Genetic diversity and evolutionary history of *Plasmodium falciparum* and *P. vivax*

Aparup Das[1,*], Ruchi Bajaj[1], Sujata Mohanty[2] and Vijaylakshmi Swain[3]

[1]Evolutionary Genomics and Bioinformatics Laboratory, National Institute of Malaria Research (ICMR), 22 Sham Nath Marg, Delhi 110 054, India
[2]Department of Biotechnology, Jaypee Institute of Information Technology University, A-10, Sector 62, Noida 201 301, India
[3]Zoology Department, University of Delhi (North Campus), Delhi 110 007, India

Understanding the genetic diversity and evolutionary history of the malaria parasite and how it is genetically distinct in different regions of the genome and populations, may open up new avenues to population-specific malaria control measures. The two principal human malaria parasites, *Plasmodium falciparum* and *P. vivax*, seem to be very different in origin and in phylogenetic resemblance to other species of *Plasmodium*. Further, the mortality and morbidity that these two parasites express are quite different, so also the percentage of nucleotide composition in their respective genomes. However, the net genetic diversity at the molecular level (in different independent genes and in both nuclear and mitochondrial genomes) seems to have very little differences, with *P. vivax* at a slightly higher scale. Whereas African populations of *P. falciparum* are highly diverse at the DNA level among other populations, high genetic diversity was found in Asian *P. vivax*; clearly depicting that *P. falciparum* has possibly originated in Africa and *P. vivax* in Asia. Furthermore, the findings of a comparably higher genomic diversity in *P. vivax* corroborate to earlier hypothesis of being older than *P. falciparum* and also a close genetic affinity with the malaria parasite species of Asian monkeys. On the other hand, *P. falciparum* is genetically much similar to the chimpanzee malaria parasite, *P. reichenowi*. The advancements in genomics and parallel statistical analyses of the DNA sequence data so far have succeeded in revealing new evolutionary information on the parasite genes and populations and generate renewed hopes for effective and common vaccines for both forms of malaria.

Keywords: Bioinformatics, evolution, genetic diversity, genomics, *Plasmodium falciparum*, *Plasmodium vivax*.

MALARIA remains an urgent problem in global public health with the annual death toll of 0.7–2.7 million, with more than 75% of the victims being African children[1]. Over the past 35 years, the incidence of malaria has increased 2–3-fold. At present, it affects 300–500 million people and causes about one million deaths, primarily in Africa. In 1955, the World Health Organization (WHO) began an ambitious programme to eradicate malaria through clinical treatment using chloroquine and by control of the mosquito population using DDT (dichlorodiphenyltrichloroethane). Phased out in the late 1960s, the programme nevertheless resulted in an important and sustained reduction in the burden of the disease in many countries throughout the world[2]. However, in many countries there has been resurgence in malaria. This has resulted essentially from the emergence and spread of drug-resistant parasites[3,4], the evolution of insecticide-resistant mosquitoes[5,6], increased population density (the world population has doubled since 1963), global warming (which has allowed the spread of vectors into areas that were previously outside their range), continuing poverty, political instability and loss of productivity due to infectious diseases[7]. All these factors undermine the maintenance of a stable public-health infrastructure for the treatment and control of malaria[8].

Human malaria is a parasitic disease that is endemic in most tropical and subtropical ecosystems worldwide[9]. Malarial parasites belong to the genus *Plasmodium* and infect many vertebrate hosts, including several species of non-human primate[10]. Four *Plasmodium* species are parasitic to humans: *Plasmodium falciparum*, *P. malariae*, *P. ovale* and *P. vivax*. Of these, *P. falciparum* and *P. vivax* are associated with most malaria morbidity and mortality, respectively. Both the *P. falciparum* and *P. vivax* genome, like all *Plasmodium* genomes studied so far, appear to be distributed among 14 haploid chromosomes[11], but the *P. vivax* genome contains a much lower AT (~55%) content[12] in comparison to the genome of *P. falciparum* (~80%)[13]. Conservation of gene synteny is extensive among all of the *Plasmodium* species and declines as the evolutionary distance between the species increases[11,14].

