

Recent Origin of *Plasmodium falciparum* from a Single Progenitor

Sarah K. Volkman,^{1,2*} Alyssa E. Barry,^{1,3*} Emily J. Lyons,^{1,3}
Kaare M. Nielsen,^{1,4,5} Susan M. Thomas,^{1,2} Mehee Choi,^{1,4}
Seema S. Thakore,^{1,2} Karen P. Day,^{1,3} Dyann F. Wirth,^{1†}
Daniel L. Hartl^{1,4}

Genetic variability of *Plasmodium falciparum* underlies its transmission success and thwarts efforts to control disease caused by this parasite. Genetic variation in antigenic, drug resistance, and pathogenesis determinants is abundant, consistent with an ancient origin of *P. falciparum*, whereas DNA variation at silent (synonymous) sites in coding sequences appears virtually absent, consistent with a recent origin of the parasite. To resolve this paradox, we analyzed introns and demonstrated that these are deficient in single-nucleotide polymorphisms, as are synonymous sites in coding regions. These data establish the recent origin of *P. falciparum* and further provide an explanation for the abundant diversity observed in antigen and other selected genes.

Plasmodium falciparum causes the most virulent form of human malaria, resulting in 200 million to 300 million infections and 1 million to 3 million deaths annually (1). Genetic variation within this human pathogen facilitates its transmission and pathogenesis and limits efforts to combat the disease. In the case of *P. falciparum*, the issue is caught up in controversy (2, 3). Genetic variation in proteins for antigenic determinants (4), drug resistance (5–8), and pathogenesis is abundant (9–13), whereas DNA variation at silent (synonymous) sites in coding sequences appears virtually absent (14). Nevertheless, microsatellite variation within and among subpopulations is widespread (15, 16). These discrepancies could be reconciled if all extant *P. falciparum* derived from a single progenitor that spread through the human population within the past few thousand years (14). Alternatively, codon usage may be so constrained that synonymous mutations are eliminated by selection. To resolve these possibilities, we analyzed 25 introns from eight independent isolates and found only eight single-nucleotide polymorphisms (SNPs), five of which occur within microsatellite repeats. In contrast, microsatellite polymor-

phisms are common within introns. Our results support the recent progenitor hypothesis and imply a high mutation rate for the creation of microsatellite repeats.

We chose introns for our analysis because introns are subject to selective constraints that differ from those for codon usage. Apart from pseudogenes, introns are among the most rapidly evolving sequences in eukaryotes (17), and they are of general utility in studies of population structure (18). We chose to analyze introns from general metabolic or housekeeping genes on chromosome 2 (19) and chromosome 3 (20), the first chromosomes completely sequenced, and examined these regions in each of eight independent isolates from diverse geographic regions including Africa, Honduras, Southeast Asia, and Papua New Guinea (21). For each intron, the target sequence was amplified by the polymerase chain reaction and the products were cloned and sequenced (21). To guard against polymerase incorporation error, we sequenced each intron in both directions from each of three clones derived from each of three independent amplifications (21).

The results demonstrate that microsatellite variation in *P. falciparum* is widespread within introns (Table 1), which is consistent with previous results (16, 22). Across all introns there are 71 microsatellite repeats, which we define as a region of eight or more tandem repeats of a sequence 1 to 8 base pairs (bp) in length. Among the microsatellite repeats, 36 (51%) are monomorphic and 35 (49%) are polymorphic with two or more alleles in the sample [Web fig. 1 (21)]. The tremendous amount of genetic diversity generated by the alteration of these repetitive sequences is illustrated by the microsatellite

genotype of each isolate with respect to these polymorphisms [Web fig. 2 (21)]. The genotype of each isolate is unique, even among contemporary isolates from the same geographic region. The potential for microsatellite diversity in *P. falciparum* is also evident by the number of distinct alleles for each polymorphic microsatellite repeat (21). These data support a high rate of microsatellite mutation within introns, presumably as a consequence of replication slippage (23, 24).

