

UC Irvine

UC Irvine Previously Published Works

Title

The Recent Origin of Allelic Variation in Antigenic Determinants of Plasmodium falciparum

Permalink

<https://escholarship.org/uc/item/8144n79h>

Journal

Genetics, 150(1)

ISSN

0016-6731

Authors

Rich, Stephen M
Ayala, Francisco J

Publication Date

1998-09-01

DOI

10.1093/genetics/150.1.515

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Letter to the Editor

The Recent Origin of Allelic Variation in Antigenic Determinants of *Plasmodium falciparum*

Stephen M. Rich* and Francisco J. Ayala†

*Department of Biology, University of Rochester, Rochester, New York 14627 and

†Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525

STUDIES of genetic variability of *Plasmodium falciparum* have often focused on antigen proteins and the genes that encode them. A consistent observation of these studies is that *P. falciparum* populations exhibit high levels of genetic polymorphism in their antigenic determinants, such as the genes encoding surface proteins of the merozoite (*Msa-1*, *Msa-2*) and the sporozoite (*Csp*). What is the age and derivation of this variation? It has often been concluded that *P. falciparum*'s allelic polymorphism is old on an evolutionary scale and that allelic variants undergo a process of mixis during the obligate sexual phase of the parasite life cycle. It has been clearly shown that these antigenic genes are under strong positive selection for evasion of human immune response (Hughes 1991, 1992; Escalante *et al.* 1998). Because natural selection causes nonuniform rates of nucleotide substitution that may be greatly accelerated, the study of such genes confounds efforts to estimate the age of the polymorphisms. Nucleotide substitutions that evolve by a neutral process are much more reliable for determining the age of allelic variants.

Recently we examined 10 gene loci (1 antigenic and 9 nonantigenic) and determined that there are no polymorphisms at silent nucleotide sites, *i.e.*, within those codons for which no amino acid replacement has occurred (Rich *et al.* 1998). Based on the observation of no polymorphism at 30,973 silent sites (10,912 fourfold and 20,061 twofold sites), we concluded with 95% confidence that the set of isolates examined, which are representative of the global geographic distribution of the parasite, must have derived from a most recent common ancestor within the last 57,500 years, although the actual time of origin might be an order of magnitude more recent (Rich *et al.* 1998).

Hughes and Verra (1998, this issue) argue that our conclusions are inconsistent with estimates that certain polymorphisms have been maintained in *P. falciparum* for millions of years. They make three points: (1) The

abundance of nucleotide polymorphisms in the antigenic peptide regions of the circumsporozoite protein (CSP) is evidence that the polymorphism is maintained by balancing selection. The age of the most divergent CSP alleles of *P. falciparum* is estimated to be 2.1 ± 1.5 million years. The merozoite surface antigen-1 (MSA-1) is also estimated to consist of ancient alleles. (2) The genomes of Plasmodium parasites are AT rich, which lowers the rate of synonymous substitution. (3) Most loci examined by Rich *et al.* (1998) are housekeeping enzymes, encoding metabolic enzymes and chaperone proteins, which are most likely evolving neutrally. We will respond to these objections in turn.

Hughes and Verra (1998) contend that balancing natural selection can maintain allele polymorphisms for very long times. They cite the major histocompatibility complex of vertebrates as a well-documented example of long-lasting polymorphisms that have been maintained for millions of years and predate speciation. It has been shown, for example, that some human alleles are phylogenetically more closely related to gorilla or macaque alleles than they are to other human alleles. The old age of these alleles precludes any extreme bottleneck in human populations, which are estimated to have long-term effective population size of the order of 10^4 – 10^5 for millions of years (Klein and Takahata 1990; Ayala 1995; Takahata *et al.* 1995; Ayala and Escalante 1996). All of this is, of course, correct. But the issue is not whether natural selection can preserve long-term polymorphisms, which it can, but whether or not the CSP polymorphisms are ancient. Natural selection in large populations also can rapidly increase the frequency of favorable mutations and maintain them at high frequencies.

