The Origin of Antigenic Diversity in *Plasmodium falciparum*

S.M. Rich, M.U. Ferreira and F.J. Ayala

Most studies of genetic variability of Plasmodium falciparum *have focused on protein antigens and the genes that encode them. The consensus is that populations exhibit high levels of genetic polymorphism, most notably the genes encoding surface proteins of the merozoite (*Msp1, Msp2*) and the sporozoite (*Csp*). The age and derivation of this variation is a subject that warrants further careful consideration, as discussed here by Stephen Rich, Marcelo Ferreira and Francisco Ayala.*

Natural selection can cause accelerated and non-uniform rates of nucleotide substitution among antigenic loci, which confounds efforts to estimate the age of the polymorphisms in these genes. More suitable for age determination of a species is the study of nucleotide substitutions that evolve by nearly neutral processes. Accordingly, we previously examined single-copy coding regions of ten genetic loci in *Plasmodium falciparum* and no polymorphisms at any silent nucleotide sites were found; ie. the only nucleotide polymorphisms are those associated with amino acid replacements. Based on the absence of neutral substitutions at 10 912 fourfold and 20 061 twofold redundant codon sites, it was concluded, with 95% confidence, that the set of *P. falciparum* isolates in the sample had derived from a single *P. falciparum* genotype within the past 57 500 years, although the real time of this coalescence might be an order of magnitude more recent[1]. This phenomenon is most likely attributable to an extreme reduction – usually referred to as a 'bottleneck' – in global *P. falciparum* population size. An independent study of ten additional loci, most of which encode antigenic determinants, has also shown a paucity of silent polymorphisms[2].

Saul[3] has argued that the paucity of synonymous substitutions is attributable to the high AT content of the *P. falciparum* genome. Although we agree that AT bias might affect substitution rates, it cannot account for the complete absence of polymorphism[4]. Three lines of evidence support this: (1) intra- and interspecific comparisons of *Plasmodium* show that synonymous substitutions have occurred, even in the lineages leading to *P. falciparum* and *P. reichenowi*; (2) among fourfold redundant codons, AT bias may lead to restriction of A/T ↔ G/C changes, although a survey of 312 coding regions shows that A ↔ T changes are definitely not restricted; and (3) in determining the age of the *P. falciparum* bottleneck, synonymous and non-synonymous substitution rates have been estimated empirically among *Plasmodium* spp, and these estimates are corrected for differential rates among two- and fourfold codons[4,5].

There can be little doubt that the ancestral *P. falciparum* propagule originated in Africa, and that its expansion within and from that continent was a consequence of human activity in historical times, starting with the Neolithic events that brought agriculture to Africa six to seven thousand years ago, with the associated formation of human settlements. Moreover, the expansion of *P. falciparum* outside Africa may have been associated with the gradual increase in global temperatures that followed the Würm glaciation, which peaked some 15 000 years ago, so that about 6000 years ago, climatic conditions in the Middle East and the Mediterranean region made possible the spread of *P. falciparum* and its vectors beyond the African tropics[6–8]. The demographic and climatic changes might, in turn, have facilitated the speciation of the highly anthropophilic *Anopheles* vectors that are now largely responsible for the effective transmission of *P. falciparum* in human populations[8]. In more recent times, colonial expansion and the slave trade might have contributed to the dispersion of *P. falciparum* outside of Africa[8,9].

Age of antigenic alleles

As expected for genes under strong diversifying selection for evasion of the human immune response[2], antigenic genes of *P. falciparum* are exceedingly polymorphic. Indeed, the high number of non-synonymous nucleotide substitutions relative to synonymous substitutions is evidence of diversifying selection[2]. Moreover, much of the amino acid polymorphism observed in antigenic genes has been mapped directly to B- and T-cell epitopes[10]. The question is: how old are these antigenic polymorphisms?

Hughes and colleagues[11,12] have hypothesized that the polymorphisms of genes encoding *P. falciparum* surface proteins [merozoite surface protein (MSP) and circumsporozoite surface protein (CSP)] are very old, perhaps older than the species itself. They estimated that the ages of the most divergent alleles of *Msp1* and *Csp* are 35 million and 2.1 million years, respectively. Balancing natural selection can maintain gene polymorphisms for millions of years; as is the case for the vast diversity of human major histocompatibility complex (MHC) molecules, some of which far pre-date the split between humans and chimpanzees[13].