Knowledge on the genetic diversity in genome is important in biomedical research in many respects. However, detecting and understanding its maintenance in populations is somewhat tricky as different evolutionary forces help in maintaining the genetic diversity in an organism and a population genetic approach is necessary for unleashing the net genetic diversity present in the genome of an organism in a population[15]. It is now quite clear that both *P. falciparum* and *P. vivax* use the genetic diversity they possess in the genome to fight against the anti-malarial drugs and host immunity and the mecha-

*For correspondence. (e-mail: aparup@mrcindia.org)

nism of maintenance of such diversities is unknown. Thus, estimating genetic variations and inferring evolutionary history of malaria parasite genes and populations is critical in understanding several mechanisms in malaria, especially in identifying drug and vaccine targets, understanding virulence pattern, host–parasite interactions and evolution of drug resistance.

Malaria parasite genetic diversity and evolution: The basics

The fact that diversity in genetic content of individual organisms bears the signature of evolutionary history of species and genes has long been documented ever since the rediscovery of the Mendelian genetics combined with Darwin's theory of evolution by natural selection. So far two basic approaches have been followed in population genetic studies of malaria parasites: (i) to understand the parasite population genetic structure and intrahost dynamics, including the stability of strains and complexity of infections, and (ii) studies assessing the diversity of specific genes encoding vaccine antigens and those associated with drug resistance. The studies on population structure quantify how alleles are spatially dispersed and how the overall genetic diversity is organized. These studies provide important indications as to the origin, dispersal and stability of multilocus genotypes. Analysis of defined genes provides information about how different alleles are generated and maintained in the population. Specifically, it addresses the relevance of factors such as intragenic recombination and natural selection on the observed polymorphisms. These approaches answer fundamental questions related to the association of specific alleles with drug resistance and how the genetic variation at a given locus affects vaccine development and deployment[16].

Evolutionary history of genes and populations of *P. falciparum*

The recent upsurge of malaria owing to a coincidence of mutually reinforcing factors might echo the first expansion of malaria, which is thought to have taken place about 10,000 years ago[17,18]. The events that could have conjoined to create this earlier expansion include: (i) climate change in Africa after the last glaciation leading to optimal warm and humid conditions in the equatorial regions about 10,000 years ago[19], (ii) increase in human population density owing to the spread of swidden ('slash-and-burn') agriculture into Africa at about the same time[20], and (iii) proliferation and rapid diversification of the highly anthropophilic *Anopheles* mosquito vectors[21,22] which efficiently transmit the malaria parasites. Determining whether a major expansion of malaria took place about 10,000 years ago is of interest because understanding the causes of this expansion might provide clues for controlling malaria today[8]. Interestingly, evidences that have a more direct bearing on the origin of epidemic malaria have come from the level of genetic variation in the parasite itself. However, the evidence is not altogether consistent across different studies involving different genetic loci. Some studies reveal a low level of genetic variation in *P. falciparum,* indicating a relatively recent expansion, whereas other studies reveal a high level of genetic variation, indicating a large population size that has been maintained over hundreds of thousands of years[23].

Inferences from independent nuclear genes

Antigen-coding genes and microsatellites: The possibility of a rapid expansion of *P. falciparum* about 10,000 years ago received renewed interest after an analysis of protein-coding sequences[24]. Rapid population expansion can leave a characteristic signature on genetic polymorphisms in the form of a reduced overall level of variation, as well as a frequency distribution of variants that is skewed towards rare alleles. Neither of these signatures is observed in *P. falciparum* for genes that encode antigens, or for microsatellite polymorphisms. However, neither the antigen-coding genes nor the microsatellites is considered as the appropriate markers as genes that encode antigens are subject to strong selection for diversity to evade the host immune response, and microsatellites have an extraordinarily high mutation rate, owing to replication slippage[25,26].

Although a dramatic population expansion 10,000 years ago does not require that all extant *P. falciparum* descended from a single common ancestor, evidence for a single common ancestor at around this time provides strong support for this model of population history of *P. falciparum*. However, the inference seems at odds with the high level of polymorphism that is found in many genes in *P. falciparum*, especially genes that encode antigens or membrane-bound proteins[27]. These antigens include MSP1 (merozoite surface protein 1), one of the most highly polymorphic proteins in eukaryotes. However, there is no doubt that selection for diversity in MSP1 is strong, owing to its interaction with the host immune system.

