The paradoxical population genetics of *Plasmodium falciparum*

Daniel L. Hartl, Sarah K. Volkman, Kaare M. Nielsen, Alyssa E. Barry, Karen P. Day, Dyann F. Wirth and Elizabeth A. Winzeler

Among the leading causes of death in African children is cerebral malaria caused by the parasitic protozoan *Plasmodium falciparum*. Endemic forms of this disease are thought to have originated in central Africa 5000–10 000 years ago, coincident with the innovation of slash-and-burn agriculture and the diversification of the *Anopheles gambiae* complex of mosquito vectors. Population genetic studies of *P. falciparum* have yielded conflicting results. Some evidence suggests that today's population includes multiple ancient lineages pre-dating human speciation. Other evidence suggests that today's population derives from only one, or a small number, of these ancient lineages. Resolution of this issue is important for the evaluation of the long-term efficacy of drug and immunological control strategies.

Daniel L. Hartl*
Kaare M. Nielsen
Dept of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Ave, Cambridge, MA 02138, USA.
*e-mail: dhartl@oeb.harvard.edu

Sarah K. Volkman
Dyann F. Wirth
Dept of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA.

Alyssa E. Barry
Karen P. Day
Peter Medawar Building for Pathogen Research and the Dept of Zoology, University of Oxford, Oxford, UK OX1 3FY.

Elizabeth A. Winzeler
Scripps Research Institute, Dept of Cell Biology, 10550 Torrey Pines Rd, La Jolla, CA 92037, USA.

Among killer infectious diseases, malaria causes fewer deaths than tuberculosis, but most of its victims are infants and toddlers. Other high-risk groups include women during pregnancy and non-immune travelers, refugees, displaced persons and migrant laborers. The disease is endemic in tropical regions of more than 100 countries and territories with a combined population of 2.4 billion. The most lethal form of the disease, caused by the protozoan parasite *Plasmodium falciparum*, has an annual clinical-case prevalence of 300–500 million. The majority of these occur in sub-Saharan Africa, amounting to 1.5–2.7 million deaths each year, primarily of children under the age of five years [1].

Various lines of evidence suggest that severe malaria caused by *P. falciparum* may have become endemic in tropical Africa coincident with the advent of slash-and-burn agriculture 5000–10 000 years ago [2]. The evidence for this RECENT ORIGIN OF MALARIA (see Glossary) includes the estimated time of diversification of the highly anthropophilic mosquitoes of the *Anopheles gambiae* and *Anopheles funestus* complexes that transmit the disease in Africa [3]. It also includes the estimated time of origin of malaria-resistant alleles of the gene for glucose-6-phosphate dehydrogenase in human red blood cells [4].

Recent data suggest that the level of genetic polymorphism found in most genes among isolates of *P. falciparum* is far lower than that would be expected of an ancient population maintaining itself at a consistently large population size. The apparently low level of genetic polymorphism in *P. falciparum* suggests that endemic malaria not only arose relatively recently, but also that it originated from only a small number of haploid individuals (in the limiting case, perhaps from a single individual). It will be convenient to refer to this hypothesis as the RECENT COMMON ANCESTOR (RCA) hypothesis. By contrast, some *P. falciparum* genes have a pattern of genetic variation that is seemingly inconsistent with this hypothesis. This finding suggests that many ancient lineages contributed to the evolution of endemic malaria, which is referred to as the ANCIENT COMMON ANCESTOR (ACA) hypothesis. Some alternatives that could reconcile the apparently conflicting data are discussed below. Finally, the article comments briefly on why the population history of *P. falciparum* and its current level of genetic polymorphism are important in evaluating strategies for immunological or therapeutic intervention.

Common ancestors: ancient versus recent

Evidently, *P. falciparum* is an ancient parasite of humans. Its closest known relative is *Plasmodium reichenowi* (known from a single isolate from a chimpanzee), from which it is estimated to have diverged ~6–10 million years ago, at approximately the same time as the human–chimpanzee divergence [5]. These species are allied with plasmodial parasites of birds and are only remotely related to those that cause other human

Glossary

Ancient common ancestor (ACA) hypothesis: The hypothesis that contemporary populations of *Plasmodium falciparum* represent a large number of ancient lineages that last shared a common ancestor at least 100 000 years ago.
Effective population size: The effective size of a population of organisms is the size of a theoretically ideal population that has the same magnitude of random drift in allele frequency as the actual population; typically the effective size is smaller, and sometimes much smaller, than the actual size [33].
Hierarchical *F* statistics (F_{IS}, F_{ST}, F_{IT}): F_{IS} measures the deficiency of heterozygous genotypes, relative to random mating, that results from inbreeding within local populations; F_{ST} measures the deficiency of heterozygous genotypes, relative to random mating, that can be attributed to geographical population structure; and F_{IT} measures the overall deficiency of heterozygous genotypes, relative to random mating, from both effects combined.
Recent common ancestor (RCA) hypothesis: The hypothesis that contemporary populations of *P. falciparum* all originated from a small number of individuals (and possibly only one) as recent as 5000 years ago and as many as 40 000 years ago [14,26].
Recent origin of malaria: The hypothesis that epidemic malaria originated in central Africa coincident in time with the development of slash-and-burn agriculture some 5000–10 000 years ago.
Virtual heterozygosity: The frequency of heterozygous genotypes that would be found in a random-mating diploid population, given the same allele frequencies as actually observed in a sample of haploid individuals.

