

Patterns of polymorphism in genomic regions flanking three highly polymorphic surface antigens in *Plasmodium falciparum*

Olukemi K. Amodu^{a,b,1,2}, Daniel L. Hartl^{a,1}, Scott William Roy^{a,c,*,1,2}

^a Department of Organismic and Evolutionary Biology, 16 Divinity Avenue, Harvard University, Cambridge, MA 02138, United States

^b College of Medicine, University of Ibadan, Ibadan, Nigeria

^c National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, United States

Received 25 May 2007; received in revised form 4 December 2007; accepted 10 December 2007

Available online 15 December 2007

Abstract

Many surface antigens of the human malaria parasite *Plasmodium falciparum* show extraordinary diversity, with different alleles being so divergent as to be unalignable in some coding regions. To better understand the population history and modes of selection on such loci, we sequenced genomic regions flanking the highly polymorphic genes merozoite surface protein-1, merozoite surface protein-2, and circumsporozoite protein, from reference isolates of *P. falciparum*. Diversity was much lower in genomic flanking regions than in the coding sequences. Average pairwise nucleotide diversity for these regions was 0.00088, similar to other genomic regions not thought to be evolving under balancing selection, suggesting against balancing selection acting on promoter regions of these genes. Most observed polymorphisms were singletons. A higher ratio of SNPs to indels than previously reported for *P. falciparum* was observed. An 11 bp repeat upstream of *msp2* showed an intriguing pattern of polymorphism possibly suggestive of purifying selection on total allele length.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Diversifying selection; Balancing selection; Population history; Population genetics

1. Introduction

Plasmodium falciparum, the most virulent form of human malaria, remains a leading cause of global morbidity and mortality, with some 300–500 million clinical cases and 1–2 million deaths annually, most in sub-Saharan Africa [1]. The genetic diversity of *P. falciparum* appears to be a major reason that patients only acquire cross-strain immunity after repeated infections [2]. For instance, in vivo and in vitro studies suggest that effective immunity to one strain of a given malaria parasite does not imply immunity to other distantly related strains. This lack of cross-strain immunity is likely a central reason why it has been so difficult to develop species-wide vaccines to malaria parasites.

In keeping with the importance of immunogenic diversity within *P. falciparum*, levels of intraspecific polymorphism in *P. falciparum* differ vastly among and within loci and between classes of mutations. Silent polymorphisms, thought to be nearly selectively neutral, are fairly rare or even absent in nuclear genes [3,4], in the 6 kb mitochondrial genome [5] and in introns [6]. By contrast, some regions of genes encoding proteins expressed on the parasite's surface show extensive protein-level polymorphism, probably due to balancing/diversifying selection driven by interaction with the host immune system [3,7–20]. Interestingly, highly divergent regions often lie directly adjacent to near-monomorphic regions [7,12].

Here, we studied three promising vaccine targets for which trials are currently under way. Merozoite surface protein-1 (MSP1) shows extensive polymorphism in two distinct sub-regions of its coding sequence [7]. One, termed block 2, is a long repetitive region, and exhibits three distinct and widely diverged allelic classes. The other, termed blocks 6–16, shows only two distinct and widely diverged allelic classes. While there was initially debate about the age of alleles at this locus, the recent report of an ortholog from the chimpanzee parasite

* Corresponding author at: National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, United States.
Tel.: +1 310 496 8941.

E-mail address: scottwroy@gmail.com (S.W. Roy).

¹ Tel.: +1 617 496 5540; fax: +1 617 496 5854.

² These two authors contributed equally.

Plasmodium reichenowi that clearly more closely resembles one of the two allelic types indicates a deep allelic divergence [21]. Strikingly, the 5' and 3' regions of the gene, as well as regions between the two highly divergent regions, show near monomorphism [7]. The second gene, MSP2, has a central repetitive region with two very distinct allelic classes, though these two classes are thought to have diverged much more recently than the MSP1 classes [10]. As with *Msp1*, the 5' and 3' regions show much lower levels of diversity. Finally, the circumsporozoite protein (CSP) is expressed in the sporozoite stage of parasite invasion [12]. The CSP gene shows extensive allelic diversity. Unlike MSP1 and MSP2, CSP does not show two distinct allelic classes, but like the other two genes 5' and 3' regions show much lower levels of diversity than the central portion of the gene. The great diversity of each of these genes seems to correspond to decreased cross-strain immune system reaction, and thus understanding the pattern of polymorphism is important.

