/ μ l) were added to 12 μ l of dNTP mixture (Promega, 107 μ mol/reaction), 3 μ l of $MgCl_2$ (25 mM), 10 μ l of PCR 10 $\times$ buffer containing 15 mM $MgCl_2$ (Perkin-Elmer), 0.5 μ l Taq polymerase (GIBCO), and 59.5 μ l double distilled water. For the first PCR, 5 μ l of DNA (adjusted to a concentration of 150 ng/ μ l) was added to the 95 μ l reaction mixture, and amplified for 25 cycles at a 55 $^{\circ}C$ annealing temperature. For each internal PCR, 5 μ l of the first (external) PCR product was added to the 95 μ l reaction mixture and amplified for 30 cycles at a 64 $^{\circ}C$ annealing temperature. Each internal PCR product was electrophoresed on a 3% Agarose-1000 gel (GIBCO) to determine the alleles present. We could consistently detect length differences of 10 bp. The methods used to detect non-recombinant K1, Mad20, and RO33 were the same as those used to detect recombinant MR, only using forward and reverse internal primers specific to each allele family (Branch et al., 2001).

2.3. Sequencing of Block 2 alleles

We sequenced products obtained from the recombinant PCR representing a range of sizes and band intensities. Products were cut from 3% Agarose-1000 gels using a clean sterile scalpel, and the DNA was purified using the Qiaquick Gel Extraction Kit (Qiagen). The purified products were sequenced on an ABI automated sequencer. It is possible that parasites could have MR alleles with the same fragment size but different nucleotide sequences, and in that case, the cut fragment could contain more than one parasite genotype; however, the sequence did not suggest mixed genotypes. Sequence analysis of the Block 2 region from 29 full-length MSP-1 clones was used to test the validity of the assumption of conserved amino acid motifs among Block 2 alleles of the same size and allele family. These clones were generated, as part of a different study, from samples collected from infected Asembo Bay infants during 1992–1993. Five K1 alleles with the same MSP1 Block 2 fragment size and 24 Mad20 alleles with the same Block 2 fragment size were examined.

2.4. Statistical analysis

We tested whether the observed distribution of Block 2 fragment length polymorphisms is neutral. We used the Ewens–Watterson test (Watterson and Perlow, 1978) that is based on Ewens sampling theory of neutral alleles (Ewens, 1972). This test calculates the sum of squared

allele frequencies (F), and tests it versus its expected value derived from a null distribution generated by simulating 1000 neutral samples following the method developed by Stewart in 1977. We also calculate the exact version of this test as developed by Slatkin (1994, 1996). These tests were performed using the program Arlequin version 2.00 (Schneider et al., 2000).

3. Results

3.1. Distribution of Block 2 alleles

We found that the different allele families were not evenly represented in the 362 Kenyan samples included in this study. The most common allele family was K1, with 95% of the samples tested containing at least one K1 allele. The R033 family was found in 79% of the samples, while MAD20 was found in 72% of the samples. The MR allele family was found in 29% of the samples under study. We also detected the MR allele family in three additional localities including Thailand (three samples out of 31), Venezuela (two samples out of 31) and India (four samples out of 31).

Fig. 1 shows the number of alleles detected for each allele family within the Kenyan samples. The K1 allele family is the most diverse with 20 different alleles, compared to 15 from the MAD-20 allele family. The new allele family MR also showed polymorphisms with seven different alleles.

In order to determine if the MR family indeed originated from a single recombination event we sequenced 40 alleles from Kenya, and two from each of the other geographic locations. The sequences of unique alleles from each geographic location are available in GenBank under the accession numbers AF462449–AF462456 and AY826427–AY826431. An alignment with alleles from the four localities is shown in Fig. 2. The recombination event occurs exactly in the same position for all the alleles sequenced. We found two synonymous substitutions: one in a 150 bp allele from Kenya and the other in a 150 bp allele from Thailand.