http://parasites.trends.com 1471-4922/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved. PII: S1471-4922(02)02268-7

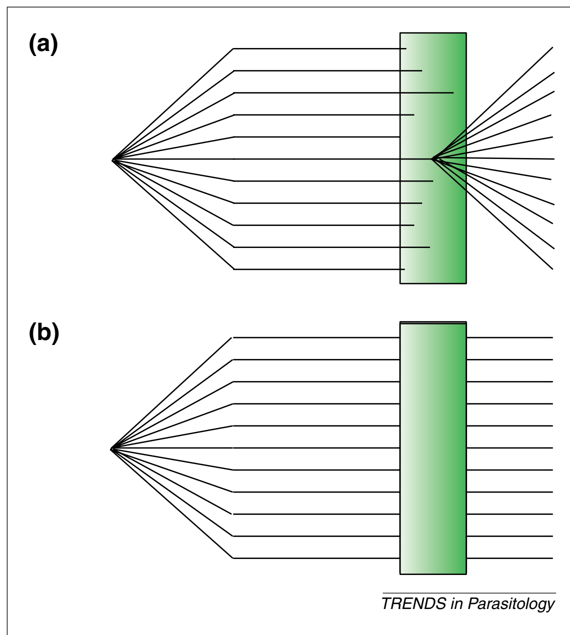

Fig. 1. Endemic malaria caused by *Plasmodium falciparum* is thought to have originated in equatorial Africa 5000–10 000 years ago. (a) Some aspects of genetic diversity among today's *P. falciparum* suggest that only a few (and perhaps only one) of the ancient lineages persisted through this period [the recent common ancestor model (RCA) hypothesis]. (b) Other aspects of genetic diversity suggest that many ancient lineages persisted through this period of time [the ancient common ancestor (ACA) hypothesis]. In (a) and (b), the time progresses from left to right.

malarias (*Plasmodium vivax*, *Plasmodium malariae* and *Plasmodium ovale*) or to plasmodial parasites isolated from other primates [5].

These relationships suggest that *P. falciparum* has been a co-evolutionary companion of humans for millions of years, perhaps even maintaining a relatively large population size throughout the early period. This early history is not incompatible with a scenario in which endemic malaria as we know it originated with the advent of slash-and-burn agriculture ~5000–10 000 years ago. Furthermore, the possibility that endemic malaria arose recently makes no specific prediction of whether the number of organisms that contributed to the evolution of endemic malaria was large (e.g. thousands of individuals) or small (e.g. fewer than ten individuals).

The RCA hypothesis asserts that endemic malaria not only arose recently, but resulted from the proliferation of a small number of haploid genomes. This hypothesis would be supported by the finding of a low level of genetic variation in most genes throughout the genome, but its support also requires that satisfactory explanations be found for certain exceptional genes that are highly variable. On the face of it, these genes support the ACA model.

Figure 1 illustrates the key distinctions between the ACA and RCA hypotheses. Both parts (Fig. 1a,b) assume that endemic malaria arose some 5000–10 000 years ago. However, in the RCA model (Fig. 1a), all but one (or a small number) of the ancient lineages became extinct either before or during the rise of endemic malaria, and those few ancestors that survived gave rise to all of the lineages of *P. falciparum* extant today. In the ACA model (Fig. 1b), many ancient lineages passed through this period and left descendants present in today's population of *P. falciparum*. The principal distinction between these two models is that the number of surviving ancestral lineages is very small in the RCA model, whereas it is very large in the ACA model.