How can we know whether the CSP alleles of *P. falciparum* are ancient or recent? If the alleles are ancient they will differ among themselves not only in the amino acid replacements targeted by selection to escape the host's immune response, but also in selectively neutral or nearly neutral substitutions. If the CSP alleles of *P. falciparum* have been around for hundreds of thousands or millions of years (Hughes and Verra 1998), they should have accumulated multiple synonymous

Corresponding author: Francisco J. Ayala, Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525. E-mail: fjayala@uci.edu

substitutions. They have not (see below); and this shows that the CSP alleles are of recent origin.

We examined 25 CSP alleles of widely divergent geographic origins and found no silent polymorphisms in a total of 5373 synonymous sites examined in the nonrepeat regions of the gene (Rich *et al.* 1998). There are, however, numerous CSP silent polymorphisms between *P. falciparum* and *P. reichenowi* (Escalante *et al.* 1998, see Table 7). Hughes and Verra (1998) calculate $d_s = 0.015$, but even this low level of silent polymorphism may be an overestimate. There is one polymorphic codon among the CSP alleles that involves two substitutions, in the second and in the third position (ACT,Thr \leftrightarrow ATC,Ile, at site 45 in Rich *et al.* 1997, Table 2). We treat this change as nonsynonymous even though one single substitution in the second position is sufficient to effect the amino acid replacement, because the substitution in the third position has not occurred independently of the amino acid replacement. Some methods of estimating synonymous substitutions between sequences (such as d_s of Nei and Gojobori 1986 used by Hughes and Verra 1998) will, however, yield nonzero values in such cases. Distance measures should not be used without taking into account the underlying events.

The MSA-1 is an antigen protein expressed on the surface of the merozoite, the parasite's stage that invades the host's red blood cells. This protein is encoded in *P. falciparum* by a gene consisting of about 1018 codons. The MSA-1 alleles form two distinct families [namely, the MAD-20 and Wellcome allele types of Tanabe *et al.* (1987) and Escalante *et al.* (1998)]. According to Hughes and Verra (1998) these two families diverged at least 35 mya, on the grounds that the mean genetic distance between the two families in region 6 (the most divergent and largest region, with about 628 codons) is very large, with $d_s = 0.681 \pm 0.122$ (Hughes 1992). These grounds are correct, but the same issues arise as for CSP. If the two sets of alleles would have diverged long ago, synonymous substitutions would have accumulated within each family. If we consider separately the two groups of alleles (three alleles in each group) in Hughes (1992), we obtain for the region in question $d_s = 0.000$ for group I and $d_s = 0.003 \pm 0.002$ for group II, showing that the allele families are of very recent origin, precisely the opposite of the claim they make. Note again that d_s estimates can be biased upward for the reason noted above, and more generally whenever the assumptions of the Jukes-Cantor model are violated (Ina 1995). The relevant assumption is that all nucleotide substitutions are equally likely, which is not the case in *P. falciparum*, given its high AT content (see below). If the split between the MAD-20 and Wellcome sets of alleles occurred long ago, the absence of silent polymorphisms within each set could only have resulted from a very recent bottleneck. It seems likely, however, that the apparent large divergence between the two allele sets in region 6 (nearly 70% of all sites) but not

elsewhere in this gene is an artifact of nonhomologous alignment within the region. The virtual absence of polymorphism within each of the MSA-1 allele families has been noted by Escalante *et al.* (1998), who analyzed a 42-kD fragment of MSA-1 for which 40 nucleotide sequences are available. The average nucleotide diversity (synonymous and nonsynonymous) per site within each allelic family is quite low, namely $\pi = 0.004$ (30 MAD-20 alleles) and $\pi = 0.001$ (10 Wellcome alleles) (Escalante *et al.* 1998).

