Copyright © 1998 by the Genetics Society of America

Letter to the Editor

Ancient Polymorphism and the Hypothesis of a Recent Bottleneck in the Malaria Parasite *Plasmodium falciparum*

Austin L. Hughes and Federica Verra

Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania 16802

Rich *et al.* (1998) have recently proposed that the current world population of the human malaria parasite *Plasmodium falciparum* consists of individuals descended from a single haploid genotype or strain (malarial Eve) that occurred between 24,500 and 57,500 years ago. They base this conclusion on an examination of synonymous and nonsynonymous nucleotide polymorphisms at 10 loci. At each of these loci, the authors examined published allelic sequences, with allele numbers ranging from 2 to 25 per locus; they reported some nonsynonymous polymorphisms but no synonymous polymorphisms at any locus. This conclusion is surprising because it ignores previous reports of extensive synonymous polymorphisms at certain loci of *P. falciparum* encoding surface proteins (Hughes 1991, 1992; Hughes and Hughes 1995a; Jongwutiwes *et al.* 1994). Furthermore, it is inconsistent with estimates that certain polymorphisms at loci encoding immunogenic proteins of *P. falciparum* have been maintained for millions of years (Hughes 1992, 1993; Hughes and Hughes 1995b).

The circumsporozoite protein (CSP) is a surface protein of the sporozoite, the infective stage of the malaria parasite. Examination of rates of synonymous and nonsynonymous nucleotide substitution at this locus showed a significantly enhanced rate of nonsynonymous nucleotide substitution in the region of the gene encoding peptides bound by host major histocompatibility complex (MHC) molecules and presented to T cells [T cell epitopes (TCE); Good *et al.* 1988; Hughes 1991]. This is evidence that CSP polymorphism is maintained by balancing selection relating to the evasion of host immune recognition (Hughes 1991). The recent discovery of altered peptide ligand (APL) antagonism between certain allelic forms of these TCE in *P. falciparum* (Gilbert *et al.* 1998) suggests a possible mechanism for such selection. When antagonistic forms of the TCE are present in the same host, an effective immune response is prevented (Gilbert *et al.* 1998). Therefore APL antagonism will give rise to overdominant selection, favoring *CSP* heterozygosity at the level of the zygote that occurs in the mosquito intermediate host and gives rise to the infective sporozoites.

It is well known that balancing selection can maintain polymorphisms for much longer than neutral polymorphisms are expected to persist (Takahata and Nei 1990). A classic example involves the MHC genes of vertebrates. The MHC genes are characterized by a greatly enhanced rate of nonsynonymous nucleotide substitution in the codons encoding the peptide-binding region (PBR) of the molecule (Hughes and Nei 1988, 1989; Hughes *et al.* 1994). This pattern of nucleotide substitution is evidence that some form of balancing selection (such as overdominant selection) acts to maintain diversity at the amino acid level in the PBR. Diversity in the PBR is apparently selectively favored because MHC heterozygosity enhances immune surveillance by enabling the individual to bind and present a wide array of foreign peptides (Hughes and Nei 1988; Hughes and Hughes 1995a). MHC loci are also characterized by long-lasting polymorphisms, which have been maintained for millions of years and predate speciation events (Klein and Takahata 1990). The fact that there are numerous ancient alleles at human MHC loci is evidence that the human population could not have gone through a recent extreme bottleneck; rather the long-term effective population size of humans has been on the order of 10^4–10^5 for the past 5 million years or so (Klein *et al.* 1990; Ayala 1995; Takahata *et al.* 1995).

Because the *CSP* locus is subject to a similar sort of balancing selection to that seen at the MHC, it would not be surprising to find that polymorphism at this locus also has been maintained for a long time. The available data support this hypothesis. For example, we computed the mean number of synonymous substitutions per site (d_S; Nei and Gojobori 1986) in the 5′ and 3′ nonrepeat regions between 13 distinct CSP alleles of *P. falciparum* and one of *Plasmodium reichenowi*, a parasite of chimpanzees. (The central tetrapeptide repeat region of the gene was excluded because it cannot be reliably aligned among alleles; see Jongwutiwes *et al.* 1994.) Mean d_S

Corresponding author: Austin L. Hughes, Department of Biology, 208 Mueller Laboratory, The Pennsylvania State University, University Park, PA 16802. E-mail: austin@hugaus3.bio.psu.edu

between *P. falciparum* and *P. reichenowi* was 0.035 ± 0.016 SE. On the other hand, d_S between the two most distant *P. falciparum* alleles was 0.015 ± 0.011. Assuming that *P. falciparum* and *P. reichenowi* diverged when human and chimpanzee diverged and that this occurred about 5 million years ago (mya), we estimate the divergence time of the most divergent known *CSP* alleles in *P. falciparum* at 2.1 ± 1.5 million years. The standard error of the estimate is large because the number of sites involved is small; nonetheless the range of two standard errors around the mean (0.6–3.6 million years) falls well beyond the very recent date for malarial Eve proposed by Rich *et al.* (1998). To accumulate the degree of synonymous difference observed between the two most divergent *P. falciparum CSP* alleles in the roughly 60,000 years proposed by Rich *et al.* (1998), a rate of synonymous substitution of about 1×10^{-7} substitutions/site/year would be required. This is two orders of magnitude higher than typical rates of synonymous substitution in eukaryotes (Li 1997).