3.2. Tests for selection

We tested if the observed distributions of the fragment size polymorphism in K1, MAD20, and MR alleles fit the expected distribution under neutrality using the Ewens–Watterson (EW) and the Ewens–Watterson–Slatkin (EWS) tests. The results of these tests are summarized in Table 1. First we analyzed K1 and MAD20 separately since no evidence has been found that these two allele families can recombine in Block 2. Both distributions rejected the null hypothesis that they can be explained as neutral allele samples. In the case of K1 the observed value of F is 0.078 while the expected is 0.212 rejecting the null hypothesis with $p < 0.0001$ for both tests. In the case of MAD20 the observed value of F is 0.119 and the expected is 0.244, the p -value for the EW test is $p = 0.004$ and for the EWS is $p = 0.001$. We also considered the MR distribution, but we could not reject the null hypothesis of neutrality. The observed F was 0.336 and the expected was 0.389 with p -values of 0.426 and 0.280 for the EW and the EWS, respectively. It is important to notice that this allele family was less frequent so it has a smaller sample size. If we combine all the alleles, given that they are from the same locus, their combined distribution cannot be explained by neutrality with an observed F value of 0.081 and an expected F value of 0.231, the p -values for both tests are < 0.0001 . When sampling infections from the same individuals over time, the possibility exists of sampling the same parasites at sequential time points. To address this, the analysis was repeated using data only from those samples separated from one another by treatment and/or aparasitemia ($n = 185$). The results were similar and the conclusions were the same.

3.3. Comparison of alleles of the same size and allele family

In order to investigate the genetic diversity in Block 2 alleles of the same size and allele family, we examined Block 2 amino acid sequences obtained from 29 full-length MSP-1 clones generated from infected Kenyan children. Out of five K1 alleles with the same size Block 2, two different amino acid motifs were observed. Out of 24 Mad20 alleles with approximately the same size Block 2 region, three different motifs were observed. These results indicate that within allele families, motifs are not conserved within alleles of the same fragment size. These differences are at the amino acid level (i.e. not just single base-pair changes). Examples of the lack of conserved amino acid motifs in alleles of the same size are depicted in Fig. 3. The sequences shown in Fig. 3 are available in Genbank under accession numbers DQ377133–DQ377137.

4. Discussion

In this study we show that diversity in Block 2 is underestimated by only genotyping three allele families as there is a fourth recombinant allele family that is distributed worldwide. We also demonstrate that the distribution of fragment size polymorphism in MSP-1 Block 2 cannot be explained by neutrality. Finally, we observe that within allele families, amino acid motifs are not conserved within alleles of the same fragment size. These results suggest a higher level of complexity that may hamper our ability to elucidate allele family specific immune responses elicited by this vaccine target and its overall use as genetic marker in other types of epidemiologic investigations.

In order to test whether the fragment size polymorphism in Block 2 is under selection, we genotyped DNA from 362 samples collected from parasitemic Kenyan children. The number of observed alleles is higher than those reported for other populations such as Tanzania (Babiker et al., 1994, 1997) with 10 K1 and five Mad20 alleles detected, Sudan (Babiker et al., 1997) with four K1 and four Mad20 alleles detected, Papua New Guinea (Paul et al., 1995) with four K1 and four Mad20 alleles detected, and Thailand (Paul et al., 1998) with three K1 and four Mad20 alleles detected. Although this overall pattern is consistent with other malarial antigens (Escalante et al., 2001, 2002), it is important to note that several of these studies used PCR followed by hybridization with allele-specific probes to genotype Block 2 (Babiker et al., 1994, 1997; Paul et al., 1995, 1998). This technique yields products that range in size from 400 to 600 bp in length. As a result, it is possible that some unique alleles were not recognized because it is difficult to differentiate tripeptide size differences in products this large. We also observed more alleles than studies that used a similar nested PCR technique detecting smaller product sizes. Jelinek et al. (1999) detected six K1 alleles and one Mad20 allele in Uganda; Snounou et al. (2000) detected four K1 alleles and five Mad20 alleles in Thailand; and Robert et al. (1996) detected 10 K1 alleles and 5 Mad20 alleles in Senegal. These data support the idea that it is difficult to make comparisons between populations whose genotyping was performed in different laboratories, using different protocols with varying sensitivities. This fact was noted by Farnert et al. (2001), who conducted a study of the comparability of genotyping results from different laboratories and emphasized the need for standardization of genotyping methods. The inability to compare results across different populations and laboratories poses an additional problem with using Block 2 as a reliable molecular marker.