Single nucleotide polymorphisms in the second and third chromosome: While the data from coding sequences were being debated, the complete sequence of *P. falciparum* chromosomes 2 and 3 became available through the Malaria Genome Project[28,29]. This important information afforded a new approach to the study of genetic diversity in *P. falciparum* as it allowed non protein-coding (non-coding) sequences to be examined. The rationale is that if there are unrecognized constraints on synonymous sites, then non-coding sequences will be a more reliable indica-

tor of the level of background polymorphism because of the relative lack of selective constraints. This led to the examination of introns[30], which are among the most rapidly evolving DNA sequences. The initial study analysed 4217 bp in 25 introns across chromosomes 2 and 3 in eight reference isolates. From the analysis, excluding the five single nucleotide polymorphisms (SNPs) located in microsatellites, the estimated time to the Most Recent Common Ancestor (MRCA) of the reference strains was calculated to be 9500–23,000 years. Additional uncertainty about the timing of the MRCA came from a study of 204 protein-coding and noncoding sequences on chromosome 3 from five diverse isolates of *P. falciparum*[31]. On the whole, these data show less polymorphism than would be expected from an organism that had maintained a very large population size for several hundred thousand years. For example, across the central region of the third chromosome, 150 of 204 genes contained no synonymous SNPs. However, the overall level of SNPs was sufficient to imply a most recent common ancestor for chromosome 3 that existed 100,000–180,000 years ago. This is much older than the previous estimates from coding sequences[32], and even older than the estimates obtained from introns[30].

Studies involving mitochondrial genome: For evolutionary studies, one problem that lies with the nuclear genes is that they can undergo recombination. Although inbreeding (mating between related genotypes) takes place in many local populations of *P. falciparum* and this reduces the effective rate of recombination, the level of inbreeding is usually low enough that recombination is effective in breaking up linked blocks of nucleotide sequences. This means that regions from different chromosomes can have distinct ancestral histories, so the MRCA can differ from one chromosomal region to the next. To this extent, the mitochondrial genome is inherited through only one parent, thus it does not undergo recombination and is therefore potentially more informative for inferring population evolutionary history than are nuclear genes. Evidence for a recent mitochondrial common ancestor emerged from an initial study of the complete sequence of the mitochondrial genome (6 kb pairs) from four isolates, which revealed an extremely low level of synonymous nucleotide polymorphism[33]. Additional DNA typing of the polymorphisms among 104 diverse isolates showed a geographical pattern that indicated migration out of Africa into Southeast (SE) Asia and South America. The inference of a rapid and recent expansion of *P. falciparum* has also been supported by sequences of the complete mitochondrial genome from 100 geographically diverse isolates[34]. The mitochondrial data for African isolates show a mismatch distribution that is typical of rapid population growth, with a model that is consistent with rapid expansion having occurred approximately 10,000 years ago[34] (Figure 1). The data also indicate that some lineages are more ancient (50,000–100,000 years) and that the parasite populations migrated to SE Asia and South America prior to the African expansion (Figure 2). The latter inference is surprising, as the pattern of microsatellite variation in SE Asia and South America indicates a much more recent invasion[32], and is also at odds with the conventional belief that malaria was introduced into South America by Europeans and the slave trade[35].

However, for any particular non-recombining DNA sequence, such as mitochondrial DNA, the pattern of sequence relationships has a large variation, which is purely due to the chance extinction of lineages that occurs over a period of time. Therefore, the mitochondrial genealogy might not necessarily coincide with that of any nuclear gene. Moreover, this analysis assumes that all nucleotide substitutions in mitochondrial DNA are selectively neutral. If some are selectively favoured, then the increased frequency of this type of substitution would make them seem substantially older than they actually

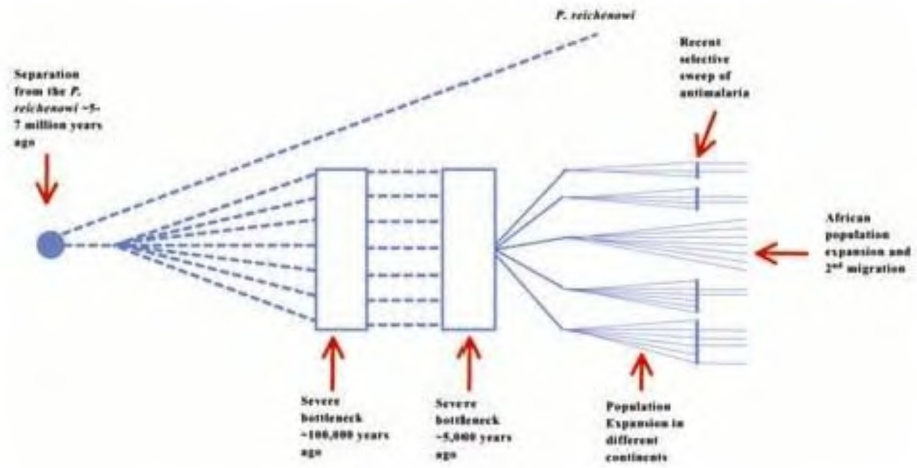

Figure 1. Speciation and putative population demographic history of *Plasmodium falciparum* based on genetic studies.