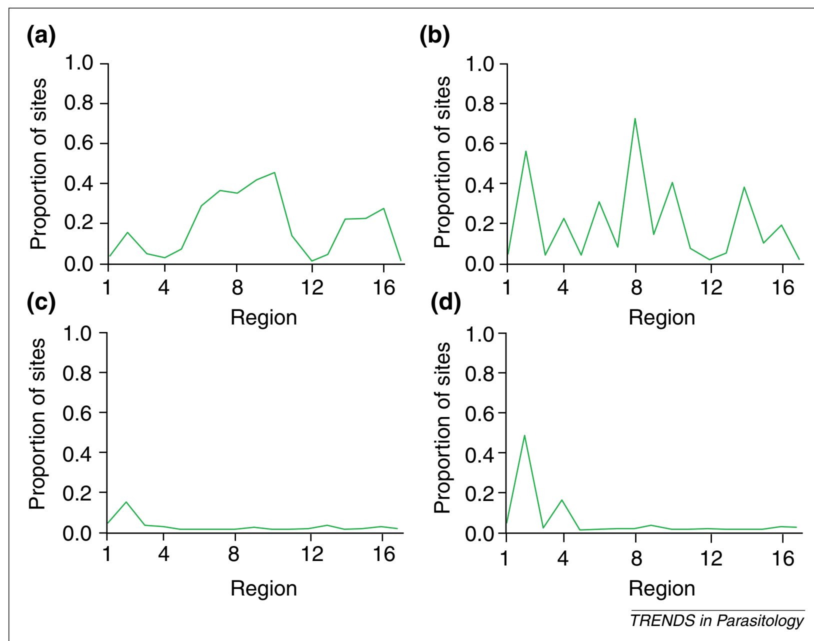

Fig. 2. Genetic diversity between and within the two major allelic classes of merozoite surface protein 1 (*Msp1*) across 17 regions within the gene [7,42]. The greatest differences are in regions 6–16. (a) Proportion of synonymous differences between classes. (b) Proportion of non-synonymous differences between classes. (c) Average proportion of synonymous differences within classes. (d) Average proportion of non-synonymous differences within classes.

Genetic polymorphisms in coding regions

Several antigenic genes in *P. falciparum* are so highly polymorphic that they are routinely used for genotyping clinical isolates [6]. Among these are those encoding merozoite surface proteins (especially MSP1 and MSP2) and circumsporozoite proteins (especially CSP1). *Msp1* is exceptionally polymorphic and includes two highly divergent classes of alleles, known as the *MAD* and *K1* types, that in some regions differ in more than 60% of their encoded amino acids. Figure 2 shows the average number of synonymous (silent) nucleotide substitutions and non-synonymous (amino acid replacement) substitutions in each of 17 regions of the MSP1 protein [7,8]. The regions range in size from ~50–200 amino acids per region, averaging about 100 amino acids. Regions 6–16 are highly divergent between the allelic classes for both synonymous and non-synomymous substitutions (Fig. 2a,b), whereas the average differences within each of the allelic classes are much smaller (Fig. 2c,d). On the basis of the divergence between the classes of alleles, Hughes has estimated that they may have diverged from a common ancestor about 48 million years ago [9]. This would make *Msp1* the oldest known polymorphism in eukaryotes, exceeding even the 20–40 million years estimated for some alleles in the primate major histocompatibility complex (MHC) class II [9].

By contrast with the high level of genetic polymorphism in *Msp1* and many other genes encoding antigenic proteins [10–13], many other genes show

exceedingly low levels of polymorphism. A recent analysis was carried out by Rich *et al.*, who examined GenBank entries for multiple sequenced alleles of ten genes in six different chromosomes [14]. These genes included regions of the polymorphic CSP antigen gene, *Csp1*, as well as the gene encoding dihydrofolate reductase, *Dhfr*, which is the target of the antimalarial drugs pyrimethamine and proguanil [15]; hence, amino acid polymorphisms among these and some other genes was to be expected. Totally unexpected was the absence of synonymous polymorphisms among 21 000 synonymous nucleotide sites [14]. For a set of coding sequences from *Escherichia coli* and *Salmonella enterica*, the comparable figure is approximately one synonymous polymorphism per 40 synonymous sites [16]. The low level of synonymous polymorphisms in *P. falciparum* was interpreted as implying that all extant *P. falciparum* derived from a single haploid progenitor, which Rich *et al.* referred to as 'malaria's Eve' [14]. From the sequence divergence between *P. falciparum* and either *P. reichenowi* or *Plasmodium berghei* (a rodent parasite), they estimated the rate of single-nucleotide mutation to be in the range 0.95×10^{-9} to 7.12×10^{-9}, and from this estimate derived a 50% confidence interval for the time to this most recent ancestor of contemporary *P. falciparum* as 5670–13 296 years.

The RCA hypothesis has recently been contested [17]. Using GenBank data for multiple alleles of 23 genes encoding nuclear proteins, many of which were also included in the previous study [14], Hughes and Verra concluded that the EFFECTIVE POPULATION SIZE of *P. falciparum* has been at least 10^5 for the past 300 000–400 000 years [17].

How can these conclusions be reconciled? One possibility for accounting for a low level of synonymous polymorphism in some genes is a strong selection for the use of preferred codons [18], such that synonymous codons are not nearly equivalent in fitness as they are in many other organisms [19,20]. Strong preferences for codon usage could relate to the fact that the *P. falciparum* genome encodes three distinct 18S RNAs that are differentially regulated in development [21]. Alternatively, some of the genes included in the Hughes and Verra analysis [17] could be subject to strong balancing selection, in spite of their attempts to exclude such genes, or could have rates or patterns of mutation that are atypical of the genome as a whole.

The RCA model received additional support from the finding of low levels of polymorphism in the complete sequences of mitochondrial DNA (5965 bp) from six isolates of *P. falciparum* [22]. Only four polymorphic sites were found, and their geographical distribution suggests spread of the species from central Africa [22]. However, mitochondrial DNA is a single, non-recombining molecule, and so its ancestry may not be representative of nuclear DNA sequences.