The apparent age incongruity between antigenic and non-antigenic genes, however, may result from the disparity of evolutionary rates between these genes, and even among segments of the individual loci. A notable feature of nearly every *P. falciparum* surface protein identified to date is the presence of repeating nucleotide sequences that encode iterative amino acid sequences[14]. These antigenic repeat regions are highly mutable[15,16]. The propensity of antigenic genes

Stephen M. Rich is at the Division of Infectious Diseases, Tufts University School of Veterinary Medicine, 200 Westboro Rd, Bldg 20, North Grafton, MA 01536, USA. Marcelo Urbano Ferreira is at the Laboratory of Molecular Parasitology, Faculty of Medicine of São Jose do Rio Preto, Av. Brigadeiro Faria Lima 5416, 15090-000 São Jose do Rio Preto (SP), Brazil. Francisco J. Ayala is at the Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA. **Tel: +1 508 887 8924, Fax: +1 508 839 7911, e-mail: srich01@emerald.tufts.edu**

 0169-4758/00/$ – see front matter PII: S0169-4758(00)01741-5

Fig. 1. Alignment of *Csp* repeat allotypes (RATs). Sequences are named by the species (Pf, *Plasmodium falciparum*; Pr, *P. reichenowi*) and the GenBank accession number. Each RAT is a unique nucleotide sequence encoding a certain amino-acid motif. Two amino-acid repeat motifs are present in *P. falciparum*: NANP and NVDP (shaded). The NANP RATs are α(aatgcaaaccca), β(aatgcaaatcct), χ(aatgcaaatcca), δ(aatgccaatcca), ϵ(aatgaaacccc), ϕ(aacgcaaacccc), γ(aatgccaaccca), η(aagcaaaccct), ι(aacgcaaaccca) and ψ(aatgccaacccc). The NVDP RATs are: μ(aatgtagatcca), υ(aatgtagatccc), ο(aacgtagatcct) and π(aatgtagatcct). Subscripts refer to the number of consecutive repeats of the RAT.

to mutate rapidly coupled with natural selection favoring novel antigens might account for the seemingly great age of the alleles.

Merozoite and sporozoite surface antigens

It is proposed here that most of the variation in antigenic genes is attributable to duplication and/or deletion of the repeated segments within the genes. This process occurs by several mechanisms, each of which is well understood at the molecular level and might involve either intra- or interhelical exchange of DNA[17]. These mechanisms will be referred to by the generic term intragenic recombination (IGR), which increases or decreases the number of repeats within a genetic locus.

The IGR process is often associated with the evolution of mini- or microsatellite DNA loci, such as those recently described in *P. falciparum*[18,19]. However, IGR has also been implicated in generating variability within coding regions in a variety of eukaryotic genes, including those encoding *Drosophila* yolk protein and human α_2-globin[20,21]. The probable effects of IGR in antigen-encoding genes of *P. falciparum* have been demonstrated, with examples of the *Csp*, *Msp1* and *Msp2* genes. These loci were chosen because: (1) they are widely used in studies of epidemiology and population structure; (2) their polymorphisms are believed to be ancient[12,22]; (3) they contain repeated DNA segments; and (4) each is a prototypical example of the various stages in the differentiation of genes by IGR.

The *Csp* gene encodes the antigenic circumsporozoite protein, which has been investigated extensively because it is a likely target for vaccine development[23,24]. The gene comprises two end-regions that are not repetitive (5′ NR and 3′ NR), which embrace a central region (CR) made up of a variable number (typically, between 40 and 50) of tandemly arranged 12 nucleotide repeats. There are no silent polymorphisms in the 5′ NR and 3′ NR regions, which is part of the evidence used to infer the recent origin of *P. falciparum* populations[1,4].

The repetitive amino acid sequences encoded within the CR are remarkably conserved (only two amino acid motifs are known in *P. falciparum*: NANP and NVDP), but there is a great deal of synonymous nucleotide polymorphism among the repeats. To quantify the degree of nucleotide difference among these motifs, Rich *et al.*[25] introduced the concept of the repeat allotype (RAT) to refer to the set of variant nucleotide sequences that encode a single amino acid motif. Using the RAT as the basic evolutionary unit, it is possible to achieve correct alignments between gene sequences and, hence, to determine their homologies[25]. Among the known *Csp* gene sequences of *P. falciparum*, there are ten RATs that encode the NANP motif and four that encode the NVDP motif (Fig. 1). Each RAT is identified by a Greek letter to distinguish its alignment from that of either nucleotides or amino acids (Fig. 1). The pattern of duplication/deletion of RATs clearly reflects the underlying IGR mechanisms that generate diversity in the CR. Identical symbols in the columns of this alignment indicate identical nucleotide sequences between alleles. Note that nearly all of the observed synonymous site differences in the CR are between RATs found within any single allele. This is a strong indication that