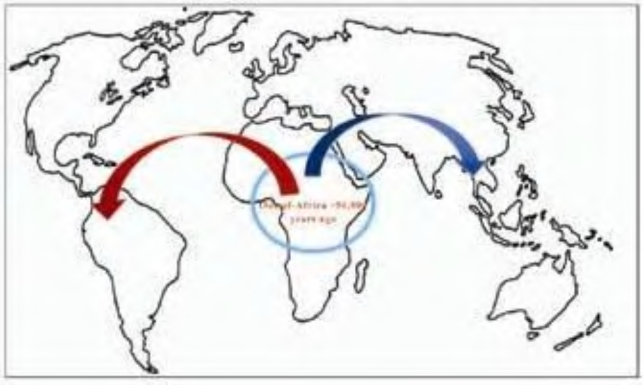

Figure 2. Proposed origin and migration out-of-Africa of *Plasmodium falciparum* based on genetic studies.

are. The possibility of natural selection, however, cannot be entirely ruled out, since the mitochondrial gene encoding cytochrome *b* seems to underlie susceptibility to certain antibiotics[36].

Inferences from whole genome re-sequencing studies

Although the *P. falciparum* whole genome has been sequenced and published five years ago from a cultured isolate (3D7), this does not tell on the amount of genetic diversity it possesses and thus necessitates a population genomic study, with representing isolates from the whole distribution area. Almost 60% of genes in *P. falciparum* have unknown functions which poses a great challenge in identifying the vaccine targets in the genome. Thus the major aim to sequence the *P. falciparum* genome is to discover the vaccine and drug targets[13]. Since the distribution of genetic diversity in the genes across the genome of this species is controversial[23] and majority of the antigenic and drug resistance genes are shown to be highly polymorphic and under various selection pressure[4,27,33,37,38], a genome-wide distribution of the genes with signature of natural selection is one of the best ways to discover the genes which are possible target to drugs and vaccines.

Genome-wide SNPs and insertion-deletion polymorphisms: Recently, Volkman and co-workers[39] have reported the extent of genetic diversity in *P. falciparum* genome. High quality draft genome from two clones of different geographic locations, Honduras and Indochina, were obtained and compared with the published genome of 3D7 isolate[13]. The comparison uncovered 26,845 SNPs, making the genomic average of one SNP in every 780 nucleotide bases, indicating high genetic diversity in *P. falciparum* genome. This is in disagreement with a previous estimate of one SNP per 910 bases in the third chromosome[40]. In addition to the SNPs, the authors[39] detected as many as 37,039 insertion-deletion (indel) polymorphisms of at least three nucleotide bases, providing evidence that indel polymorphisms are as much abundant in *P. falciparum* genome as the SNPs and were equally highly frequent. Twelve additional lines of *P. falciparum* from different world-wide locations were also studied with light genomic coverage and 12,530 additional SNPs were detected.

Whole genome of two *P. falciparum* isolates (one field isolate and one laboratory clone) and *P. reichenowi* was completely sequenced in yet another recent study[41] and compared with the published genome sequences of the 3D7 isolate[13]. Comparing all the three genome sequences of *P. falciparum* uncovered 27,169 non-redundant SNPs and as many as 216,619 fixed differences between the 3D7 isolate and *P. reichenowi*. There were also large numbers of indel events. The results are clearly indicative of a high genetic variation in *P. falciparum* genomes and the divergence between *P. falciparum* and *P. reichenowi* is about ten-fold greater than the intra-specific polymorphisms. The findings revealed an excess of insertion mutations in 3D7 which have been accumulated during the parasite culturing process. The indel events were however underrepresented in the coding regions.