The genome of Plasmodium parasites is AT rich. In *P. falciparum*, the AT content is 71.7% overall and 83.6% in the third position (Nakamura *et al.* 1998). This AT excess lowers the rate of synonymous substitution, but does not altogether eliminate it (Sharp and Li 1989; Ticher and Graur 1989; Sharp 1991). There are three reasons why AT richness cannot account for the virtual absence of synonymous polymorphism in *P. falciparum*. (1) In the case of fourfold redundant codons, the bias is for codons terminating in A or T at the expense of G or C, but this does not by itself impede A \leftrightarrow T (or G \leftrightarrow C) mutations. The mean T/A ratio in 3rd position for fourfold codons is 1.1, while the ratio for C/G is 2.2 (based on data from Nakamura *et al.* 1998 for 312 complete coding sequences and 69,120 codons total). Thus, while the A/T \leftrightarrow G/C changes are restricted and G \leftrightarrow C may also be restricted, there is no evidence that A \leftrightarrow T changes are restricted. (2) Other Plasmodium species are also AT rich and should have the same mutational constraints as *P. falciparum*, yet they exhibit abundant synonymous intraspecific, as well as interspecific variation. The constraints imposed by AT richness do not substantially affect the incidence of synonymous substitutions in these species. (3) Comparisons between *P. falciparum* and its closest relative, *P. reichenowi*, at each of five genes for which data are available in both species indicate high numbers of synonymous substitutions (average $K_s = 0.072$ and $K_n = 0.046$, for synonymous and nonsynonymous substitutions, respectively, calculated from Escalante *et al.* 1998, Table 7). Synonymous substitutions have accumulated between these two lineages over the last 5–8 million years since their divergence, AT richness notwithstanding.

The third objection raised by Hughes and Verra (1998) is that most of the loci examined by Rich *et al.* (1998) code for enzymes and other proteins that may not be subject to balancing selection. "Most," they write, "are probably neutral." They may or may not be neutral, but for the present purposes the more nearly neutral they are, the better. We have limited our investigation to synonymous substitutions, precisely because these are neutral or nearly neutral. Inferences about coalescence processes and dates can, thus, be confidently made, because properties of neutral evolution are generally well known, and there is no need to make assumptions about the value, direction, and constancy of selective coefficients, all of which are at best tentative.

How often meiotic recombination occurs in *P. falciparum* is a matter of debate. We have argued that CSP variation comes about by a clonal rather than sexual process (Rich *et al.* 1997, 1998; Ayala 1998). The 10 genes we have investigated are located on at least six different chromosomes and all yield the same result, namely, absence of synonymous substitutions that supports derivation from a recent single ancestral gene in every case. Collectively, the 10 loci warrant the inference that the world populations of *P. falciparum* have derived from one cenacestor that lived only a few thousand years ago, with 24,000 to 57,000 years as the upper 95% confidence boundaries for the cenacestor's age.

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.
- Ayala, F. J., and A. A. Escalante, 1996 The evolution of human populations: a molecular perspective. *Mol. Phylogenet. Evol.* **5**: 188–201.
- Escalante, A. A., A. A. Lal and F. J. Ayala, 1998 Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* **149**: 189–202.
- 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., and F. Verra, 1998 Ancient polymorphism and the hypothesis of a recent bottleneck in the malaria parasite *Plasmodium falciparum*. *Genetics* **150**: 511–513.
- Ina, Y., 1995 New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *J. Mol. Evol.* **40**: 190–226.
- Klein, J., and N. Takahata, 1990 The major histocompatibility complex and the quest for origins. *Immunol. Rev.* **113**: 5–25.
- 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 nucleotide 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, M. M., and W.-H. Li, 1989 On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* **28**: 398–402.
- Takahata, N., Y. Satta and J. Klein, 1995 Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* **48**: 198–211.
- Tanabe, K., M. Mackay, M. Goman and J. G. Scaife, 1987 Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **195**: 273–287.
- 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