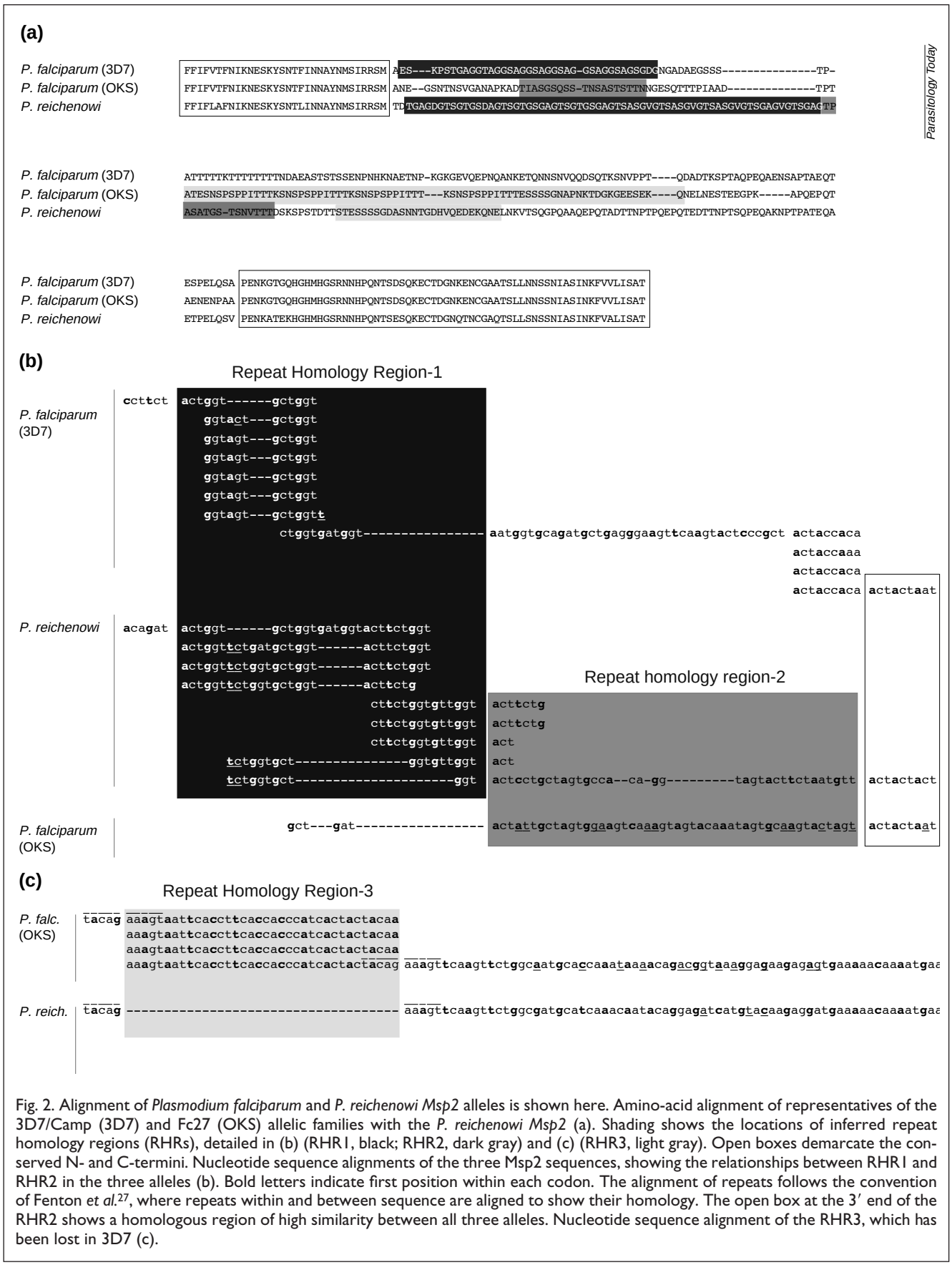


Fig. 2. Alignment of *Plasmodium falciparum* and *P. reichenowi Msp2* alleles is shown here. Amino-acid alignment of representatives of the 3D7/Camp (3D7) and Fc27 (OKS) allelic families with the *P. reichenowi Msp2* (a). Shading shows the locations of inferred repeat homology regions (RHRs), detailed in (b) (RHR1, black; RHR2, dark gray) and (c) (RHR3, light gray). Open boxes demarcate the conserved N- and C-termini. Nucleotide sequence alignments of the three Msp2 sequences, showing the relationships between RHR1 and RHR2 in the three alleles (b). Bold letters indicate first position within each codon. The alignment of repeats follows the convention of Fenton *et al.*[27], where repeats within and between sequence are aligned to show their homology. The open box at the 3′ end of the RHR2 shows a homologous region of high similarity between all three alleles. Nucleotide sequence alignment of the RHR3, which has been lost in 3D7 (c).

although RAT diversity might have an ancient origin, it has been maintained within individual alleles and can therefore withstand even the most constricted bottleneck. For example, all 25 *Csp* CR alleles contain at least one copy of each of the most common RATs (α, β, χ, δ, ϕ and γ), which constitute more than 93% of all NANP repeats. If any one of these sequences were the sole survivor following a bottleneck, it alone would possess nearly all the diversity currently known for the species. After some cell generations, IGR rearrangements

of these RATs generate size polymorphisms in the resulting alleles. This process has presumably occurred numerous times in the evolution of the species, and might continue to do so, given the nature of the parasite life style and its propensity for being confronted by population bottlenecks. Interestingly, the single-known *Csp* CR of *P. reichenowi*, is more variable than all known *P. falciparum* alleles combined, in that it has three amino acid repeat motifs: NVNP as well as the two *P. falciparum* motifs (NANP and NVDP).