It might seem surprising that we obtained nonzero values of d_S for the *CSP* alleles, when Rich *et al.* (1998) counted no synonymous differences in the same data set. The difference is due to their method of counting and illustrates the pitfalls of merely counting codon differences rather than estimating the number of differences by a method like Nei and Gojobori's (1986). For example, in the 5′ nonrepeat region of the *CSP* gene of *P. falciparum*, there is a codon position at which certain alleles have ATC (isoleucine), while the rest have ACT (threonine). Rich *et al.* (1998) presumably counted this as a nonsynonymous difference. However, Nei and Gojobori's method considers the pathways of point mutation by which one can get from ATC to ACT (or vice versa). The possible intermediates are ATT (isoleucine) and ACC (threonine). Thus, by whichever mutational pathway we proceed from ATC to ACT (or vice versa), we must accumulate one synonymous mutation and one nonsynonymous mutation. By ignoring mutational pathways, Rich *et al.* (1998) came to the misleading conclusion that no synonymous substitutions have occurred among *CSP* alleles.

Merozoite surface antigen-1 (MSA-1) is another antigenic surface protein of Plasmodium. It is encoded by a single polymorphic locus and expressed on the surface of the merozoite, the stage of the parasite that invades host red blood cells (Hall *et al.* 1984). The evolutionary history of *MSA-1* alleles is complex because of numerous past events of interallelic recombination; thus different gene regions have different evolutionary histories (Hughes 1992). (Note that interlocus recombination cannot have been a factor here because there are no related loci in the *P. falciparum* genome; Hall *et al.* 1984.) In one extensive region [called region 6 by Hughes (1992)], the known *MSA-1* alleles of *P. falciparum* form two distinctive families, which are highly divergent from each other. Hughes (1992, 1993) estimated that these two families diverged from each other at least 35 mya, making this one of the oldest polymorphisms known in any organism. Mean d_S in region 6 between the two families of *P. falciparum MSA-1* alleles is 0.681 ± 0.122 (Hughes 1992). Thus, alleles from these two families are nearly as divergent at synonymous sites as are orthologous genes of human and mouse. To accumulate this degree of difference in 60,000 years would require a synonymous substitution rate of 6×10^{-6} substitutions/site/year—over three orders of magnitude greater than the typical eukaryotic rate.

The genomes of Plasmodium parasites are extraordinarily AT rich. The 334 coding regions of *P. falciparum* now in the database are 71.7% AT (Nakamura *et al.* 1998). This nucleotide content bias affects codon usage, with codons having A or T in the third position being preferentially used. Of over 200,000 *P. falciparum* codons in the database, 83.6% have A or T in the third position. It is known from other systems that a strong codon bias will lower the rate of synonymous substitution (Sharp and Li 1989; Ticher and Graur 1989; Sharp 1991). Thus, the rate of synonymous substitution is expected to be lower in *P. falciparum* than in organisms such as humans, whose codon usage is less biased. The fact that the rate of synonymous substitution is reduced may explain the observation of nonsynonymous but not synonymous polymorphisms at certain loci in this species (Hughes 1991; Hughes and Hughes 1995b).

The existence of long-lasting polymorphisms and high levels of intraspecific synonymous polymorphism in *P. falciparum* is strong evidence against the hypothesis of malarial Eve. Most of the loci examined by Rich *et al.* (1998) are housekeeping genes, encoding metabolic enzymes, chaperone proteins, and so forth. The existence of balancing selection at such loci is not expected. Most are probably evolving neutrally. It is not expected that the pattern of coalescence at a neutral locus will be the same as that at a locus under balancing selection; rather, polymorphism at the former will be much less long-lasting than at the latter (Takahata and Nei 1990). Of course, if recent directional selection has led to fixation of an allele at any of the loci studied by Rich *et al.* (1998), the coalescence time of current alleles could be even more recent than that at a neutrally evolving locus. In either case, the coalescence pattern at one or a few loci cannot be used to estimate the population history of a species (Takahata *et al.* 1995). Alleles at *CSP*, *MSA-1*, and other loci that are under balancing selection clearly have much earlier coalescence times than do alleles at neutrally evolving loci, just as in humans MHC loci have much earlier coalescence times than do mitochondrial DNA.

The fact that long-lasting polymorphisms have been maintained at certain loci of *P. falciparum* but not at others is also strong evidence that interlocus recombination has been an important factor in the evolution of this species, contrary to the hypothesis of Rich *et al.*

(1997) and Ayala (1998). If haploid genomes did not recombine, polymorphism at all loci would hitchhike together, causing polymorphism even at selectively neutral loci to be as ancient as that at *MSA-1.* The evidence for past interallelic recombination at both *CSP* and *MSA-1* (Hughes 1992; Jongwutiwes *et al.* 1994) loci also shows the importance of recombination in this sexually reproducing diploid.