Genetic polymorphisms in introns

The genome size of *P. falciparum* is 30 Mb with a base composition of 82% AT [23]. It has 14 chromosomes [24] ranging in size from 0.65 to 3.4 Mb [23]. Chromosomes 2 and 3 were the first to be completely sequenced [23,25]. On the basis of these sequences, Volkman *et al.* undertook an analysis of 25 introns in each of eight independent isolates of *P. falciparum*, totaling approximately 32 000 nucleotides [26]. Approximately 17% of the intron sequences consisted of any of the 71 microsatellite repeats, defined as regions comprising eight or more consecutive repeats of any sequence of 1–8 bp. More than 50% (31) of the microsatellites are polymorphic for length variants, in which the alleles in different isolates differ in the number of repeating units. These arise by what is often called replication slippage or slip-strand mispairing, which takes place when the DNA replication complex detaches from the template strand at the location of one repeating unit and reunites with the template strand at another repeating unit [27]. Both expansions and deletions of the number of repeats can be generated by this process. The significance of the high level of microsatellite polymorphisms in *P. falciparum* is discussed below.

Surprisingly, among the eight single-nucleotide polymorphisms (SNPs) found in the introns, five were present in microsatellite regions [26]. This represents a significant excess of mutations in these regions, which may be caused by the involvement of the mismatch repair system in correcting the mismatched heteroduplex created by replication slippage [28].

Among the intronic regions not consisting of microsatellites, three SNPs were found [26]. Two of these were discounted because they were present in a single isolate in one small intron of a gene that apparently undergoes alternative splicing. This leaves one SNP in 27 336 bp of the non-microsatellite intron sequence [26], a number consistent with the low level of synonymous polymorphism discussed above [14]. Taking the intron sequences and synonymous sites at face value, together with the estimated single-nucleotide mutation rate based on divergence between *P. falciparum* and either *P. reichenowi* or *P. berghei*, the data imply that the most recent ancestor of all extant *P. falciparum* probably lived 3200–7700 years ago [26]. This value is consistent with the RCA model.

Microsatellite variation

As noted above, length variation in microsatellite repeats in *P. falciparum* is abundant. Intron sequences include microsatellite repeats spaced, on average, every 50–100 bp, and about 50% of these are polymorphic [26]. The high level of polymorphism of microsatellite repeats implies that the rate of replication slippage mutations is substantially greater than that of single-nucleotide substitutions. The finding of five new microsatellite mutations among 35 progeny of a cross genotyped for 901 microsatellite loci [29] implies a mean rate of microsatellite mutation of 1.59×10^{-4} [30], which is greater than the average rate of single-nucleotide substitutions by a factor of 10^5–10^6. However, as

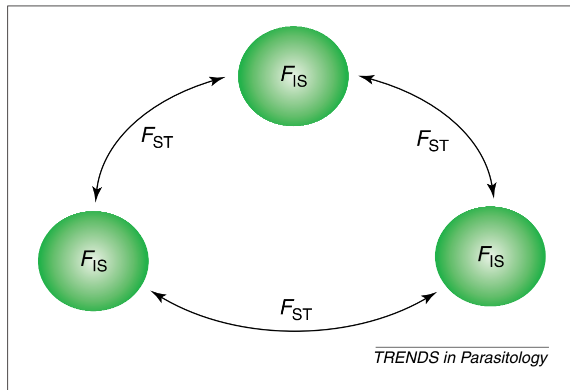

Fig. 3. Effects of population structure. F_{IS} measures the deficiency of heterozygous zygotes as a result of self-fertilization within *Plasmodium* populations, F_{ST} measures the deficiency of heterozygotes as a result of limited migration between populations. The overall deficiency of heterozygotes from both effects combined is given by $F_{IT} = 1 - (1 - F_{IS})(1 - F_{ST})$ [33].

noted, the rate of single-nucleotide substitutions is enhanced in microsatellite regions [26].

The important practical implication of widespread microsatellite polymorphisms is that microsatellites provide abundant genetic markers for drug resistance and other kinds of studies. A subset of microsatellite polymorphisms already form the framework of a genetic map derived from a laboratory cross [29,31]. Furthermore, *P. falciparum* shows a great deal of recombination per unit of physical distance, averaging approximately 1 cM per 15–30 kb. This rate of recombination per kb is somewhat smaller than that for budding yeast (1 cM per 3 kb), but much greater than that for fruit flies (1 cM per 500 kb) or humans (1 cM per 1000 kb).

Self-fertilization versus outcrossing

There is a great deal of interest in the current population size and geographical structure of *P. falciparum* [32]. However, these issues have no bearing on the RCA hypothesis because the hypothesis of a recent bottleneck in the history of the parasite population makes no prediction about its subsequent expansion or geographical subdivision.

Plasmodium falciparum is an organism with an obligate sexual cycle. Gametocytes formed in the human host are taken up in the bloodmeal of a mosquito. They complete their sexual development and fuse in pairs to form oocysts in the mosquito gut, where meiosis takes place to form haploid cells that multiply to form sporozoites that pass through the midgut wall. The sporozoites make their way to the salivary glands, from which they can continue the haploid phase of the life cycle when the mosquito bites another person.