The approach used to determine the evolution of the *Csp* CR is not applicable to all *P. falciparum* antigenic determinants. For example, the *Msp2* of *P. falciparum* shows much greater variability in length, amino acid content and number of repeats; therefore, the number of nucleotide sequences encoding one given identical amino acid motif is limited. Nonetheless, the pattern of allele polymorphism in *Msp2* is consistent with the IGR model.

Similar to CSP, the MSP-2 protein is characterized by N- and C-termini with 43 and 74 residues, respectively[26]. Bracketed within these conserved segments is the highly variable repeat region. Two allelic families have been identified and named after the isolates in which they were first identified. The FC27 family is characterized by at least one copy of a 32-amino acid sequence and a variable number of a 12-amino acid repeat; the 3D7/Camp family contains tandem amino acid repeats of 4–10 amino acids in length[27].

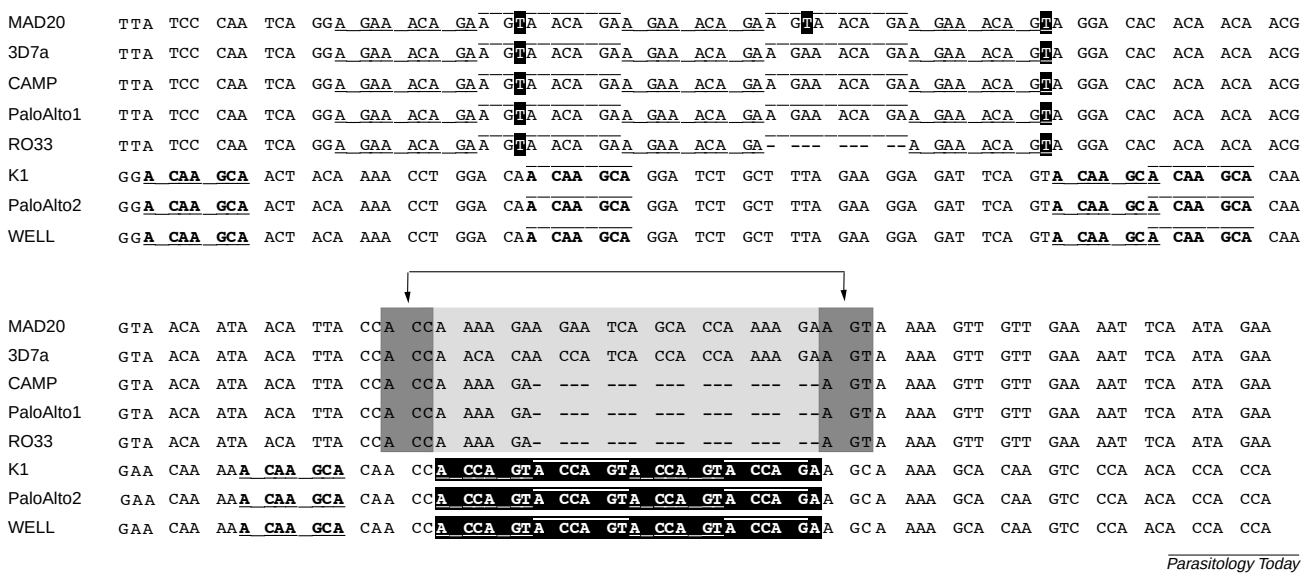
The 3D7/Camp alleles are more variable in length and sequence of repeat types than are those of the FC27 family[16]. Fenton *et al.*[28] proposed a model to explain the origin of repeat diversity within the 3D7/Camp family of alleles. The 3D7/Camp family was divided into distinct allelic subclasses, which included types A1 and A3, distinguished by amino acid repeats of different lengths. For example, A1 alleles possess four amino acid motifs, whereas a repeating eight amino acid motif occurs in A3. Fenton *et al.* have shown that the allelic subclasses within the 3D7/Camp family are derived from a common ancestral nucleotide sequence and that the diversity arises from duplication and deletion of repeat subunits[28].