In conclusion, there is strong evidence against a recent extreme population bottleneck in *P. falciparum*, involving a single haploid genotype, as proposed by Rich *et al.* (1998). At this point we cannot rule out the possibility of a moderate bottleneck, which would have allowed continued maintenance of polymorphism at the *CSP*, *MSA-1*, and other loci subject to balancing selection (Hughes and Hughes 1995b). We probably do not yet have sufficient data to estimate reliably the long-term effective population size of *P. falciparum*, as has been done for humans on the basis of polymorphism at MHC loci (Takahata *et al.* 1995). To address this question, we need extensive samples of sequence polymorphism from worldwide populations of *P. falciparum.* It is instructive that the only study that attempts to assay *CSP* allelic variation in a local population of this species—in Thailand by Jongwutiwes *et al.* (1994)—revealed several previously undescribed alleles, in fact nearly doubling the number of known alleles at this locus. This suggests that many more alleles at this locus remain to be discovered, which in turn would imply a substantial long-term effective population size. We predict that similar studies in other parts of the species' range will reveal extensive polymorphism, particularly in Africa, where *P. falciparum* presumably originated.

LITERATURE CITED

Ayala, F. J., 1995 The myth of Eve: molecular biology and human origins. Science **270:** 1930–1936.

Ayala, F. J., 1998 Is sex better? Parasites say "no." Proc. Natl. Acad. Sci. USA **95:** 3346–3348.

Gilbert, S. C., M. Plebanski, S. Gupta, J. Morris, M. Cox *et al.*, 1998 Association of malaria parasite population structure, HLA, and immunological antagonism. Science **279:** 1173–1177.

Good, M. F., D. Pombo, L. A. Quakyi, E. M. Riley, R. A. Houghten *et al.*, 1988 Human T-cell recognition of the circumsporozoite protein of *Plasmodium falciparum*: immunodominant T-cell domains map to the polymorphic regions of the molecule. Proc. Natl. Acad. Sci. USA **85:** 1199–1203.

Hall, R., J. E. Hyde, M. Goman, D. L. Simmons, I. A. Hope, M. Mackay *et al.*, 1984 Major surface antigen gene of a human malaria parasite cloned and expressed in bacteria. Nature **311:** 379–382.

Hughes, A. L., 1991 Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics **127:** 345–353.

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.

Hughes, A. L., 1993 Coevolution of immunogenic proteins of *Plasmodium falciparum* and the host's immune system, pp. 109–127 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer, Sunderland MA.

Hughes, A. L., and M. K. Hughes, 1995a Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics **42:** 233–243.

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

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at MHC class I loci reveals overdominant selection. Nature **335:** 167–170.

Hughes, A. L., and M. Nei, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86:** 958–962.

Hughes, A. L., M. K. Hughes, C. Y. Howell and M. Nei, 1994 Natural selection at the class II major histocompatibility complex loci of mammals. Philos. Trans. R. Soc. Lond. B Biol. Sci. **345:** 359–367.

Jongwutiwes, S., K. Tanabe, M. K. Hughes, H. Kanbara and A. L. Hughes, 1994 Allelic variation in the circumsporozoite protein of *Plasmodium falciparum* from Thai field isolates. Am. J. Trop. Med. Hyg. **51:** 659–668.

Klein, J., and N. Takahata, 1990 The major histocompatibility complex and the quest for origins. Immunol. Rev. **113:** 5–25.

Klein, J., J. Gutknect and N. Fischer, 1990 The major histocompatibility complex and human evolution. Trends Genet. **6:** 7–11.

Li, W.-H., 1997 *Molecular Evolution.* Sinauer, Sunderland, MA.

Nakamura, Y., T. Gojobori and T. Ikemura, 1998 Codon usage tabulated from the international DNA sequence databases. Nucleic Acids Res. **26:** 334.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol. Biol. Evol. **3:** 418–426.

Rich, S. M., R. R. Hudson and F. J. Ayala, 1997 *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. Proc. Natl. Acad. Sci. USA **94:** 13040–13045.

Rich, S. M., M. C. Licht, R. R. Hudson and F. J. Ayala, 1998 Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum.* Proc. Natl. Acad. Sci. USA **95:** 4425–4430.

Sharp, M. M., 1991 Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33:** 23–33.

Sharp, P. M., and W.-H. Li, 1989 On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Evol. **28:** 398–402.

Takahata, N., and M. Nei, 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics **124:** 967–978.

Takahata, N., Y. Satta and J. Klein, 1995 Divergence time and population size in the lineage leading to modern humans. Theor. Pop. Biol. **48:** 198–211.

Ticher, A., and D. Graur, 1989 Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J. Mol. Evol. **28:** 286–298.

Communicating editor: N. Takahata