We sequenced 1000–1500 bp regions flanking the three genes from a variety of *P. falciparum* lab isolates. We find that levels of polymorphism in these regions are similar to previously reported values for a variety of loci not thought to be evolving under balancing selection. Interestingly, the ratio of SNP to indel polymorphisms is lower in these regions than previously reported for a variety of loci. The distribution of polymorphisms suggests that these regions themselves are likely not under balancing selection, and that recombination between the genes' coding regions and flanking regions has homogenized these flanking regions. This pattern is in stark contrast to some previously studied cases of balancing selection.

2. Methods

2.1. *Plasmodium falciparum* strains sequenced

For all three genes we sequenced: HB3 (Honduras), 7G8 (Brazil), D6 (Sierra Leone), W2 (Laos), and 3D7 (The Netherlands). For *mSP1*, we additionally sequenced Muz12.4 (Papua New Guinea); for *mSP2* and *cSP*, we additionally sequenced D10 (Papua New Guinea). For the region upstream of *mSP1* we additionally sequenced FCB-1 (Columbia), FCR-3 (The Gambia), 106/1 (Sudan), T-994 (Thailand), D10 and RO33 (Ghana). Parasites were grown in culture for short periods using standard culture techniques. DNA was isolated as previously described [6].

2.2. Primer design, PCR amplification, cloning and sequencing

Primers (Table 1) were designed based on the published 3D7 genome sequence as obtained from PlasmoDB using Primer 3 [21]. PCR amplifications were performed using the HotStarTaq DNA Master Mix (Qiagen, CA) with 35 cycles of 2 min at 94 °C, 1 min at 57–60 °C (usually 57 °C), and 5 min at 72 °C. Single PCR bands were purified using shrimp alkaline phosphatase and exonuclease 1 (USB Biochemicals). The Invitrogen TOPO TA cloning kit was used for constructing plasmids which were

Table 1
Primers for the three *P. falciparum* genes

| Locus | Primers |
|--------------|-----------------------------|
| <i>Msp-1</i> | |
| Ends5d.f | CCTTTTCAATATATTCACCTTTGAACC |
| Ends5d.r | GCACTTTATCGAATATGAATTTACAC |
| Ends5e.f | CATTATTTTTGCATTTTACATTAGCC |
| Ends5e.r | GGGAAATTTGTGAAATTCATATTC |
| Ends3a.f | CCATCAAATTCTTATTTATTTTTCC |
| Ends3a.r | CGATGGTATTTTACTGCAGTTCC |
| Ends3b.f | CATCTCCCATAAAAACAATTTTCC |
| Ends3b.r | TAAAACAACATATGAAGAAAAGAAGAA |
| <i>Msp-2</i> | |
| Ms2_5a.f | AAAAGACACTGCAAGTTGG |
| Ms2_5a.r | TTCTTTCATTTTAAAACATTGAC |
| Ms2_5b.f | TGTGAGATTTTACAGATACATTGG |
| Ms2_5b.r | TTTGACAAATAGCTTTCGCTA |
| Ms2_a.f | CAATGTTTTAATTACCTTCATTTTGA |
| Ms2_a.r | TCGACTAATCAATTTACAATTTTCCA |
| Ms2_b.f | TTTTGTGTGTGTCATTTTGA |
| Ms2_b.r | TCCCTCGGATAATTTTTACC |
| <i>Csp</i> | |
| Cs_5b.f | CCATTTAAAAATAGAACACCGTATGA |
| Cs_5b.r | CAACTGAAACCTTTAAATGATCCTAA |
| Cs_5a.f | GGATCATTTAAAGGTTTCAGTTGTG |
| Cs_5a.r | GAATATTTAATTACGTTTTCGCGATT |
| Cs_3a.f | CGTGGTTTCTACTTATTTTTACACGA |
| Cs_3a.r | TTTTACACATGCGATTTGGAT |
| Cs_3b.f | TTTTGGTTTACCACCAATCCA |
| Cs_3b.r | TTTCCTTTTTAAAGCACGTGATAA |

purified using the Perfectprep® Plasmid 96 Vac, Direct Bind from Eppendorf. Plasmids were sequenced using the TOPO TA cloning kit primers. Products from two separate colonies were sequenced in both directions on a 3100 DNA capillary sequencer (Applied Biosystems) using Big Dye chemistry. Editing of raw chromatograms and multiple sequence alignments was carried out using the Sequencher version 4.1 software. Final alignments were carried out with the ClustalW program and alignments were adjusted by hand and analyzed by eye.