Recently, Dubbeld *et al.*[29] have cloned and sequenced the *Msp2* gene of *P. reichenowi* (*PrMsp2*), which is a 'unique mosaic of *P. falciparum* allelic forms and species-specific elements'. The methods described in Ref. 28 have been used to determine whether *PrMsp2* provides insight into the ancestry of the FC27 and 3D7/Camp families. Figure 2a shows the amino acid sequence alignment of two *P. falciparum* MSP-2 proteins with the *PrMSP2*. The *P. falciparum* alleles from the 3D7 and OKS isolates are representative of the 3D7/Camp and FC27 families, respectively. The two *P. falciparum* alleles are identical at nucleotide sites encoding the N- and C-termini, but exhibit little similarity, even at the amino acid level, in the intervening repeat region. A closer look at the nucleotides within this central portion reveals homology at three distinct regions – the repeat homology regions (RHRs). RHR1 shows common ancestry between the *PrMsp2* and the

Table 1. Nucleotide diversity within and between Group I and II alleles of the *Plasmodium falciparum Msp1* genes[a]

Block	Length (codons)[b]	πSynonymous[c]			πNon-synonymous		
		Group I only	Group II only	Group I & Group II	Group I only	Group II only	Group I & Group II
1	55	0.019	0.021	0.017	0.017	0.010	0.013
2	55	0.106	0.185	0.150	0.449	0.497	0.553
3	202	0.038	0.006	0.042	0.018	0.000	0.023
4	31	0.031	0.000	0.020	0.307	0.000	0.215
5	35	0.000	0.000	0.070	0.000	0.000	0.026
6	227	0.000	0.000	0.282	0.004	0.001	0.300
7	73	0.000	0.000	0.361	0.003	0.000	0.072
8	95	0.000	0.000	0.338	0.000	0.003	0.711
9	107	0.000	0.023	0.409	0.005	0.043	0.126
10	126	0.008	0.000	0.448	0.011	0.000	0.394
11	35	0.000	0.000	0.128	0.000	0.000	0.068
12	79	0.000	0.000	0.000	0.000	0.000	0.000
13	84	0.000	0.042	0.040	0.005	0.007	0.052
14	60	0.000	0.018	0.212	0.002	0.005	0.371
15	89	0.000	0.000	0.216	0.001	0.003	0.089
16	217	0.002	0.032	0.277	0.005	0.027	0.185
17	99	0.002	0.019	0.007	0.010	0.027	0.016

[a] Abbreviation: Msp, merozoite surface protein.
[b] Block length may vary between Group I and II alleles, the given value indicates the average length of Group I and II alleles.
[c] Shading indicates the relative degree of amino acid polymorphism for each block as reported by Tanabe *et al.*[31] Unshaded, conserved; light gray, semi-conserved; dark gray, variable.



Parasitology Today

Fig. 3. Partial alignment of *Msp1* (Block 8) Group I and II alleles are shown here. Alternating odd and even occurrence of a repeat is indicated by underline and overbar, respectively. A 9-bp sequence (shown in italics) appears in the five Group II alleles as five tandem repeats in all but the RO33 allele, where one copy has been lost. Two repeats, measuring 7 bp (bold) and 6 bp (black shading), are found in Group I alleles. Some of the 7-bp repeats are separated by several codons, while the 6-bp repeats occur in tandem. There are no repeat sequences shared between Group I and II; however, the 6-bp repeat in Group I alleles clearly derives from a deletion of the intervening lightly shaded portion of Group II alleles, followed by IGR duplication of the resulting accgat motif (junction is indicated by arrows). In this regard, the Camp, Palo Alto-1 and RO33 alleles are intermediate between MAD20/3D7 and K1/Palo Alto-2?Wellcome alleles. The alleles shown are from the GenBank database as follows: MAD20 (X05624), 3D7 (Z35327), Camp (X03831), PaloAlto1 (m37213), RO33 (Y00087), K1 (X03371), PaloAlto2 (X15063) and Wellcome (A04562).

3D7 *Msp2* alleles (Fig. 2b). Diversity within this region results from proliferation of the GGTGCT hexamer, as described by Fenton *et al.*[28] This hexamer is ancestral to the 3D7/Camp and *PrMsp2* allelic repeats within this region. Although conservation of these codons is clear among these two alleles, it appears that they have been lost altogether in the FC27-like alleles. However, the region adjacent to RHR1 in the *PrMsp2* sequence is similar to the first 21 amino acids of the 32 amino acid repeat found within the FC27 family, and this sequence is the basis for the inferred RHR2 (Fig. 2b). The last nine nucleotides of RHR2 also manifest homology between all three sequences, including the short stretch following the $[actaccaaa]_4$ repeat in 3D7. Note also the overlap between repeating nucleotides of *PrMsp2* in both RHR1 and RHR2.

A third RHR is located further downstream, and shows the relationship between the 12 amino acid repeats of OKS and *PrMsp2* (Fig. 2c). The repeat region in OKS is surrounded on either side by a 10-bp sequence (tacagaaagt), which occurs as only a single 5′ copy in the *PrMsp2* allele. Despite the lengthy repeat insertion in the OKS sequence, the homology of OKS and *PrMsp2* in the region downstream of this repeat is apparent. Therefore, it appears that the repeats were generated some time after the split between *P. falciparum* and *P. reichenowi*.