The role of genetic recombination in generating the observed polymorphisms in *P. falciparum* has been a focus of controversy for more than 10 years (Tibayrenc et al., 1991; Babiker et al., 1997; Rich et al., 1997; Conway et al., 1999; Escalante et al., 2001). Studies examining this issue have focused mostly on antigens that are targets of potential antimalarial vaccines. Unfortunately, because these genes are under positive natural selection, it is difficult to differentiate the effect of selection from recombination using point mutations (McCutchan et al., 1992; Escalante et al., 1998). Nevertheless, intragenic recombination has been implicated

as a major factor to explain the observed polymorphisms in MSP-1 (Conway and McBride, 1991; Jongwutiwes et al., 1991; Hughes, 1992; Qari et al., 1998; Conway et al., 1999; Sakihama et al., 1999). Polymorphisms in tandem repeat regions, such as Block 2, have often been considered evidence of recombination (McCutchan et al., 1988); however, alternative hypotheses such as mitotic intragenic recombination have been raised for MSP-2 (Irion et al., 1997) and CSP (Rich et al., 1997). Indeed, other investigations suggest that the number of repeats appears to be independent of the number of recombination events and that the number of repeats is relatively high compared with the number of documented recombination events in the surrounding area (Sakihama et al., 2004). Nevertheless, our finding of a recombinant in Block 2 of MSP-1 between the Mad20 and RO33 allele families supports the notion that sexual intragenic recombination is an important factor in the evolution of genetic diversity in this repetitive region. This kind of event cannot be clouded by the action of natural selection and cannot be explained by mitotic intragenic changes.

In a previous study, we have described the possible origins of the MR allele family, and postulated that this family arose from a single recombination event (Takala et al., 2002). In this study, we have shown that the MR family is found worldwide and that the recombination site occurs at the same position in MR alleles from all the localities tested. These data support the hypothesis that the MR family derives from a single recombination event. The alternative hypothesis that this recombinant appears de novo at different locations exactly in the same place is a less parsimonious explanation. The worldwide distribution of the MR allele family as well as the observed allelic polymorphisms in this family suggest that this recombination event could be as old as the population expansion of *P. falciparum* outside of Africa. However, our data do not allow us to quantitate how often these recombination events take place in a given population. Such measures demand different kinds of studies (Conway et al., 1999; Su et al., 1999); however, we are providing evidence that this process has generated a new allele family in Block 2 of MSP1. Further investigation is also needed to understand the processes responsible for the expansion and spread of the MR family in the population.

Our analysis has also shown that the number of tandem repeats in Block 2 appears to be under positive natural selection. The EW and EWS tests are capable of detecting departure from neutrality due to the action of balancing selection or advantageous alleles. We found evidence that the fragment size polymorphisms in the K1 and MAD20 allele families are not neutral. The tests failed to detect departure from neutrality in the case of MR, even though the distribution of fragment sizes had a similar shape to that of K1 and M20. A plausible explanation for this result is reduced statistical power due to the low frequency of MR alleles. A larger study would be required to apply the tests under conditions comparable to the other allele families.

Our results provide new insight about genetic polymorphisms that complement those of Conway (Conway et al., 2000), who used Wright's F_{st} test to demonstrate that the frequency of the Block 2 allele families is under selection. Although clear evidence was provided that natural selection acts on the frequency of the allele families themselves (MAD20-like and K1-like, respectively), no evidence has been reported on how the fragment size polymorphism is maintained within each allele family. Immune responses are usually considered the driving selective force acting on antigens; however, it is not clear how the number of repeats could modulate the quality of the immune response. If there is not a clear phenotype that can be recognized by the putative selective force then we cannot convincingly argue for positive natural selection (Escalante et al., 2004). An alternative explanation is that the number of repeats could affect the stability of the protein itself. Such an alternative has been explored in CSP where simulations demonstrated that structure could explain the observed frequency of fragment sizes (Escalante et al., 2002).