In contrast to the microsatellite variation, SNPs within the introns are rare (Table 1). Altogether, we observed eight SNPs in the introns. Among these, five were located within microsatellite polymorphisms. Across all of the 4217 bp of intron sequence (counting with respect to the 3D7 reference sequence), only 800 bp (19%) are located in polymorphic microsatellites. The excess of SNPs in the microsatellite repeats is, therefore, highly significant ($P = 0.008$, Fisher's exact test). This finding strongly suggests that the process of replication slippage that generates microsatellite variation (24) also increases the rate of single-nucleotide substitutions. Therefore, we have ignored the five SNPs associated with microsatellite polymorphisms in estimating the time since the most recent common ancestor of all extant *P. falciparum*. The excess SNPs within polymorphic microsatellite repeats may also explain the relatively high frequency of synonymous polymorphisms within amino acid repeat sequences in certain proteins, such as the circumsporozoite protein (24). Introduction of SNPs within repetitive sequence occurs in both introns and exons, but selective constraints on coding sequences results in fewer polymorphisms within coding regions of the genome. Our findings suggest that antigenic variation associated with these repeated amino acid sequences has occurred within *P. falciparum*, rather than by lateral transfer or some other mechanism.

Discounting the intron sequence located in microsatellite polymorphisms, we sequenced 3417 bp in each of eight isolates (total of 27,336 bp) and identified only one certain SNP. The remaining two SNPs are found in one small intron predicted by GlimmerM (25) and only in the D6 isolate, where there is evidence of alternative splicing [Web fig. 3 (21)] within the aspartyl protease gene. Therefore, we present the statistical analysis both with and without the D6 SNPs. Combining our data with those from Rich *et al.* (14), and assuming that the rate of nucleotide substitution in unique intron sequence and monomorphic microsatellites is equal to that for fourfold-degenerate sites in coding regions, we estimate the age of the most recent common ancestor (MRCA) of all extant *P. falciparum* to be in the range of 3200 to 7700 years [the estimate was obtained with equa-

¹The Harvard-Oxford Malaria Genome Diversity Project. ²Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA. ³Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, UK. ⁴Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. ⁵Department of Botany, Norwegian University of Science and Technology, N-7491, Trondheim, Norway.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: dfwirth@hsph.harvard.edu

REPORTS

Table 1. Summary of microsatellite (MS) repeat polymorphisms and SNPs observed in *P. falciparum*. The number of MS repeats and SNPs for each of 25 introns predicted within nine housekeeping genes is shown. The length of the intron analyzed is indicated in base pairs (bp).

Gene	Intron	Length (bp)	Number of MS repeat polymorphisms	SNPs within MS repeats*	SNPs outside MS repeats†	Other‡	Reference§
Ribosomal protein L37A	I	196	4	1 ^a	0		AE001396
	II	98	1		0		
Chromatin binding protein	I	252	2		0		AE001422
Clathrin coat assembly	I	126	4	1 ^b	0		AE001419
	II	200	1		0		
Ser/Thr protein kinase	I	505	0		0	1 ^h	AE001405
	II	168	1		0		
Ubiquitin conjugating enzyme	I	182	1		0		AL034558
Aspartyl protease	I	195	2	1 ^c	0		AL008970
	II	75	0		0		
	III	126	1	1 ^d	0		
	IV	99	2		1 ^f		
	V	173	1		0		
	VI	138	0		0		
	VII	107	0		0	1 ⁱ	
	VIII	79	1		0		
	IX	169	2		0		
	X	113	1		0		
	XI	104	3		2 ^g	1 ^j	
	XII	112	1		0		
	XIII	101	1		0		
	XIV	208	4		0		
Acyl carrier protein	II	175	1		0		AE001391
WEB-1 ortholog, WD40	I	255	2	1 ^e	0		AE001407
RRM-type RNA binding	I	150	0		0		AE001383
Totals	25	4106	36	5	3	3	