Analysis of the single *P. reichenowi* sequence allows us to approximate the ancestral sequence of the two *P. falciparum Msp2* allele families. Indeed, comparison of the three RHRs discloses that, although the precursor sequences for the various repeats probably derive from the common *P. falciparum–P. reichenowi* ancestral species, the extant diversity among the *Msp2* alleles has occurred since the divergence of the two species. The distinctive dimorphism of the two *P. falciparum* alleles results from proliferation of repeats in two different regions of the molecule. Presumably, because the overall MSP-2 molecule is constrained in size, the proliferation of repeats leads to loss of other regions; ie. the 3D7/Camp repeat precursors were lost in FC27 alleles, and the FC27 repeat precursors were lost in the 3D7 alleles.

The repetitive DNA sequences found within the *Csp* and *Msp2* genes, as well as those among other *P. falciparum* antigenic determinants, are clearly subject to much higher rates of mutation than are nonrepeat sequences found within the same locus. Indeed, the paucity of silent substitutions within the nonrepetitive regions indicates that IGR events have generated repeat diversity in a relatively short period of time. Empirical estimates of mutation rates among repetitive DNA sequences, such as satellite DNA, are as high as 10^{-2} mutations per generation and therefore several orders of magnitude greater than rates for point mutations[30]. These high mutation rates, coupled with strong selection for immune evasion, yield an extremely accelerated evolutionary rate for *P. falciparum* antigens.

The *Msp1* gene has been cited as an apparent exception to the rule of the association between extreme antigenic polymorphism and occurrence of repetitive DNA. Like *Msp2*, *Msp1* exhibits considerable substitution and length variation between two allelic classes (Group I and Group II), but much less variation within each class[11,31]. The two classes are commonly designated by the strains in which they were originally identified: K1 (Group I) and MAD20 (Group II). Tanabe *et al.*[31] partitioned the MSP-1 protein into 17 blocks, based on the degree of amino acid polymorphism; seven are highly variable, five are semi-conserved and five are conserved. Table 1 is a summary of the synonymous and

non-synonymous nucleotide diversity (π) for each of these 17 blocks. Note that within either group, non-synonymous and synonymous polymorphisms are absent or rare in most regions, with the notable exception of Block 2, which encodes a set of repetitive tripeptides, and is thus subject to the same type of diversity-generating IGR found in *Msp2* and *Csp*.

However, most blocks exhibit far greater nucleotide polymorphisms between than within groups. Based on the diversity in the region encompassing Blocks 4–10, Hughes[22] concluded that the divergence between Group I and II alleles occurred about 35 million years ago. However, he inferred an age of 0.5 million years for a small region within Block 3 (which Hughes referred to as Region 4). Hughes contends that this 70-fold difference in age of allelic blocks, which are separated by <200 bp, is attributable to high recombination between blocks and a strong balancing selection that has maintained these alleles throughout half of the evolution of the genus. This scenario is extraordinarily improbable, and seems not to fit the observations. Specifically, if the Block 4–10 region was in fact tens of millions of years old, we would expect to see considerable within-group synonymous site polymorphism – but this is not the case.

Rather, it is proposed that it is the rate of evolution, and not the age of these blocks, that is so vastly different. Here too, it is the repetitive DNA regions that are implicated in the rate difference. The dimorphism among Group I and II repeats within Block 2 has been shown to result from processes exactly analogous to those within the *Msp2* repeat region[32,33]. The occurrence of repetitive DNA within other blocks has not been described to date. However, repeats within several of the most polymorphic *Msp1* blocks have been identified, in particular, Blocks 4, 8 and 14, which were previously characterized as non-repetitive[35].

Work focused on the repeats detected within Block 8, which is the block identified by Tanabe *et al.* as showing the lowest amino acid similarity between groups (10%), and which, in our analysis, is the most polymorphic in terms of non-synonymous nucleotide diversity (π = 0.711)[35]. The presence of three group-specific repeats within this block (Fig. 3) was reported[35]. One 9 bp repeat (R2a) is found in all Group II alleles (the five uppermost alleles in Fig. 3); and two repeats, of 6 bp (R1a) and 7 bp (R1b), are present in all Group I alleles. It is hypothesized that the occurrence of these repeats within this very short stretch of DNA is a highly significant departure from chance, and this was tested by searching the recently completed genomic sequences of *P. falciparum* chromosomes 2 and 3. The nucleotide sequences of repeats R1a, R1b and R2a appear 25, 116 and 11 times, respectively, within the 947 kbp of chromosome 2. Within the 1060 kbp of chromosome 3, the R1a, R1b and R2a repeats are present 39, 52 and seven times, respectively. None of the three nucleotide repeats ever appears in tandem on either chromosome 2 or 3. Moreover, the average distance between each occurrence on these chromosomes is >20 kb, demonstrating that their repeated occurrence in the short 147 bp segment of *Msp1* Block 8 is a strong departure from random expectation. The *Msp1* gene is located on chromosome 9, which has not yet been assembled as a complete nucleotide sequence; nonetheless, the distribution of these nucleotide repeats is not likely to differ markedly between chromosomes by chance alone.