Although the sexual cycle is obligatory, sexual recombination is not. In principle, if the population density is sufficiently low that infected mosquitoes usually take up gametocytes derived from only one haploid genome, then reproduction will be predominantly by self-fertilization. The deficiency of heterozygous genotypes, relative to a random mating population, in a local population as a consequence of inbreeding is typically measured by the inbreeding coefficient F_{IS} (HIERARCHICAL *F* STATISTICS), which is also the probability that two homologous alleles in a zygote are identical by descent (i.e. are replicas of a single ancestral allele) [33]. Direct estimates of F_{IS} based on the deficiency of heterozygous genotypes in diploid oocysts dissected from mosquitoes are in the range of 0.9 in Papua New Guinea [34] and 0.3 in Tanzania [35]. A value of $F_{IS} = 0.33$ for Tanzania was obtained by an independent method [36], which implies that about two-thirds of the zygotes in this population can form recombinant genotypes.

Because fewer heterozygous genotypes are formed with inbreeding there are also fewer opportunities for recombination, and the effective rate of recombination between markers decreases from any value, c, to a new value equal to $c' = c(1 - F_{IS})$ [37]. Because the departure from random association between genetically linked markers is expected to decrease in each generation (which is about two months for *P. falciparum*) in proportion to the effective rate of recombination [33], the degree of association between markers (linkage disequilibrium) can be used to estimate the effective rate of recombination and, therefore, F_{IS}. Studies of the pattern of linkage disequilibrium across the gene *Csp1* were inconsistent with values of F_{IS} as small as 0.3 [14,38]. However, a similar study of blocks 1–5 and 17 in the gene *Msp1* (Fig. 2) in five African populations showed essentially random association between polymorphic nucleotides separated by 1 kb, but over a larger distance in a Sudanese population with a lower level of malaria endemicity [37]. A rapid decrease in linkage disequilibrium with distance has also been reported in the gene *AMA1* for apical membrane antigen 1 [13]. Reconciliation of the *Csp1* and *Msp1* data is somewhat problematic because each is a single gene with a complex genetic structure and is probably subjected to strong selection.

Geographical differentiation among subpopulations

In addition to self-fertilization of the individuals within a local population, measured by F_{IS}, there is another type of inbreeding that results from geographical isolation and limited migration among subpopulations [33]. Relative to a situation in which the subpopulations were undergoing random mating, geographical isolation causes a reduction both in average heterozygosity and in the effective level of recombination; this deficiency attributed to geographical population structure is measured by another type of inbreeding coefficient symbolized as F_{ST} (Fig. 3).

A large-scale study of 12 microsatellite loci in 465 infections from nine locations worldwide has helped clarify the population structure of *P. falciparum* [30]. In regions of low transmission intensity, such as South America, where fewer than 20% of infected patients contain multiple genotypes, the level of inbreeding (F_{IS}) within populations is relatively high and there is also significant geographical differentiation among subpopulations ($F_{ST} = 0.3$–0.4) [30,39]. In regions of high transmission intensity, such as central Africa, where more than 45% of infected patients contain

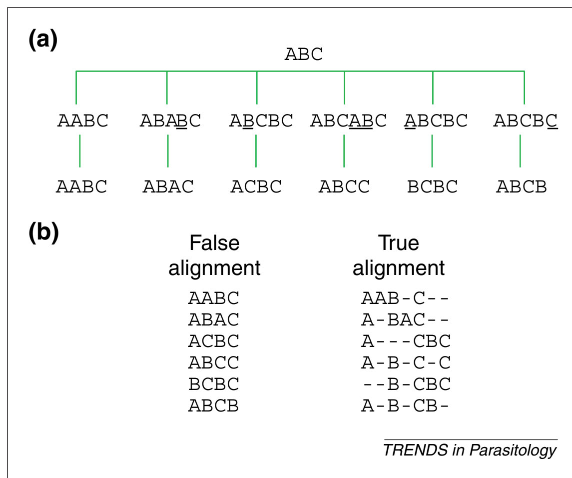

Fig. 4. Genetic diversity caused by replication slippage. (a) A, B and C represent repeat allotypes (RATs) [38]. RATs are short, repeated, in-frame coding sequences that are sufficiently similar to undergo replication slippage. The sequence ABC (top) can give rise to a diversity of allelic types through this process (middle row), and subsequent replication slippage or deletion can eliminate certain RATs (underlined), resulting in the sequences shown in the bottom row. (b) An alignment of the alleles that ignores their evolutionary history overestimates the divergence between the alleles (because A is compared with B, and so forth), relative to the true alignment.

multiple genotypes, the level of inbreeding (F_{IS}) within populations is relatively low and there is no significant geographical differentiation among subpopulations (F_{ST} = 0.007) [30]. Areas with intermediate transmission intensity, such as Thailand, have intermediate values [30]. There is also substantially more genetic variation among African isolates than elsewhere. The VIRTUAL HETEROZYGOSITY is 0.76–0.80 in African samples, 0.51–0.65 in Southeast Asian and Pacific samples, and 0.30–0.40 in South American samples (Ref. [30]). Correspondingly, the current effective population sizes in Africa are 5–10 times greater than those in South America, and in Africa they are in the range 5000–20 000 [30].