3. Results

We sequenced 6–12 strains for 1000–1500 bp of the genomic regions directly flanking the surface antigens *Msp1*, *Msp2*, and *Csp*. These sequences are available in GenBank as accession numbers DQ286461–DQ286502. The results are summarized in Table 2.

3.1. *Msp1*

We sequenced 1119 bp upstream of the *Msp1* coding sequence from 12 *P. falciparum* lab isolates from four different continents. In all, we found 3 SNPs and 13 indels. This region showed an average θ of 4.5×10^{-4} . All three of the SNPs, as well as all six of the indels which exhibited only two distinct alleles, were singletons. We also sequenced 1057 bp downstream of the *Msp1* coding sequence from six *P. falciparum* lab isolates from four different continents and found six indels but no SNPs. Two

Table 2
Observed polymorphisms in analyzed regions

| | Length (bp) | SNPs | Indels (2-allele) | Ins:Del | Singletons | | Singleton distribution | θ |
|-------------------------|-------------|------|-------------------|---------|------------|--------|------------------------|----------|
| | | | | | SNPs | Indels | | |
| Msp1 | | | | | | | | |
| 5' ^a | 1119 | 3 | 13 (6) | 2:4 | 3/3 | 6/6 | (3,2,2,1,1) | 0.00045 |
| 3' | 1057 | 0 | 6 (3) | 1:2 | – | 2/3 | (2) | 0.00000 |
| Msp2 | | | | | | | | |
| 5' | 1290 | | | | | | (5,2,2,1) | |
| Msp5 CDS | 314 | 3 | 0 (0) | – | 3/3 | – | (3) | |
| Msp5 intron | 136 | 0 | 0 (0) | – | – | – | – | 0.00000 |
| Intergenic ^b | 840 | 3 | 11 (5) | 1:5 | 3/3 | 4/5 | (2,2,2,1) | 0.00117 |
| 3' | 1116 | | | | | | (2,2,2,1) | |
| Intergenic | 655 | 2 | 3 (2) | 2:0 | 2/2 | 2/2 | (2,2) | 0.00102 |
| A.L CDS | 461 | 3 | 0 (0) | – | 3/3 | – | (2,1) | |
| Csp | | | | | | | | |
| 5' | 1081 | 5 | 5 (4) | 2:2 | 4/5 | 3/4 | (3,2,1,1) | 0.00173 |
| 3' | 1113 | 6 | 8 (5) | 1:4 | 6/6 | 5/5 | (3,3,2,2,1) | 0.00180 |
| Total intergenic | 5689 | 25 | 46 (25) | 9:17 | 24/25 | 22/25 | | 0.00088 |

n, number of different isolates sequenced; indels, all indels and, in parentheses, those in regions with only two alleles so that a single local change may confidently be inferred; ins:del, among indel sites with only two alleles, the number of cases in which the rarer allele represents an insertion, or deletion, relative to the more common one; singletons, number of polymorphisms in which one allele is represented by a single sequence in the sample; singleton distribution, the number of singletons per sequence: for instance (3,2,2,1,1,1) indicates that out of the ten total singletons, one sequence had three, two had two, and three had one (thus six of the twelve sequences had none). *Msp5* and A.L. (adenylosuccinate lyase) are genes flanking *msp2* that overlap the sequenced regions.

^a 12 isolates were sequenced for the 5' region of *msp1*, 6 for all other regions.

^b Excluding highly polymorphic 16-copy 11 bp repeat described in text.

of the indels were singletons, while three others exhibited more than two distinct alleles.