It is worth noting that R1a and R2a also exist as clustered repeats outside of *Msp1*, but they are in both cases located within encoded surface proteins. Thus, on chromsome 2: (1) five of the 11 R2a repeats are located within a 558 bp region corresponding to a predicted secreted antigen that appears similar to the glutamic acid-rich protein gene; and (2) within the *pfEMP* member of the *var* gene family, there are 67 repeats, each 39 bp long and the 3′ terminus of each of the 67 repeats is an R1a sequence. The biological significance of the occurrence of these repeat motifs within multiple antigens is difficult to interpret, but these tantalizing observations lead us to wonder whether these repeats are random products of IGR events, or whether they play some important role in recombination, as would be the case if they were involved in site-specific recombinase activity. In any case, what is clear from the observation of highly significant repeats within regions of the *Msp1* gene previously thought to be non-repetitive is that the extensive polymorphism is attributable to the same kinds of repeat variation and rapid divergence known in the other antigenic determinants.

Conclusions

Homologous comparisons among allelic variants of antigenic genes reveal that most of the observed variation is directly attributable to rapid mutational processes associated with IGR. The increased rate of evolution among these genes reconciles the recent origin of extant *P. falciparum* populations with the abundance of antigenic diversity observed globally and locally. Conclusions regarding the evolutionary origin of antigenic diversity in *P. falciparum* have bearing on determining the mechanisms for generating the novel antigen alleles that ensure the long-term survival of the parasite[35]. What remains is to ascertain the relevance of the various IGR mechanisms that underlie the diversification process. It has been noted that IGR can result from either intra- or interhelical events. An example of intrahelical recombination is that of mitotic, slipped-strand mismatch repair (SSM), which is considered to be the principal source of variation in repetitive units such as satellite DNA. Interhelical recombination derives from the classic process of meiotic crossing over and recombination within or between loci on homologous chromosomes.

Both of these processes clearly occur in *P. falciparum*. Kerr *et al.*[34] have shown that meiotic, interhelical recombination occurs between mixed *Msp2* genotype parasites passaged in laboratory animals. Indeed, this process constitutes the basis for generating linkage maps of *P. falciparum* chromosomes[18]. But it has been shown that, despite the abundant intragenic recombination within *Csp* CR, there is an apparent absence of recombination between 5′ and 3′ NR, suggesting that the duplication and deletion of RATs occur by mitotic processes such as SSM[25]. SSM has also been implicated[28] as the cause of repeat variation in *Msp2*. However, it is interesting to note that among >100 field isolates from which *Msp2* has been sequenced and entered in GenBank, only six have hybrid 3D7/Camp-FC27 sequences, despite the strong bias towards sequencing isolates with unusual serotyping results.

The debate over the relevance of sexual recombination between *P. falciparum* types has been contentious and will probably remain so for some time. However,

as with most controversies centering upon mutually exclusive, dichotomous viewpoints, the final resolution may come from conciliation. In any case, it is becoming increasingly clear that the population structure of *P. falciparum* might not be uniform throughout the species, but dependent upon local factors related to parasite, vector and host biology[36–39]. An accurate determination of these factors is contingent upon careful analysis of parasite genotypes and appropriate determination of homologous comparisons.