Previous studies have made the observation that fragment size polymorphisms of antigens are not good markers for certain types of molecular epidemiological studies (e.g. studies of evolutionary history) since they may have their own dynamics due to the action of selection (Anderson et al., 2000). We have shown that at least in the Kenyan population included in this study, the fragment size polymorphisms in Block 2 are not neutral. These results support the use of other markers, such as microsatellites, to conduct such studies. The role played by natural selection, if any, in the fixation and worldwide distribution of the MR allele family demands further investigation.

We have also demonstrated that within allele families, alleles of the same size may have different amino acid motifs. The frequency of this phenomenon cannot be accurately determined from this study as the number of alleles examined is small. However, this result has significant implications for studies using Block 2 fragment length polymorphism as a marker of diversity. For example, studies examining the association between Block 2 alleles and immunity may not observe a correlation between fragment size and immune response because correlations may be masked by one fragment size containing a mixture of repeats. In addition, studies that rely on allele frequencies (e.g. studies of selection, linkage disequilibrium, or population structure) may also be affected by grouping together alleles based on fragment size that are actually different from each other.

In summary, we have shown that the new allele family, MR, is present worldwide, and that it appears to be the product of a single recombination event. This evidence suggests that sexual intragenic recombination is a mechanism that can generate new genetic variants of malarial antigens, and that it should be considered in the study of genes encoding vaccine and drug targets. We have also shown that the fragment size polymorphisms in Block 2 of MSP-1 are under positive natural selection. Finally, we demonstrate the amino acid sequences in alleles of the same size and allele family are not conserved. These results suggest that fragment size may not be an accurate marker for genetic diversity within MSP1 Block 2 and support the use of other markers such as microsatellites to conduct molecular epidemiological studies.

Acknowledgements

We thank the study participants from Kenya, Thailand, India, and Venezuela for their willingness to participate in the study. We also thank Amanda Poe for technical support. S. Takala was supported by the Emerging Infectious Diseases Fellowship Program administered through APhL and CDC. A.A. Escalante was supported by the grant NIH R01 GM60740.

References

- Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* 2000;17 (10):1467–1482. [PubMed: 11018154]
- Ariey F, Chalvet W, Hommel D, Peneau C, Hulin A, Mercereau-Puijalon O, Duchemin JB, Sarthou JL, Reynes JM, Fandeur T. *Plasmodium falciparum* parasites in French Guiana: limited genetic diversity and high selfing rate. *Am J Trop Med Hyg* 1999;61 (6):978–985. [PubMed: 10674682]
- Babiker HA, Ranford-Cartwright LC, Currie D, Charlwood JD, Billingsley P, Teuscher T, Walliker D. Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitology* 1994;109 (4):413–421. [PubMed: 7800409]
- Babiker HA, Lines J, Hill WG, Walliker D. Population structure of *Plasmodium falciparum* in villages with different malaria endemicity in east Africa. *Am J Trop Med Hyg* 1997;56 (2):141–147. [PubMed: 9080871]
- Bloland PB, Ruebush TK, McCormick JB, Ayisi J, Boriga DA, Oloo AJ, Beach R, Hawley W, Lal A, Nahlen B, Udhayakumar V, Campbell CC. Longitudinal cohort study of the epidemiology of malaria