3.2. *Csp*

We sequenced 1081 bp upstream and 1113 downstream of the *Csp* coding sequenced from six *P. falciparum* isolates from four different continents. In the upstream region, were five indels and five SNPs ($\theta=0.00173$). One indel showed more than two alleles. Otherwise, 4/5 SNPs and 3/4 indels were singletons. In the downstream region, there were eight indels and six SNPs ($\theta=0.0080$). Aside from three indels that showed more than two alleles, all polymorphisms were singletons.

3.3. *Msp2*

The genomic regions flanking *Msp2* are more complicated, owing to nearby upstream and downstream protein-coding genes and to the presence of a 16-copy repeat of an 11 bp variable sequence. In total we sequenced 1116 of flanking downstream sequence from six isolates from four continents. In the immediately adjacent 655 bp of intergenic sequence, we found two SNPs and three indels. Aside from one indel which showed more than two alleles, all polymorphisms were singletons. In 461 coding bases of the adjacent adenylosuccinate lyase gene, there were three SNPs, all singletons.

The genomic region upstream of *Msp2* comprises four distinct regions (Fig. 1A). Within 314 bp of coding sequence from the flanking *Msp5* gene, we found 3 singleton SNPs and no

indels. No polymorphism was found within a 136 bp intron within *Msp5*. Within the intergenic region, there was a 11 bp repeat region below that showed complex patterns of polymorphism (see below). Outside of that region, there were 3 SNPs and 5 indels in 840 bp. 7/8 of the observed polymorphisms were singletons.

3.4. Repeat region upstream of *Msp2*

The single highly polymorphic region encompassed 16 copies of a variable 11 bp repeat, in the region upstream of *Msp2*. The observed pattern of polymorphism in this region is summarized in Fig. 1, and the full alignment given online in Supplementary Material, online. Three repeat variants (TAATAAATATA, TTATAAATATA, and TAATATAAATA) are observed, and each allele shows exactly sixteen copies, the only length difference being attributable to a 1 bp insertion in strain D10. The sequence alignment yields a surprisingly large number of SNPs, 13 over only 110 bp. However, the variation in the pattern of usage of the three different 11-mers suggests strand slippage and/or unequal crossing over and gene conversion acting at the locus. For instance, D6 shows seven copies of the first of the three 11-mers (grey boxes in Fig. 1B) versus only five in 7G8; similarly, D6 shows only six copies of the third 11-mer (white), versus eight in most of the other alleles.

Another intriguing pattern concerns the positions of the few cytosines in the region, each of which is found in either the 6th position within an 11-mer (8th copy in all alleles except D6, 9th copy in all alleles except D6 and 7G8) or in the 8th