In order to detect genome-wide variations and identify possible vaccine targets, Mu and co-workers[40] recently followed an evolutionary approach to survey about 65% of the predicted genes in *P. falciparum* genome (3539 genes) for polymorphisms in four cloned isolates (Dd2, Hb3, D10 and 7G8) and compared with the homologous regions in the 3D7 published whole genome[13]. They have detected as many as 3918 SNPs. About 54% of the total genes surveyed had one or more SNPs and the majority of them were non-synonymous. Further re-sequencing of about 45 kb of DNA fragment covering 183 already-reported known SNPs in chromosome 3 in 99 worldwide isolates of *P. falciparum* detected 108 SNPs (out of the 183 reported SNPs) and discovered 185 completely new SNPs, of which 29 are common SNPs (minor allele frequency ≥ 0.05)[42]. In addition, a high abundance of microsatellites in the *P. falciparum* genome was detected, averaging one polymorphic microsatellite per 1.3 genes[42].

Rate and distribution of genomic diversity: Taking into account the whole data set, the genomic average of nucleotide diversity becomes equal to 1.16×10^{-3}. Also, genetic diversity varies widely across different chromosomes; from 8.2×10^{-4} (chromosome 4) to 2.41×10^{-3} (chromosome 14). Though the level of diversity is not uniformly distributed across the chromosomes, there seem to be some diversity hotspots as well as coldspots. While the majority of the diversity hotspot regions seem to be dominated by the highly diverse surface molecule responsible for cytoadherence and antigen-coding genes (e.g. *var*, *stevor*, *rifin*), the coldspots are the places where low-diverse genes responsible for electron transport chain and mitochondrial function (e.g. genes encoding for essential metabolic enzymes, such as, oxidases) were present[27].

In order to infer if recent positive selection has left any footprint on the gene responsible for conferring drug resistance in *P. falciparum*, the data were further dissected. Data from four chloroquine-resistant isolates were compared with six sensitive varieties and about 60–100 kb of DNA fragments in chromosomes 5, 7 and 11 with fairly low diversity were found[27]. Interestingly, the low diversity region in chromosome 7 included the *pfcrt* gene (conferring chloroquine resistance) that is suggested to be under selective sweep model of natural selection. When similar experiments were conducted for pyrimethamine, two regions in chromosome 13 and 14 were found to have significantly low diversity. The well-known gene *dhfr* had reduced diversity, but not significant to be considered under selective sweep. Another hitherto unidentified region spanning around 160 kb clearly reflected the signature of positive selection, as low genetic diversity was estimated.

SNPs were not found to be distributed randomly across the chromosomes; some regions have consecutive genes with SNPs and some regions do not bear SNPs at all. The percentage of genes with SNPs varies across *P. falciparum* genome, from 47.1% (chromosome 13) to 67.5% (chromosome 7). There was distribution heterogeneity in the number of SNPs per gene between chromosomes that differ more than two-fold, averaging 0.85 to 2.08 SNPs per sequenced gene with large chromosomes having fewer SNPs. Jeffares and co-workers[41] detected a negative correlation between the chromosome size and the number of SNPs per gene in all chromosomes except chromosome 7. Further, nucleotide diversity was statistically significantly higher in the genes located at the end of the chromosomes than the rest of the genes. In contrast to earlier hypotheses, a relatively high frequency of recombination with several hot and cold spots in the genome were found, with most hot spots being clustered in sub-telomeric regions, conforming to the previous findings in the third chromosome[40].

Genomic diversity, gene expression and evolution of P. falciparum: In order to estimate the selective effects acting on individual genes, the ratio of the non-synonymous (nucleotide substitution with change in amino acid) substitution rate to the synonymous (nucleotide substitution without the change in amino acid) for the three pairwise comparison of each gene with at least 100 codons was calculated and it was found that most genes are under purifying selection and some under positive selection[41]. Further, correlating evolutionary rate of protein-coding genes with mRNA and protein expression level, as described in other organisms, the authors found that the rapidly evolving genes are expressed at a low level and evolutionarily conserved genes are highly expressed. This study and the others indicate that highly specialized genes in *P. falciparum* genome are most likely to undergo accelerated selective process either because of relaxed constraint or directional selection. Significantly elevated rates of non-synonymous changes in two types of gene clusters related to the merozoite stage of parasite (merozoite invasion and early ring transcript) were observed. Between the 3D7 and *P. reichenowi*, comparisons followed by other statistical tests indicated that these proteins are highly variable within *P. falciparum* isolates.