Possibilities to reconcile apparently conflicting data

With regard to the hypothesis that contemporary *P. falciparum* derives from the proliferation of a single haploid progenitor on the order of 6000 years ago, it is unlikely to be laid to rest unless a satisfactory explanation can be found that reconciles the exceptionally high polymorphism of some sequences, such as those encoding *Csp1* and *Msp1*, with the exceptionally low polymorphism of other sequences, such as synonymous sites and introns. This section considers some possibilities.

First, the low polymorphism could be due to strong selective constraints on DNA sequence at synonymous sites and in introns. However, this possibility seems a little strained in view of the high level of microsatellite length polymorphism in introns [26].

Second, there might be a single progenitor with some introgression in the early period due to backcrossing with lineages yet to be displaced. This process would introduce some ancient alleles into the same genetic background, especially those subject to strong selection.

Third, there might be no single progenitor but multiple selective sweeps [40] in which favorable mutations go rapidly to fixation and eliminate genetically linked variation in the process. In this model, many parts of the genome could be monomorphic because of recent selective sweeps, but some highly polymorphic regions would remain, especially if subject to strong balancing selection.

Fourth, mutation and selection might occur under conditions of accelerated mutation rates. The rate of replication slippage appears to be 10^5–10^6 times greater than that of single-nucleotide substitutions. In certain genes, such as *Csp1*, the coding region contains tandem repeats of similar sequences that could undergo slip-strand mispairing to generate new variants [8,41,42] (Fig. 4a). These sequences have been called repeat allotypes (RATs) [38]. Through time, different RATS can be gained or lost by replication slippage or deletion in different lineages, with the result that the alignment of allelic sequences does not correctly reflect the ancestral homologies between the RATs (Fig. 4b). The result of the false alignment is an overestimate of the amount of genetic variation generated by nucleotide substitutions. Although this process could contribute to the apparently large divergence in the repeat regions of *Csp1* alleles, nonrepeat regions of *Csp1* are also highly polymorphic, and RATS can account for only part of the extreme variation in *Msp1*.

Finally, reversion of paralogous genes to homologous alleles might occur. Variation as extreme, and apparently as ancient, as found in the *Msp1* alleles could arise by gene duplication and divergence (Fig. 5). Especially if the MAD-like and K1-like antigens are mutually antagonistic, chromosomes with paralogous copies of *MAD*-like and *K1*-type alleles would be strongly favored under conditions of low transmission. Under conditions of high transmission with many multiple infections, the selection would be less intense and, through either deletion or unequal crossing-over, the different lineages could lose one or the other paralog (Fig. 5). This would create a situation in which the *Msp1* alleles in the contemporary population are highly divergent and extremely ancient, even though

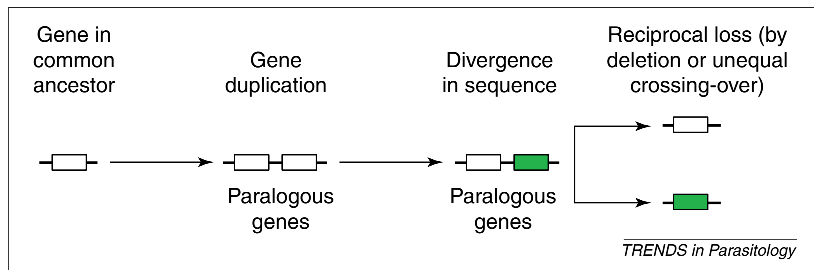

Fig. 5. Paralogous genes are related through gene duplication (indicated by open boxes), and the duplicate copies can become highly divergent in sequence (indicated by shaded boxes). However, later in evolutionary history, two chromosomal lineages could be created in which each has lost a different paralog through replication slippage or deletion. The paralogs again become alleles, but the alleles retain their high degree of sequence divergence.

Acknowledgements
We gratefully acknowledge Elena R. Lozovsky for her many helpful comments for improving the manuscript. This work was supported by NIH grant GM61351 (S.K.V., D.F.W., D.L.H., K.M.N., K.P.D. and A.E.B.), a Burroughs Wellcome Fund grant for New Initiatives in Malaria Research (D.L.H. and K.M.N.), The Wellcome Trust (K.P.D. and A.E.B.), and a grant from the Ellison Medical Foundation New Scholars Award (E.A.W.). We are grateful to Mehee Choi, Jacob Kasper, Kalsang Namgyal, Emily Lyons, Susan Thomas and Seema S. Thakore for their contributions to this research.

they derived from a single haploid genome when they were still paralogs. However, while this process might account for extreme divergence of individual genes like *Msp1*, it is unlikely to be a general explanation for high levels of polymorphism of antigenic determinants.

Conclusions

To distinguish between the RCA and ACA hypotheses, SNPs in *P. falciparum* need to be examined on a genome-wide scale in a diversity of strains with a strong representation from Africa. Efficient, high-throughput and cost-effective methods for such studies have only recently become available through, for example, high-density oligonucleotide arrays (DNA chips) suitable for assaying nucleotide matches and mismatches with a reference strain [43]. Chips containing oligonucleotides for chromosome 2, which include 4167 probes to 1 Mb, have already been produced and validated on test strains (S.K. Volkman *et al.*, unpublished), and chips with more-dense coverage of the entire genome (350 000 probes to 26 Mb) will soon be available.