- infections in an area of intense malaria transmission I. Description of study site, general methodology, and study population. *Am J Trop Med Hyg* 1999;60 (4):635–640. [PubMed: 10348240]
- Branch OH, Takala S, Kariuki S, Nahlen BL, Kolczak M, Hawley W, Lal AA. *Plasmodium falciparum* genotypes, low complexity of infection, and resistance to subsequent malaria in participants in the Asembo Bay Cohort Project. *Infect Immun* 2001;69 (12):7783–7792. [PubMed: 11705960]
- Cavanagh DR, Doodoo D, Hviid L, Kurtzhals JA, Theander TG, Akanmori BD, Polley S, Conway DJ, Koram K, McBride JS. Antibodies to the N-terminal block 2 of *Plasmodium falciparum* merozoite surface protein 1 are associated with protection against clinical malaria. *Infect Immun* 2004;72 (11):6492–6502. [PubMed: 15501780]
- Conway DJ, McBride JS. Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology* 1991;103 (1):7–16. [PubMed: 1719468]
- Conway DJ, Roper C, Oduola AM, Arnot DE, Kremsner PG, Grobusch MP, Curtis CF, Greenwood BM. High recombination rate in natural populations of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1999;96 (8):4506–4511. [PubMed: 10200292]
- Conway DJ, Cavanagh DR, Tanabe K, Roper C, Mikes ZS, Sakihama N, Bojang KA, Oduola AMJ, Kremsner PG, Arnot DE, Greenwood BM, McBride JS. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat Med* 2000;6 (6):689–692. [PubMed: 10835687]
- Da Silveira LA, Dorta ML, Kimura EA, Katzin AM, Kawamoto F, Tanabe K, Ferreira MU. Allelic diversity and antibody recognition of *Plasmodium falciparum* merozoite surface protein 1 during hypoendemic malaria transmission in the Brazilian amazon region. *Infect Immun* 1999;67 (11):5906–5916. [PubMed: 10531247]
- Escalante AA, Lal AA, Ayala FJ. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 1998;149 (1):189–202. [PubMed: 9584096]
- Escalante AA, Grebert HM, Chaiyaroj SC, Magris M, Biswas S, Nahlen BL, Lal AA. Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. *X Asembo Bay Cohort Project Mol Biochem Parasitol* 2001;113 (2):279–287.
- Escalante AA, Grebert HM, Isea R, Goldman IF, Basco L, Magris M, Biswas S, Kariuki S, Lal AA. A study of genetic diversity in the gene encoding the circumsporozoite protein (CSP) of *Plasmodium falciparum* from different transmission areas–XVI. Asembo Bay Cohort Project *Mol Biochem Parasitol* 2002;125 (1–2):83–90.
- Escalante AA, Cornejo OE, Rojas A, Udhayakumar V, Lal AA. Assessing the effect of natural selection in malaria parasites. *Trends Parasitol* 2004;20 (8):388–395. [PubMed: 15246323]
- Ewens WJ. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 1972;3 (1):87–112. [PubMed: 4667078]
- Farnert A, Rooth I, Svensson A, Snounou G, Bjorkman A. Complexity of *Plasmodium falciparum* infections is consistent over time and protects against clinical disease in Tanzanian children. *J Infect Dis* 1999;179 (4):989–995. [PubMed: 10068596]
- Farnert A, Arez AP, Babiker HA, Beck HP, Benito A, Bjorkman A, Bruce MC, Conway DJ, Day KP, Henning L, Mercereau-Puijalon O, Ranford-Cartwright LC, Rubio JM, Snounou G, Walliker D, Zwetyenga J, do Rosario VE. Genotyping of *Plasmodium falciparum* infections by PCR: a comparative multicentre study. *Trans R Soc Trop Med Hyg* 2001;95 (2):225–232. [PubMed: 11355566]
- Holder AA, Blackman MJ, Burghaus PA, Chappel JA, Ling IT, McCallum-Deighton N, Shai S. A malaria merozoite surface protein (MSP1)-structure, processing and function. *Mem Inst Oswaldo Cruz* 1992;87 (Suppl 3):37–42. [PubMed: 1343716]
- Hughes AL. Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. *Mol Biol Evol* 1992;9 (3):381–393. [PubMed: 1584009]
- Irion A, Beck HP, Felger I. New repeat unit and hot spot of recombination in FC27-type alleles of the gene coding for *Plasmodium falciparum* merozoite surface protein 2. *Mol Biochem Parasitol* 1997;90 (1):367–370. [PubMed: 9497062]
- Jelinek T, Kilian AH, Westermeier A, Proll S, Kabagambe G, Nothdurft HD, von Sonnenburg F, Loscher T. Population structure of recrudescant *Plasmodium falciparum* isolates from western Uganda. *Trop Med Int Health* 1999;4 (7):476–480. [PubMed: 10470338]