In order to understand the evolutionary history of the genes involved in direct interaction with host cells, the evolutionary rates of genes by their intracellular localization was examined and it was found that the proteins localized to the nucleus, cytoplasm and the mitochondrion are generally conserved, whereas apicoplast-localized proteins, predicted membrane-spanning proteins and predicted export proteins have evolved significantly rapidly[41]. These findings are also corroborated by previous studies indicating strong positive selection in the genes related to immunity and defense between humans and chimpanzee[43]. All these studies are consistent with the model of 'evolutionary arm race' between the mammalian immune system and the exposed proteins of *Plasmodium*. Corroborating this model, the genes annotated as 'antigens' in *P. reichenowi* and 'cell communication' category of *P. falciparum* showed significant excess of non-synonymous polymorphisms, providing evidence for rapid evolution[41]. However, it is not evident at first hand that in which species positive selection has shaped the observed diversity in these genes. It seems from the biology of these two parasite species that positive selection has more probably driven these genes in *P. falciparum*.

Identification of vaccine candidates: While organizing the pattern of polymorphism according to the function of genes, Mu and co-workers[42] detected the genes encoding surface antigens and cell adhesion molecules and proteins involved in drug interactions are mostly polymorphic. This indicates, as described before, that the antigenic groups of genes are possibly under balancing, diversifying or partial directional selection. One of the interesting findings is that, chromosomal regions flanking the *var* genes (the *var* genes encode a family of variant antigens called PfEMP1 that are important for immune invasion and disease pathogenesis) are mostly polymorphic than the genome-wide average, indicating that these genes are under strong balancing selection pressure from the host immune response[44], thus maintaining more genetic variation than expected under neutrality. Further, approximately 40% of the 83 loci was found to be with five or more consecutive polymorphic genes that contain genes encoding known antigens[42]. Also, 56 genes were detected to have high nucleotide diversity with one or more SNPs. More than half of the genes encode proteins of unknown functions, only 10 encode known antigens. Majority of the genes encode proteins with signal peptides and transmembrane domains which signify that these potential membrane and/or surface localization may be recognized by the host immune system.

Further, 108 (56 plus 52 genes that have five or more SNPs) genes encoding a predicted signal peptide and/or

transmembrane domain were expressed using an *Escherichia coli* cell-free rapid expression system[42]. Sixty-five proteins were successfully expressed and 11 were recognized by pooled human sera, seven of these were previously unknown antigens which could be potential vaccine targets. The study also showed that different *var* gene clusters are under different immune selective pressure, which is important to consider while designing *var*-based vaccines.

The genomic diversity studies[39,41,42] thus opened up the field on evolutionary approach to drug and vaccine target identification in the genome and demand further evolutionary studies in world-wide populations to understand both population history of the parasites and carry out genotype–phenotype association studies which could lead to new strategies to control this dreadful disease. In this respect, one clear gap in the sampling of *P. falciparum* from malaria-endemic regions of the world is the Indian subcontinent, for which no whole-genome sequence data yet exist[45]. Further, additional coverage of *P. reichenowi* would also do much to answer questions concerning the quality of the existing sequence data, highlighting the importance of quality as well as quantity in these studies[45]. While comparable whole-genome sequencing studies of large numbers of individual parasite isolates from the field would be one of the next desirable steps for the malaria community, the foundations laid in these three studies[39,41,42] reinforce the importance of a combination of modern biological approaches to biomedical research.

Evolutionary history of *P. vivax*

Until the middle of the 1900s, *P. vivax* was the most globally widespread and arguably the most prevalent of the four malaria parasite species that infect humans. The ability of this parasite to complete its sporogonic cycle at a minimum lower temperature of 16°C, compared with 21°C for *P. falciparum*, has substantially contributed to its success in establishing stable foci of transmission in the temperate zones. Although responsible for less mortality than *P. falciparum, P. vivax* causes considerable morbidity. *P. vivax* confounds control measures owing to the presence of dormant stages in the liver, which lead to relapses of the infection, weeks after the initial episode[46]. Alarmingly, an increasing number of clinical studies has shown treatment failure of the first-line *P. vivax* antimalarial, chloroquine[47]. *P. vivax* seems to have extraordinary phenotypic diversity, especially in its relapse patterns, and is found in a broad range of ecotypes, from Russia to the tropical regions of Asia, the Pacific, and South and Central America[46]. The relatively stable GC-rich genome of *P. vivax* (which has a GC content double that of *P. falciparum*) has few di-nucleotide microsatellite markers, and the (TA) and (CA) motifs commonly seen in *P. falciparum* and other organisms such as yeasts, are rare in *P. vivax*[12].