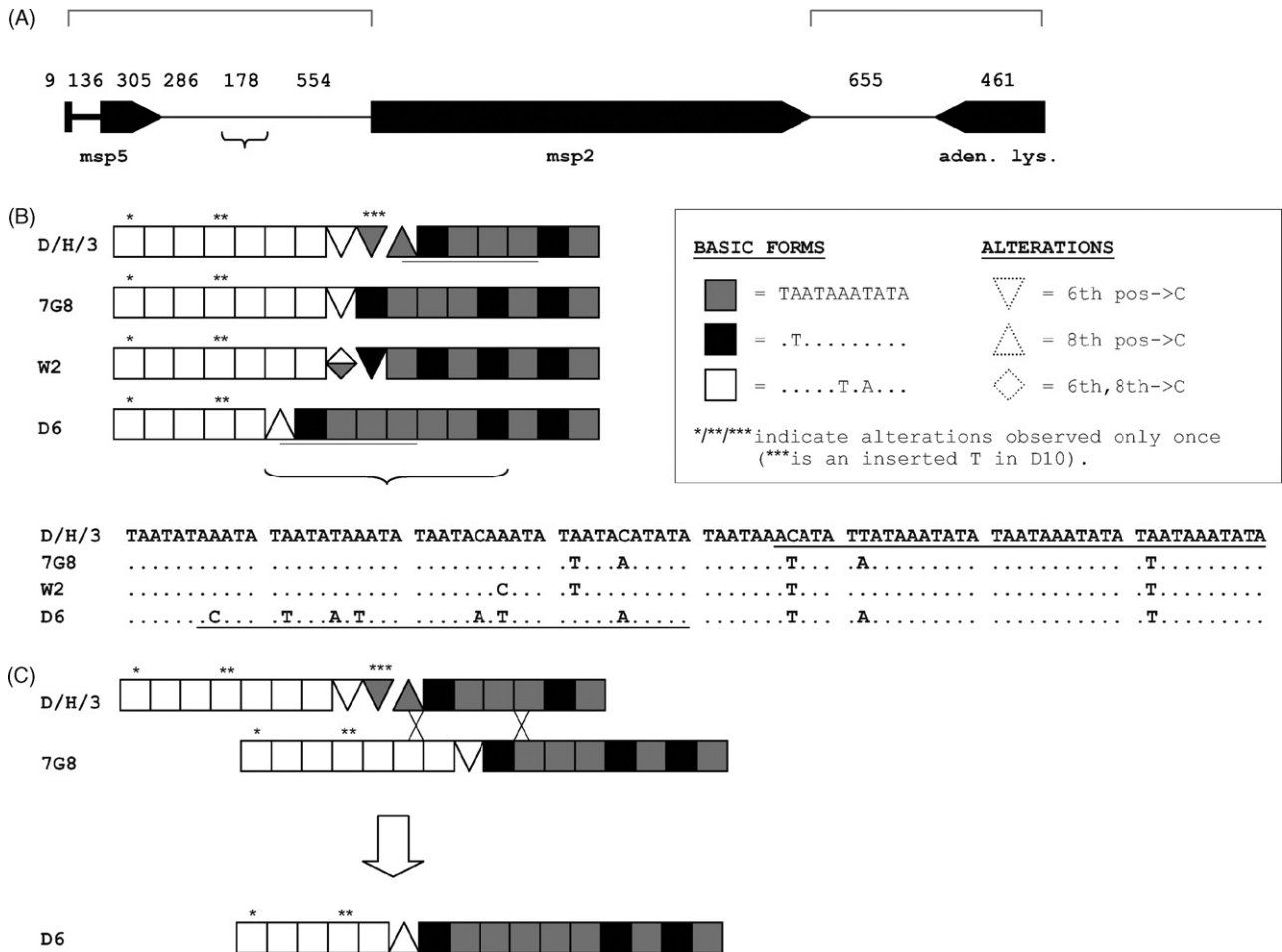


Fig. 1. Structure and repeat polymorphism flanking the MSP2 locus. (A) Layout of the *Msp2* locus. Numbers give numbers of basepairs, square brackets show sequenced regions. Boxes indicate exons, thick lines introns and thin lines intergenic regions. The curly bracket underneath the sequence indicates the repetitive region illustrated in parts B and C. (B) Pattern of polymorphism at a microsatellite upstream of *msp2*. Colored boxes indicate the different common repeat forms. Other shapes indicate alterations to the major forms, thus a grey triangle pointing up indicates an 11-mer identical to the grey box 11-mer, but with an 8th position C (i.e. TAATAAACATA). The grey/white diamond indicates “TAATACACATA”. D/H/3 indicates the allele found in D10, HB3, and 3D7 (with the exception of a single inserted “T” in the D10 sequence relative to the other two). The bracket indicates the region of alignment given below the diagram. The underlined sequences of the D/H/3 and D6 alleles are identical. (C) Possible double recombination leading to the observed D6 sequence.

position (6th repeat copy in D6, 8th copy in W2, 10th copy in D10, HB3, and 3D7). This pattern either requires recurrent mutation to C at the same two positions of the 11-mer or alternatively strand slippage, unequal crossing over, or gene conversion. One possibility for such an event involves the sequence identity between the repeats 6–10 of D6 and 10–14 of the common D10/HB3/3D7 allele (hereafter the “3D7 allele”). Indeed, the D6 sequence is the exact product expected by a gene conversion between 3D7 and 7G8 (Fig. 1C). Though the actual relationship between alleles in highly polymorphic repetitive regions is notoriously hard to infer, the use of the same repeats in different patterns and the observed positions of cytosine suggests expansions and deletions of repeats in the history of this locus.