If naturally occurring SNPs are rare in most genes, and resistance to a drug requires multiple amino acid replacements, then it may take the organism many years to acquire the necessary mutations and recombine them into a resistance allele. In this way, a low level of SNPs in *P. falciparum* buys time for the development of new generations of antimalarials. For example, chloroquine was an effective antimalarial for over 20 years despite its widespread use and probable under-dosage [44]. However, if a drug or immunological target contains RATs, then new coding sequences can evolve extremely rapidly by replication slippage, exacerbated by the increased rate of nucleotide substitutions in microsatellite repeats [26]. In many cases, genetic polymorphisms with partial or complete resistance may preexist in the parasite population. It is through such considerations that knowledge of the population history and genetic structure of *P. falciparum* is important in assessing its possible responses to drug and immunological control strategies.

References

1 Phillips, R.S. (2001) Current status of malaria and potential for control. *Clin. Microbiol. Rev.* 14, 208–226

2 Livingstone, F.B. (1958) Anthropological implications of sickle-cell gene distribution in West Africa. *Am. Anthropol.* 60, 533–562

3 Coluzzi, M. (1999) The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. *Parassitologia* 41, 277–283

4 Tishkoff, S.A. *et al.* (2001) Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* 293, 455–462

5 Escalante, A.A. and Ayala, F.J. (1994) Phylogeny of the malaria genus *Plasmodium* derived from rRNA gene sequences. *Proc. Natl. Acad. Sci. U. S. A.* 91, 11373–11377

6 Farnert, A. *et al.* (2001) Genotyping of *Plasmodium falciparum* infections by PCR: a comparative multicentre study. *Trans. R. Soc. Trop. Med. Hyg.* 95, 225–232

7 Tanabe, K. *et al.* (1987) Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 195, 273–287

8 Rich, S.M. *et al.* (2000) The origin of antigenic diversity in *Plasmodium falciparum*. *Parasitol. Today* 16, 390–396

9 Hughes, A.L. (1999) *Adaptive Evolution of Genes and Genomes*, Oxford University Press

10 Hughes, M.K. and Hughes, A.L. (1995) Natural selection on *Plasmodium* surface proteins. *Mol. Biochem. Parasitol.* 71, 99–113

11 Verra, F. and Hughes, A.L. (2000) Evidence for ancient balanced polymorphism at the Apical Membrane Antigen-1 (*AMA-1*) locus of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 105, 149–153

12 Conway, D.J. *et al.* (2001) Extreme geographical fixation of variation in the *Plasmodium falciparum* gamete surface protein gene *Pfs48/45* compared with microsatellite loci. *Mol. Biochem. Parasitol.* 115, 145–156

13 Polley, S.D. and Conway, D.J. (2001) Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* 158, 1505–1512

14 Rich, S.M. *et al.* (1998) Malaria's Eve: evidence of a recent population bottleneck throughout the world population of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4425–4430

15 Le Bras, J. (1999) Mechanisms and dynamics of *Plasmodium falciparum* drug resistance. *Bull. Soc. Pathol. Exot.* 92, 236–241

16 Hartl, D.L. *et al.* (2000) The glean machine: what can we learn from DNA sequence polymorphisms? In *Genomics and Proteomics* (Suhai, S., ed.), pp. 37–49, Plenum Press

17 Hughes, A.L. and Verra, F. (2001) Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. R. Soc. London B Biol. Sci.* 268, 1855–1860

18 Saul, A. (1999) Circumsporozoite polymorphisms, silent mutations and the evolution of *Plasmodium falciparum*. *Parasitol. Today* 15, 38–39

19 Hartl, D.L. *et al.* (1994) Selection intensity for codon bias. *Genetics* 138, 227–234

20 Akashi, H. (1995) Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* 139, 1067–1076

21 Li, J. *et al.* (1997) Regulation and trafficking of three distinct 18S ribosomal RNAs during development of the malarian parasite. *J. Mol. Biol.* 269, 203–213

22 Conway, D.J. *et al.* (2000) Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* 111, 163–171

23 Gardner, M.J. *et al.* (1998) Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* 282, 1126–1132

24 Triglia, T. *et al.* (1992) Towards a high-resolution map of the *Plasmodium falciparum* genome. *Parasitol. Today* 8, 225–229

25 Bowman, S. *et al.* (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* 400, 532–538

26 Volkman, S.K. *et al.* (2001) Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 293, 482–484

27 Kunkel, T.A. (1986) Frameshift mutagenesis by eucaryotic DNA polymerases *in vitro*. *J. Biol. Chem.* 261, 13581–13587

28 Strand, M. *et al.* (1993) Destabilization of tracts of simple repetitive sequences in yeast by mutations affecting DNA mismatch repair. *Nature* 365, 274–276

29 Su, X-Z. *et al.* (1999) A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* 286, 1351–1353

30 Anderson, T.J.C. *et al.* (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* 17, 1467–1482

31 Walker-Jonah, A. *et al.* (1992) An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favored linkage groups in a genetic cross. *Mol. Biochem. Parasitol.* 51, 313–320