Genetic diversity, origin and population history of P. vivax

In general, very little is known of the global genetic diversity and population structure of *P. vivax*. The major obstacle to this is the lack of appropriate genetic markers for the *P. vivax* genome, which has severely hampered an in-depth analysis of the population structure and evolutionary history of the parasite and prevented efforts to map determinants contributing to important parasite phenotypes such as drug resistance and relapse patterns[48]. However, recent studies on the mitochondrial genome and the microsatellite loci provide a clear picture on the evolutionary history of *P. vivax*. Thus, phylogenetic studies[49,50] have shown that this parasite originated from a malarial parasite of non-human primates as a result of a host switch. Further, some SNPs have also been reported across the genome.

Inference from the mitochondria genome. Two recent independent studies[51,52] involving the mitochondrial genome provide a clear picture on the evolutionary history of *P. vivax*. The first study[51] includes 176 sequences and the second study includes 106 sequences[52]. There was some overlap in the estimates of Time to MRCA (TMRCA); one study[51] concluded that *P. vivax* originated between 53,000 and 265,000 years ago while in other study[52] it was estimated between 217,000 and 304,000 years. The estimation of origin differs due to differential analytical methods and the use of geographically different samples. The data from these two studies were compiled, giving a total of 282 sequences with 93 haplotypes covering the known geographical distribution of *P. vivax*[53]. Analyses of this combined dataset revealed a haplotype network and the results were similar to those reported in one of these studies[51]. Overall, Asian haplotypes were found to be more diverse than those of other regions. Although the number of haplotypes *per se* does not provide evidence of the location of the ancient population or of the population that shares an older common ancestor, both Melanesian and Asian populations have similar numbers of haplotypes (Figure 3). Further, the available data regarding complete mitochondrial genomes have been analysed to study the phylogeny of SE Asian malarial parasites of primates and the history of *P. vivax* populations[53]. The phylogeny estimated from complete mitochondrial genomes confirms previous results[49,50], indicating that *P. vivax* originated from a malarial parasite of non-human primates that is related to the species currently found in Asian macaques. Further, the Asian populations of *P. vivax* consistently seem to be older than other populations[53]. These lines of evidence (phylogenetic and population genetics) are most parsimoniously explained by an Asian origin of *P. vivax* (Figure 3).

Inference from the microsatellite markers. The *P. vivax* genome is very diverse but has a relatively low abundance

of microsatellites. Although *P. vivax* microsatellites are less frequent, a study of a single microsatellite sequence demonstrated a high degree of allelic diversity[54]. The genome of *P. vivax* was also found to be abundant in polymorphic tandem repeats (TRs), with one approximately for every 3 kb[48]. However, a controversy has arisen from the use of microsatellite diversity data to reconstruct the evolutionary history of *P. vivax*. While in certain studies[55] it has been shown that the di-nucleotide repeats have a low level of polymorphism, suggesting a recent bottleneck event in the evolutionary history of *P. vivax*, others[56] reported a very high level of microsatellite diversity. Variations in the length of 11 microsatellites were compared for the diversity of *P. vivax* from Thailand, India and Columbia (82 isolates in total) using the draft of the unpublished *P. vivax* genome[56]. Interestingly, a high degree of polymorphism in all of the microsatellite loci examined was found[56]. The data showed that *P. vivax* had a high allelic diversity, between seven and 18 alleles per locus and a high proportion of individuals being heterozygous at a particular locus in each of the three countries studied (Thailand 0.77, India 0.76 and Columbia 0.68). Microsatellite variation seems to be dependent on the length of the repetitive sequence[55] meaning longer arrays are more diverse than shorter ones because slippage mutations become exponentially more common with an increase in array length. These findings[55,56] clarified the seeming disparity between the high degree of SNP diversity and meagre microsatellite diversity, and provide little support for a recent evolutionary origin for *P. vivax*[57].