- Jongwutiwes S, Tanabe K, Nakazawa S, Uemura H, Kanbara H. Coexistence of GP195 alleles of *Plasmodium falciparum* in a small endemic area. *Am J Trop Med Hyg* 1991;44 (3):299–305. [PubMed: 2035751]
- Konate L, Zwetyenga J, Rogier C, Bischoff E, Fontenille D, Tall A, Spiegel A, Trape JF, Mercereau-Puijalon O. Variation of *Plasmodium falciparum* msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. *Trans R Soc Trop Med Hyg* 1999;93 (Suppl 1):21–28. [PubMed: 10450422]
- McCutchan TF, de la Cruz VF, Good MF, Wellems TE. Antigenic diversity in *Plasmodium falciparum*. *Prog Allergy* 1988;41:173–192. [PubMed: 2457215]
- McCutchan TF, Lal AA, do Rosario V, Waters AP. Two types of sequence polymorphism in the circumsporozoite gene of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1992;50 (1):37–45. [PubMed: 1542315]
- Miller LH, Roberts T, Shahabuddin M, McCutchan TF. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). *Mol Biochem Parasitol* 1993;59 (1):1–14. [PubMed: 8515771]
- Paul RE, Packer MJ, Walmsley M, Lagog M, Ranford-Cartwright LC, Paru R, Day KP. Mating patterns in malaria parasite populations of Papua New Guinea. *Science* 1995;269 (5231):1709–1711. [PubMed: 7569897]
- Paul RE, Hackford I, Brockman A, Muller-Graf C, Price R, Luxemburger C, White NJ, Nosten F, Day KP. Transmission intensity and *Plasmodium falciparum* diversity on the northwestern border of Thailand. *Am J Trop Med Hyg* 1998;58 (2):195–203. [PubMed: 9502604]
- Polley SD, Tetteh KK, Cavanagh DR, Pearce RJ, Lloyd JM, Bojang KA, Okenu DM, Greenwood BM, McBride JS, Conway DJ. Repeat sequences in block 2 of *Plasmodium falciparum* merozoite surface protein 1 are targets of antibodies associated with protection from malaria. *Infect Immun* 2003;71 (4):1833–1842. [PubMed: 12654798]
- Qari SH, Shi YP, Goldman IF, Nahlen BL, Tibayrenc M, Lal AA. Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen. *Mol Biochem Parasitol* 1998;92 (2):241–252. [PubMed: 9657329]
- Rich SM, Hudson RR, Ayala FJ. *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. *Proc Natl Acad Sci USA* 1997;94 (24):13040–13045. [PubMed: 9371796]
- Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, Sarthou JL, Mercereau-Puijalon O. Extensive genetic diversity of *Plasmodium falciparum* isolates collected from patients with severe malaria in Dakar, Senegal. *Trans R Soc Trop Med Hyg* 1996;90 (6):704–711. [PubMed: 9015525]
- Sakihama N, Kimura M, Hirayama K, Kanda T, Na-Bangchang K, Jongwutiwes S, Conway D, Tanabe K. Allelic recombination and linkage disequilibrium within Msp-1 of *Plasmodium falciparum*, the malignant human malaria parasite. *Gene* 1999;230 (1):47–54. [PubMed: 10196473]
- Sakihama N, Matsuo T, Mitamura T, Horii T, Kimura M, Kawabata M, Tanabe K. Relative frequencies of polymorphisms of variation in Block 2 repeats and 5' recombinant types of *Plasmodium falciparum* msp1 alleles. *Parasitol Int* 2004;53:59–67. [PubMed: 14984836]
- Schneider, S.; Roessli, D.; Excoffier, L. Geneva, Genetics and Biometry Lab. Dept. of Anthropology, University of Geneva.; 2000. Arlequin: A software for population genetics data analysis.
- Slatkin M. An exact test for neutrality based on the Ewens sampling distribution. *Genet Res* 1994;64 (1):71–74. [PubMed: 7958833]
- Slatkin M. A correction to the exact test based on the Ewens sampling distribution. *Genet Res* 1996;68 (3):259–260. [PubMed: 9062082]
- Snounou G, Zhu X, Siripoon N, Jarra W, Thaithong S, Brown KN, Viriyakosol S. Biased distribution of msp1 and msp2 allelic variants in *Plasmodium falciparum* populations in Thailand. *Trans R Soc Trop Med Hyg* 2000;94 (1):65.
- Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, Wootton JC, Wellems TE. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* 1999;286 (5443):1351–1353. [PubMed: 10558988]
- Takala S, Branch O, Escalante AA, Kariuki S, Wootton J, Lal AA. Evidence for intragenic recombination in *Plasmodium falciparum*: identification of a novel allele family in block 2 of merozoite surface