*A total of 35 MS repeat polymorphisms were identified, four of which contain SNPs (a, A → G in W2; b, A → G in 37.4; c, A → C in all isolates except 37.4; d, A → C in W2, 12.1, 37.4, and 51.1; e, A → G in HB3, 1776, 12.1, and 37.4). †Outside of repetitive sequences, three additional SNPs were identified (f, C → A in W2, 12.1, 37.4, and 51.1; g, C → G and A → G in D6). ‡Other polymorphisms included h, a deletion of one of three 31-nucleotide repeats in D6 and HB3; i, an AC deletion in W2, 1776, 37.4, and 51.1; and j, a single (A) nucleotide insertion in D6. §GenBank accession numbers are listed as a reference for each gene sequence.

tion 1 (14)]. The smaller estimate is based on the number of synonymous substitutions in an estimated 55 million years (My) since the divergence between *P. falciparum* and *P. berghei* (a rodent parasite), and the larger estimate is based on the number of synonymous substitutions in an estimated 7 My since the divergence between *P. falciparum* and *P. reichenowi* (26) (a chimpanzee parasite) (14). If the two suspect SNPs are also included, the estimated age of the MRCA, on the basis of the same assumptions, is in the range of 9500 to 23,000 years. These results are consistent with the hypothesis that all extant *P. falciparum* derived from a recent common ancestor. Our data are also consistent with the finding that mitochondrial DNA is virtually identical in nucleotide sequence among diverse isolates of *P. falciparum* (27) and with historical biogeography (28–30). Our estimates coincide with the establishment of slash-and-burn agriculture in the African rainforest less than 6000 years ago, which could have provided suitable expansion conditions for the mosquito vectors of *P. falciparum* (31) and adequate human population size to maintain transmission.

Although a single recent common ancestor is the most parsimonious interpretation of all the data, alternative explanations have been proposed [see (2, 3) for discussion].

These include a series of severe population bottlenecks, which seems to be inconsistent with the high level of microsatellite variation; or a series of selective sweeps, which would predict that regions of the genome left unaffected may contain a high frequency of SNPs that were present in the ancestral population. To date, there has been insufficient sequence analysis to reveal regions with excess SNPs. Some observations are inconsistent with a single common ancestor. For example, MSP1 alleles are widely divergent from one another, suggesting an ancient origin of *falciparum* malaria (32). Our data do not directly address MSP1; however, others (33) have suggested that a mechanism of replication slippage followed by immune selection could result in the current allelic variants. Our data are not only consistent with this model, but also provide evidence for an increased frequency of SNPs in regions of polymorphic microsatellites. An alternative explanation is that multiple tandem paralogs of MSP1 were present in progenitors and that biological and immune selection resulted in the selection of different MSP1 genes. There is evidence for expansion and contraction of the *var* gene family, another biologically and immunologically selected gene (11–13).

The relative lack of SNP polymorphisms in *P. falciparum* is seemingly inconsistent

with the ability of the parasite to evade chemotherapeutic and immunological interventions by the rapid evolution of drug resistance and antigenic variation. Inasmuch as most of the genes examined so far in *P. falciparum* have been studied because they are associated with drug resistance or antigenic determinants, it is not surprising that most amino acid replacements so far reported are associated with adaptation to drug therapy or the immunological status of the host (34). Our data go well beyond these classes of genes in implying that *P. falciparum* may be unique among pathogens in having essentially no nucleotide polymorphisms apart from those that can be explained by recent mutation and positive selection. Identification of SNPs throughout the genome of *P. falciparum* may, therefore, reveal whether changes in regulatory or protein-coding sequences have been predominant in the evolution of human pathogenesis, drug resistance, and immune evasion.