It is of note, then, that all alleles show a variation in length of only a single basepair through the entire region. Repetitive regions in noncoding sequences tend to change rapidly in length, leading to a great range in lengths of

repetitive regions among alleles. The near complete lack of variation in allele length through this apparently very active repetitive region is therefore surprising. One possibility is that this region is important for transcriptional regulation of the downstream *msp2*, and that purifying selection is acting to maintain the total length of the region. Further investigation should clarify the possible role of selection in this region.

4. Discussion

We sequenced genomic regions flanking three antigens expressed on the surface of *P. falciparum*. We found levels of nucleotide diversity similar to those previously found for various loci not thought to be evolving under balancing selection, and a high frequency of singleton polymorphisms. Both patterns suggest that balancing selection is not acting in these regions.

4.1. Level of polymorphism

The first major expectation for a locus under ancient balancing selection is that the level of polymorphism should be high, due to the greater age of alleles (e.g. [11]). However, pairwise nucleotide diversity for all sequenced regions fell between zero and 0.00180, with a mean of 0.00088, consistent with previous estimates from 204 genes not known to be evolving under balancing selection (0.00103 and 0.00072 [4]) (Table 2), and comparable to that found in a many-isolate genome-wide study (0.00116 [23]). Across all six regions, there is one SNP per 310 bp and one indel per 128 bp in the intergenic regions.

4.2. High fraction of low-level polymorphisms

A second expectation under balancing selection is that the fraction of low-frequency polymorphisms should be low, due to maintenance of different allelic classes by balancing selection [24]. Instead, 92% (46/50) of two allele polymorphisms were singletons. This extended both to SNPs (24/25) and to two-allele indel loci (22/25). Thus, both the low overall level of polymorphism and the high level of low-frequency polymorphisms argue against the idea that regions flanking surface antigens in *P. falciparum* are under ancient balancing selection.

4.3. High SNP/indel ratio

Interestingly, the ratio of SNPs to indels in intergenic regions (25/46) is much higher than previously reported (5/36 from Ref. [6], 31/165 from Ref. [4], $P=0.0005$ by a two-tailed Fisher's exact test). Non-significantly more singleton indels were deletions than insertions relative to the common allele (17 vs. 9, $P>0.05$). Roughly half of indel polymorphic sites (21/46) showed more than two alleles, consistent with rapid mutation at microsatellites. While this difference could simply reflect differential evolution of different genomic regions, another possibility is that, as suggested by the case of the repeat region upstream of *Msp2*, there is purifying selection acting on total sequence length in regions with roles in regulating transcription of these genes.

4.4. Recombination at loci under balancing selection

These findings underscore the differences between cases of balancing selection in *P. falciparum* and another classic instance of balancing selection, the self-incompatibility alleles of some plants. At those loci, selection against recombinant alleles seems to have led to a mechanistic repression of recombination over kilobases of flanking genome sequence, leading to elevated levels of sequence diversity in flanking regions (e.g. [25,26]). The history of recombination at the loci studied here is less straightforward. On the one hand, recombinants within the highly diverged regions of *Msp1* are rarely observed. On the other, the more average levels of sequence divergence in other genic and flanking intergenic regions (average sequence diversity of 0.00088 in flanking regions, compared to 0.00103 and 0.00072 at other loci; see above) suggest significant recombination within

the locus. If recombination is not suppressed, the generally high rate of recombination in *P. falciparum* (estimated 6×10^{-4} per kb per meiosis [27]), consistent with the relatively short range of linkage disequilibrium in the species, would be expected to erase the signatures of balancing selection over short genomic regions. Thus it appears that recombination at these loci are not particularly low, and that recombinants within highly diverged regions of *Msp1* are suppressed by selection, not physical recombination rate.

4.5. Conclusions

We report the first study of intraspecific polymorphism in genomic regions flanking surface antigens in *P. falciparum*. Sequence divergence in these regions are similar to those found at loci not under balancing selection. Thus balancing selection appears to be restricted to certain regions of the coding sequences themselves, and local recombination rates have been sufficient to homogenize these flanking regions.

Acknowledgments

This work was supported by the Ellison Medical Foundation, and by NIH grants GM61351 and GM079536. SWR thanks Eugene Koonin and David Penny and their groups for intellectual and financial support, and for fostering an environment of open intellectual inquiry.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.molbiopara.2007.12.004](https://doi.org/10.1016/j.molbiopara.2007.12.004).