32 Awadalla, P. *et al.* (2001) The question of *Plasmodium falciparum* population structure. *Trends Parasitol.* 17, 351–353

33 Hartl, D.L. (2000) *A Primer of Population Genetics*, Sinauer

34 Paul, R.E.L. *et al.* (1995) Mating patterns in malaria parasite populations of Papua New Guinea. *Science* 269, 1709–1711

35 Babiker, H.A. *et al.* (1994) Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitology* 109, 413–421

36 Hill, W.G. *et al.* (1995) Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet. Res.* 65, 53–61

37 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

38 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

39 Urdaneta, L. *et al.* (2001) Evidence for clonal propagation in natural isolates of *Plasmodium falciparum* from Venezuela. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6725–6729
40 Maynard Smith, J. and Haigh, J. (1974) The hitch-hiking effect of a favorable gene. *Genet. Res.* 23, 23–35
41 Jongwutiwes, S. *et al.* (1994) Allelic variation in the circumsporozoite protein of *Plasmodium falciparum* from Thai field isolates. *Am. J. Trop. Med. Hyg.* 51, 659–668
42 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
43 Winzeler, E.A. *et al.* (1998) Direct allelic variation scanning of the yeast genome. *Science* 281, 1194–1197
44 Desowitz, R.S. (1991) *The Malaria Capers: More Tales of Parasites and People, Research and Reality*, W.W. Norton

How protozoan parasites evade the immune response

Sergio Zambrano-Villa, Disney Rosales-Borjas, Julio César Carrero and Librado Ortiz-Ortiz

Protozoan pathogens such as *Plasmodium*, *Leishmania*, *Trypanosoma* and *Entamoeba* are responsible for several of the most widespread and lethal human diseases. Their successful survival depends mainly on evading the host immune system by, for example, penetrating and multiplying within cells, varying their surface antigens, eliminating their protein coat, and modulating the host immune response. Immunosuppression is sometimes caused directly by parasite products and sometimes involves antigenic mimicry, which often appears in association with parasitic diseases. However, one of the most sophisticated mechanisms of evasion is the selective activation of a subset of T helper cells.

Parasites that cause malaria, sleeping sickness, Chagas disease, amoebiasis and leishmaniasis primarily affect the populations of developing countries, causing high rates of morbidity and mortality [1]. The host's mechanisms of defense range from primary barriers to the most elaborate devices, which involve a wide variety of cells and molecules capable of the specific recognition and elimination of invasive agents. Parasites with extracellular stages are the target of the humoral immune response, whereas those with intracellular stages are susceptible to attack by the cell-mediated immunity. Despite the large amount of antigen presented by the parasite to the host and the host's immune and inflammatory response, parasites manage to survive within the host for long periods by using multiple evasion mechanisms that have been acquired during millions of years of evolution (Table 1). Parasites have evasion mechanisms that depend on factors such as their life-cycle stage, the route of penetration and the microenvironment in which they are established inside the host.

Sergio Zambrano-Villa
Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, Jalisco, México.

Disney Rosales-Borjas
Hospital Universitario Dr Miguel Oraá, Guanare, Edo. Portuguesa, Venezuela.

Julio César Carrero
Librado Ortiz-Ortiz*
Departamento de Inmunología, Instituto de Investigaciones Biomédicas, UNAM, México.
*e-mail: librado@servidor.unam.mx

Malaria

In humans, malaria is caused by *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale* and *Plasmodium falciparum*. The infection involves a complex life cycle with intra- and extracellular stages. During the very early stage of infection, the host response is poor because of the relatively low density of sporozoites and their rapid migration to the liver. In order to survive, the parasite must reach the hepatocytes and transform into merozoites, acquiring new biological features to confront the immune system. Once the *Plasmodium* parasite is established in the host, it evades the immune response (IR) by changing its surface antigens as it passes through the stages of its life cycle. Thus, each phase of the cellular cycle is associated with the expression of stage- and species-specific proteins, many of which are inserted into the membrane surface of the parasite or the red blood cells they invade.

The development of an effective IR is hampered by the fact that stage-specific proteins tend to be highly polymorphic or antigenically variable. Sequence polymorphism is common in several *Plasmodium* proteins, especially in those with extensive repetitive regions, such as the circumsporozoite protein (CSP), an antigen with multiple tandem repeats and that is located on the surface of malaria sporozoites. Antigens with multiple repeats could downregulate antibody isotype maturation and the production of high-affinity antibodies by acting as B-cell superantigens and predominantly inducing a polyclonal thymus-independent humoral response [2]. Despite the short time of circulation of the malaria extracellular phases, when an antibody response does develop against sporozoites, plasmodia overcome the response by sloughing off their surface CSP coat, rendering the antibodies ineffective. The versatility of CSP is also evident in its RNA-binding motifs, which block RNA translation in mammalian cells [3].

Other proteins, such as the *P. falciparum* merozoite surface protein (MSP-1) 1, induce 'blocking antibodies' that bind anywhere on MSP-1, preventing

 1471-4922/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved. PII: S1471-4922(02)02289-4