protein-1: Asembo Bay Area Cohort Project XIV. Mol Biochem Parasitol 2002;125 (1–2):163–171. [PubMed: 12467983]

Tanabe K, Mackay M, Goman M, Scaife JG. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J Mol Biol 1987;195 (2):273–287. [PubMed: 3079521]

Tibayrenc M, Kjellberg F, Arnaud J, Oury B, Breniere SF, Darde ML, Ayala FJ. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. Proc Natl Acad Sci USA 1991;88 (12):5129–5133. [PubMed: 1675793]

Watterson GA, Perlow J. Homozygosity in three-allele populations. Theor Popul Biol 1978;13 (2):268–275. [PubMed: 694786]

WHO, Roll Back Malaria. What is Malaria?. 2002.
<http://www.who.int/inf-fs/en/InformationSheet01.pdf>

Fig. 1.
Distribution of fragment sizes for three MSP-1 Block 2 allele families in Kenyan infections:
(A) K1 alleles detected; (B) Mad20 and MR alleles detected.

Fig. 2.

Alignment of MR alleles from four geographic regions. One representative of each unique allele is included in the alignment. Dashes indicate identity with the Kenya 130 bp allele. The repeat GGTTCAGGT has been replaced with *****. The end of the Mad20-like sequence and the beginning of RO33-like sequence (i.e. the postulated recombination site) is indicated by a vertical line |. The primer sequences are indicated in bold text.

(A) K1 alleles with the same fragment size:

S599 EEITT KGASAQSGTS AQSGASQSG TSGTSGPSGP SGTSPSSRSN (3)
 704AB EEITT KGASAQSGTS GTSGTSGTSG TSGTSGPSGP SGTSPSSRSN (2)

(B) Mad20 alleles with the same fragment size:

687H AVTTSTPGSG GSVTSGGSVT SGGSGGSVAS VASVAASGGS GNSRRTNPSD (3)
 706B AVTTSTPGSG GSVTSGGSVT SGGSGGSVAS VASVA-SGGS GNSRRTNPSD (2)
 620B AVTTSTPGSG GSVTSGGSGG SVASVASGGS GGSVA-SGGS GNSRRTNPSD (19)

Fig. 3.

Amino acid motifs are not conserved within alleles of the same fragment size and allele family: (A) K1 alleles with the same fragment size but different amino acid motifs. (B) Mad20 alleles with the same fragment size but different amino acid motifs. Numbers in parentheses indicate the number of times the sequence was observed among the alleles examined in this study.

Table 1
Results of Ewens–Watterson (EW) and Ewens–Watterson–Slatkin (EWS) tests

	Observed F	Expected F	p -values [*]	
			EW	EWS
K1	0.078	0.212	<0.0001	<0.0001
Mad20	0.119	0.244	0.004	0.001
MR	0.336	0.389	0.426	0.280
All alleles	0.081	0.231	<0.0001	<0.0001

* H_0 = neutrality.