References and Notes

1. WHO Tech. Rep. Ser. 892, 1 (2000).
2. A. L. Hughes, F. Verra, *Genetics* 150, 511 (1998).
3. S. M. Rich, F. J. Ayala, *Genetics* 150, 515 (1998).
4. D. J. Kemp, A. F. Cowman, D. Walliker, *Adv. Parasitol.* 29, 75 (1990).
5. S. Biswas, A. Escalante, S. Chaiyaroj, P. Angkasekwinai, A. A. Lal, *Trop. Med. Int. Health* 5, 737 (2000).
6. S. J. Foote et al., *Nature* 345, 255 (1990).

7. X. Su, L. A. Kirkman, H. Fujioka, T. E. Wellems, *Cell* **91**, 593 (1997).
8. D. A. Fidock et al., *Mol. Cell* **6**, 861 (2000).
9. B. A. Biggs et al., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9171 (1991).
10. D. J. Roberts et al., *Nature* **357**, 689 (1992).
11. J. D. Smith et al., *Cell* **82**, 101 (1995).
12. X. Z. Su et al., *Cell* **82**, 89 (1995).
13. D. I. Baruch et al., *Cell* **82**, 77 (1995).
14. S. M. Rich, M. C. Licht, R. R. Hudson, F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4425 (1998).
15. X. Su, T. E. Wellems, *Genomics* **33**, 430 (1996).
16. T. J. Anderson, X. Z. Su, M. Bockarie, M. Lagog, K. P. Day, *Parasitology* **119**, 113 (1999).
17. W.-H. Li, *Molecular Evolution* (Sinauer, Sunderland, MA, 1997).
18. T. Lehmann, C. R. Blackston, S. F. Parmley, J. S. Remington, J. P. Dubey, *J. Parasitol.* **86**, 960 (2000).
19. M. J. Gardner et al., *Science* **282**, 1126 (1998).
20. S. Bowman et al., *Nature* **400**, 532 (1999).
21. Supplementary Web material (Methods and Figures) is available on Science Online at www.sciencemag.org/cgi/content/full/293/5529/482/DC1.
22. X. Su et al., *Science* **286**, 1351 (1999).
23. T. A. Kunkel, *J. Biol. Chem.* **261**, 13581 (1986).
24. S. M. Rich, R. R. Hudson, F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13040 (1997).
25. S. L. Salzberg, M. Perlea, A. L. Delcher, M. J. Gardner, H. Tettelin, *Genomics* **59**, 24 (1999).
26. A. A. Escalante, E. Barrio, F. J. Ayala, *Mol. Biol. Evol.* **12**, 616 (1995).
27. D. J. Conway et al., *Mol. Biochem. Parasitol.* **111**, 163 (2000).
28. F. B. Livingstone, *Am. Anthropol.* **60**, 533 (1958).
29. S. Jarcho, *Bull. Hist. Med.* **37**, 1 (1964).
30. L. J. Bruce-Chwatt, *Bull. WHO* **32**, 363 (1965).
31. M. Coluzzi, *Parassitologia (Rome)* **41**, 277 (1999).
32. A. L. Hughes, *Mol. Biol. Evol.* **9**, 381 (1992).
33. S. M. Rich, F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6994 (2000).
34. T. F. McCutchan, A. P. Waters, *Immunol. Lett.* **25**, 23 (1990).
35. We thank J. Hume and G. Ramirez for maintaining the parasite cultures. Supported by funding from NIH, the Burroughs-Wellcome Fund, and the Wellcome Trust. K.M.N. was supported by the Research Council of Norway.

14 February 2001; accepted 23 May 2001

Methane-Consuming Archaea Revealed by Directly Coupled Isotopic and Phylogenetic Analysis

Victoria J. Orphan,^{1*} Christopher H. House,^{2*†} Kai-Uwe Hinrichs,³ Kevin D. McKeegan,⁴ Edward F. DeLong^{1†}

Microorganisms living in anoxic marine sediments consume more than 80% of the methane produced in the world's oceans. In addition to single-species aggregates, consortia of metabolically interdependent bacteria and archaea are found in methane-rich sediments. A combination of fluorescence in situ hybridization and secondary ion mass spectrometry shows that cells belonging to one specific archaeal group associated with the *Methanosarcinales* were all highly depleted in ¹³C (to values of −96‰). This depletion indicates assimilation of isotopically light methane into specific archaeal cells. Additional microbial species apparently use other carbon sources, as indicated by significantly higher ¹³C/¹²C ratios in their cell carbon. Our results demonstrate the feasibility of simultaneous determination of the identity and the metabolic activity of naturally occurring microorganisms.