References

- [1] World Health Organization. WHO expert committee on malaria. World Health Organ Tech Rep Ser 2000;892:1–74.
- [2] Ferreira MU, de Silva Nunes M, Wunderlich G. Antigenic diversity and immune evasion by malaria parasites. Clin Diagn Lab Immunol 2004;11:987–95.
- [3] Rich SM, Ferreira MU, Ayala FJ. The origin of antigenic diversity in *Plasmodium falciparum*. Parasitol Today 2000;16:390–6.
- [4] Mu J, Duan J, Makova KD, et al. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. Nature 2002;418:323–6.
- [5] Conway DJ, Fanello C, Lloyd JM, et al. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. Mol Biochem Parasitol 2000;111:163–71.
- [6] Volkman SK, Barry AE, Lyons EJ, et al. Recent origin of *Plasmodium falciparum* from a single progenitor. Science 2001;293:482–4.
- [7] 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;195:273–87.
- [8] McCutchan TF, de la Cruz VF, Good MF, Wellems TE. Antigenic diversity in *Plasmodium falciparum*. Prog Allergy 1988;41:173–92.
- [9] Lockyer MJ, Marsh K, Newbold CI. Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epitopes of the circumsporozoite protein. Mol Biochem Parasitol 1989;37:275–80.
- [10] Thomas AW, Carr DA, Carter JM, Lyon JA. Sequence comparison of allelic forms of the *Plasmodium falciparum* merozoite surface antigen MSA2. Mol Biochem Parasitol 1990;43:211–20.

- [11] Hughes AL. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* 1991;127:345–53.
- [12] Hughes AL. Positive selection and intrallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. *Mol Biol Evol* 1992;9:381–93.
- [13] Shi YP, Alpers MP, Povoas MM, Lal AA. Diversity in the immunodominant determinants of the circumsporozoite protein of *Plasmodium falciparum* parasites from malaria-endemic regions of Papua New Guinea and Brazil. *Am J Trop Med Hyg* 1992;47:844–51.
- [14] Riley EM, Hviid L, Theander TG. Malaria. In: Kierszenbaum F, editor. *Parasitic infections and the immune system*. San Diego: Academic Press, Inc.; 1994. p. 119–43.
- [15] Hughes MK, Hughes AL. Natural selection on *Plasmodium* surface proteins. *Mol Biochem Parasitol* 1995;71:99–113.
- [16] Escalante AA, Lal AA, Ayala FJ. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 1998;149:189–202.
- [17] Rich SM, Licht MC, Hudson RR, Ayala FJ. Malaria's eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1998;95:4425–30.
- [18] Rich SM, Ayala FJ. The recent origin of allelic variation in antigenic determinants of *Plasmodium falciparum*. *Genetics* 1998;150(1): 515–7.
- [19] Hughes AL, Verra F. Ancient polymorphism and the hypothesis of a recent bottleneck in the malaria parasite *Plasmodium falciparum*. *Genetics* 1998;150:511–3.
- [20] Felger I, Irion A, Steiger S, Beck HP. Genotypes of merozoite surface protein 2 of *Plasmodium falciparum* in Tanzania. *Trans R Soc Trop Med Hyg* 1999;93(Suppl. 1):3–9.
- [21] Polley SD, Weedall GD, Thomas AW, Golightly LM, Conway DJ. Orthologous gene sequences from merozoite surface protein 1 of (MSP1) from *Plasmodium reichenowi* and *P. gallinaceum* confirm an ancient divergence of *P. falciparum* alleles. *Mol Biochem Parasitol* 2005;142: 25–31.
- [23] Volkman SK, Sabeti PC, DeCaprio D, et al. A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genet* 2007;39:113–9.
- [24] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;123:585–95.
- [25] Kamau E, Charlesworth D. Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Curr Biol* 2005;15:1773–8.
- [26] Hagenblad J, Bechsgaard J, Charlesworth D. Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. *Genetics* 2006;173:1057–73.
- [27] Su XZ, Ferdig MT, Huang Y, et al. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* 1999;286:1351–3.