References

1 Rich, S.M. *et al.* (1998) Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4425–4430
2 Escalante, A.A. *et al*. (1998) Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 149, 189–202
3 Saul, A. (1999) Circumsporozoite polymorphisms, silent mutations and the evolution of *Plasmodium falciparum*. *Parasitol. Today* 15, 38–39
4 Rich, S.M. and Ayala, F.J. (1998) The recent origin of allelic variation in antigenic determinants of *Plasmodium falciparum*. *Genetics* 150, 515–517
5 Rich, S.M. and Ayala, F.J. (1999) Reply to Saul. *Parasitol. Today* 15, 39–40
6 de Zulueta, J. *et al*. (1973) Entomological aspects of receptivity to malaria in the region of Navalmoral of Mata. *Rev. Sanid. Hig. Publica (Madr.)* 47, 853–870
7 de Zulueta, J. (1994) Malaria and ecosystems: from prehistory to posteradication. *Parassitologia* 36, 7–15
8 Coluzzi, M. (1997) Interazioni evolutive uomo-plasmodio-anofele, in *XXII Seminario su Evoluzione Biologica & i Grandi Problemi della Biologia*, pp 263–285, Accademia dei Lincei
9 Sherman, I.W. (1998) A brief history of malaria and the discovery of the parasite's life cycle, in *Malaria: Parasite Biology, Pathogenesis, and Protection* (Sherman, I.W., ed.), pp 3–10, American Society of Microbiology
10 Anders, R.F. *et al.* (1993) Molecular variation in *Plasmodium falciparum*: polymorphic antigens of asexual erythrocytic stages. *Acta Trop.* 53, 239–253
11 Hughes, A.L. (1992) Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. *Mol. Biol. Evol.* 9, 381–393
12 Hughes, M.K. and Hughes, A.L. (1995) Natural selection on *Plasmodium* surface proteins. *Mol. Biochem. Parasitol.* 71, 99–113
13 Ayala, F.J. (1995) The myth of Eve: molecular biology and human origins. *Science* 270, 1930–1936
14 Anders, R.F. *et al*. (1988) Antigens with repeated amino acid sequences from the asexual blood stages of *Plasmodium falciparum*. *Prog. Allergy* 41, 148–172
15 Arnot, D.E. (1991) Possible mechanisms for the maintenance of polymorphisms in *Plasmodium* populations. *Acta Leiden* 60, 29–35
16 Felger, I. *et al*. (1997) Sequence diversity and molecular evolution of the merozoite surface antigen 2 of *Plasmodium falciparum*. *J. Mol. Evol.* 45, 154–160
17 Levinson, G. and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221
18 Su, X. and Wellems, T.E. (1996) Toward a high-resolution *Plasmodium falciparum* linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* 33, 430–444
19 Anderson, T.J. *et al*. (1999) Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology* 119, 113–125
20 Oron-Karni, V. *et al*. (1997) A novel mechanism generating short deletion/insertions following slippage is suggested by a mutation in the human alpha2-globin gene. *Hum. Mol. Genet.* 6, 881–885
21 Ho, K.F. *et al*. (1996) Phylogenetic analysis of DNA length mutations in a repetitive region of the Hawaiian Drosophila yolk protein gene Yp2. *J. Mol. Evol.* 43, 116–24
22 Hughes, A.L. (1993) Coevolution of immunogenic proteins of *Plasmodium falciparum* and the host's immune system, in *Mechanisms of Molecular Evolution* (Takahata, N. and Clark, A.G., eds), pp 109–127, Sinauer Association
23 Zevering, Y. *et al*. (1998) Human and murine T-cell responses to allelic forms of a malaria circumsporozoite protein epitope support a polyvalent vaccine strategy. *Immunology* 94, 445–454
24 Gramzinski, R.A. *et al*. (1997) Malaria DNA vaccines in Aotus monkeys. *Vaccine* 15, 913–915
25 Rich, S.M. *et al*. (1997) *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13040–13045
26 Smythe, J.A. *et al*. (1991) Structural diversity in the *Plasmodium falciparum* merozoite surface antigen 2. *Proc. Natl. Acad. Sci. U. S. A.* 88, 1751–1755
27 Felger, I. *et al*. (1994) *Plasmodium falciparum*: extensive polymorphism in merozoite surface antigen 2 alleles in an area with endemic malaria in Papua New Guinea. *Exp. Parasitol.* 79, 106–116
28 Fenton, B. *et al*. (1991) Structural and antigenic polymorphism of the 35- to 48-kilodalton merozoite surface antigen (MSA-2) of the malaria parasite *Plasmodium falciparum*. *Mol. Cell. Biol.* 11, 963–974
29 Dubbeld, M.A. *et al*. (1998) Merozoite surface protein 2 of *Plasmodium reichenowi* is a unique mosaic of *Plasmodium falciparum* allelic forms and species-specific elements. *Mol. Biochem. Parasitol.* 92, 187–192
30 Schug, M.D. *et al*. (1998) Mutation and evolution of microsatellites in *Drosophila melanogaster*. *Genetica* 102/103, 359–367
31 Tanabe, K. *et al*. (1987) Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 195, 273–287
32 Frontali, C. (1994) Genome plasticity in *Plasmodium*. *Genetica* 94, 91–100
33 Frontali, C. and Pizzi, E. (1991) Conservation and divergence of repeated structures in *Plasmodium* genomes: the molecular drift. *Acta Leiden* 60, 69–81
34 Kerr, P.J. *et al*. (1994) Proof of intragenic recombination in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 66, 241–248
35 Rich, S.M. and Ayala, F.J. (2000) Population structure and recent evolution of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6994–7001
36 Paul, R.E. *et al*. (1995) Mating patterns in malaria parasite populations of Papua New Guinea. *Science* 269, 1709–1711
37 Babiker, H. and Walliker, D. (1997) Current views on the population structure of *Plasmodium falciparum*: implications for control. *Parasitol. Today* 13, 262–267
38 Conway, D.J. *et al*. (1999) High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4506–4511
39 Sakihama, N. *et al*. (1999) Allelic recombination and linkage disequilibrium within Msp-1 of *Plasmodium falciparum*, the malignant human malaria parasite. *Gene* 230, 47–54