Microbes critically impact global geochemical cycles. Although the general ecological importance of microbial activity is well recognized, the identity and involvement of microbes in specific biogeochemical cycles are often poorly understood. For example, the anaerobic oxidation of methane (AOM) is a widespread and geochemically well documented process [e.g., (1)], yet very little is known about the physiology, biochemistry, and identity of the microbes involved. One reason for this is that often the ecologically

relevant microorganisms are difficult to isolate in pure culture. Approaches that combine phylogenetic and stable isotope analyses have considerable potential for linking microbial diversity with in situ activity (2–4). Recent

studies that combine phylogenetic surveys of ribosomal RNAs (rRNAs) with structural and stable isotopic analyses of lipids have revealed new information about methane-oxidizing microbes in anoxic marine sediments (3, 5, 6). However, the stable isotopic and phylogenetic methods used in these studies were uncoupled, so identification of the specific microbes mediating AOM is based mainly on indirect lines of evidence.

Here, we report a cultivation-independent study of marine microbial assemblages in anoxic methane-rich sediments that combined microbial cell identification using ribosomal RNA-targeted fluorescent in situ hybridization (FISH) (7) with secondary ion mass spectrometry (SIMS) (8). After rRNA-targeted probes were applied to identify microbial cells, the stable isotope composition of the identified cells was determined by using SIMS. Coupled FISH-SIMS provided a measure of the stable carbon isotope composition of individual phylogenetically identified cell aggregates. Uncultured, naturally occurring microbial cells that utilize methane as the source of cell carbon could therefore be identified unambiguously.

Previous studies suggested that cell aggregates of archaea belonging to the *Methanosarcinales* (ANME-2 group), surrounded by sulfate-reducing bacteria related to the *De-*

Table 1. Carbon isotopic compositions (versus PDB) and source assignments of selected extracted archaeal ether lipids and bacterial fatty acids (FAs). SRB, sulfate-reducing bacteria.

Compound	Eel River Basin, 3 to 5 cm		Santa Barbara hydrocarbon seep	
	δ ¹³ C (‰)	Source assignment	δ ¹³ C (‰)	Source assignment
Archaeol	−104.1	ANME-2	Traces/not analyzed	Archaea
sn-2-Hydroxyarchaeol	−107.6	ANME-2	Traces/not analyzed	Archaea
n-C _{14:0} FA	−69.1	SRB	−22.7	Algae, bacteria
i-C _{15:0} FA	−51.3	SRB, other bacteria	−25.4	Bacteria mainly
ai-C _{15:0} FA	−51.9	SRB, other bacteria	−19.0	Bacteria mainly
n-C _{16:1(ω7)} FA*	−62.8	SRB	−26.0	Algae, bacteria
n-C _{16:1(ω5)} FA*	−76.1	SRB	−21.2	Algae, bacteria
n-C _{16:0} FA	−44.1	SRB, other bacteria, algae	−20.3	Algae, bacteria
10-Me-C _{16:0} FA	−69.4	SRB	Not present	—

¹Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA. ²Department of Geosciences, The Pennsylvania State University, University Park, PA 16802, USA. ³Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. ⁴Department of Earth and Space Sciences, University of California, Los Angeles, CA 90095, USA.

*These authors contributed equally to the work.

†To whom correspondence should be addressed. E-mail: chouse@geosc.psu.edu and delong@rmbi.org

*Double-bond positions were determined from nearby samples with almost congruent fatty acid distributions (24) and refer only to the sample from the Eel River Basin.