Inferences from antigen-coding genes. Like *P. falciparum*, high levels of polymorphism have been described for many *P. vivax* genes, particularly antigenic genes that might be under selection by host immunity. Polymorphic markers encoding parasite surface antigens (e.g. genes of the merozoite surface proteins (MSP), PfMSP1 and PfMSP2), repeat region size polymorphism (genes for the glutamate-rich protein PfGLURP and the circumsporozoite protein (CSP) PfCSP or PCR/restriction fragment size polymorphism (e.g. PfMSP2 and PvMSP3a)) have been used as genetic markers to study parasite diversity. Currently, DNA markers for genotyping *P. vivax* include the genes encoding MSP1 and MSP3a, apical membrane antigen 1 (AMA1), CSP, GAM1 and dihydrofolate reductase (DHFR). To date, genes encoding seven MSPs have been cloned from *P. vivax*, all named in accordance with the *P. falciparum* orthologues. PvMSP1 and PvMSP3a are highly polymorphic and are currently used for genotyping *P. vivax* field isolates[58,59]. AMA1, a protein essential for erythrocyte invasion, is highly conserved in *Plasmodium*. Similar to PfAMA1, PvAMA1 shows limited sequence polymorphism and displays limited genetic diversity within a geographic region[60,61]. Similar to that of *P. falciparum*, the *P. vivax* CSP is one of the most extensively studied antigens and is now considered to be a major vaccine candidate. Analysis of PvCSP sequences revealed that parasites have repeats belonging to one of two types of repeat units (GDRAA/DGPQA or ANGAGNQPG). Both variants have a worldwide distribution. Analysis of the PvCSP sequences from SE Asian isolates indicates very limited polymorphisms outside the repeat region[16]. The *PvGAM1* gene, named because of its expression in gametocytes, displays polymorphic deletions near the 3' end in a limited number of Sri Lankan isolates[62] and its allelic dimorphism has recently been reported in Indian isolates[63]. Since the information available is derived from antigens, geographical differentiation caused by positive natural immune selection rather than reduced gene flow might be the cause of high genetic diversity in these genes.

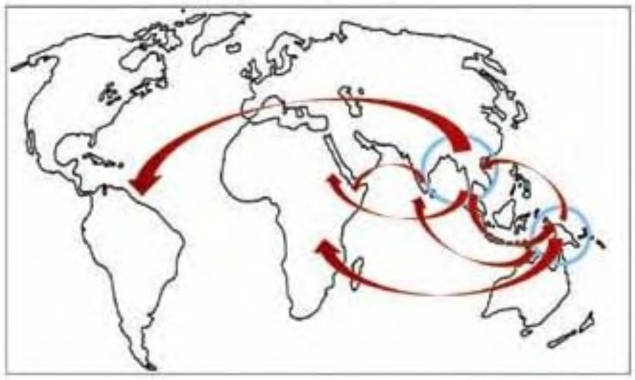

Figure 3. Proposed origin and migration routes of *Plasmodium vivax* as reflected from genetic studies.

Inference from genome-wide SNPs. A study involving the genome-wide SNPs has been made relatively recently representing a large-scale, multigene survey of variations within the *P. vivax* genome[48]. A ~100-kb contiguous chromosome segment from five isolates was analysed revealing 191 SNPs and 44 size polymorphisms. Whereas the majority (~63%) of the SNPs was in intergenic regions, introns contain significantly less SNPs. Furthermore, these SNPs were not evenly distributed across the segment; rather there were both high and low diversity hotspots. In fact, SNPs appear to be present at a higher frequency in the *P. vivax* genome than in *P. falciparum* genome. Twenty of the 26 genes in the DNA segment were polymorphic among the isolates studied, with two thirds of the coding SNPs identified as resulting in amino acid changes. In addition, size polymorphisms are abundant, with half of the 33 tandem repeats identified as occurring in coding regions[48]. One interesting observation is that very high frequencies of SNPs are found in noncoding regions flanked by well-conserved genes. The data from this study indicate that some introns in *P. vivax* may be under functional constraint. A total of five SNPs were found in 25 introns from 10 genes (total 4719 bp), averaging one SNP per ~1 kb. This study provides useful information for further understanding the genome diversity

of *P. vivax* and the overall data presenting a highly polymorphic genome in *P. vivax* might pose some challenges for future drug and vaccine development.

Even before the molecular biology era, it was recognized that *P. vivax* has an incredible range of phenotypes, both in terms of its morphology and relapse patterns. Considering the data from mitochondrial DNA and other nuclear genes including SNPs, *P. vivax* is thought to have diverged from its MRCA about 314,000 years ago[48]. The discovery of 191 SNPs and 44 size polymorphisms in a portion of the *P. vivax* genome fits a model proposing an ancient ancestor better than a model with a recent one[48]. The high degree of heterozygosity and allelic diversity would suggest that if faced with a major bottleneck event, *P. vivax* would be able to respond both immediately to the selection event and to other long-term selection pressures lasting many generations.

Conclusion

The increase in incidences of malaria has made it a priority for the international health community and is now the focus of several new control initiatives. The efficacy of malaria control measures is generally assessed from their impact on the number of infected individuals/mosquitoes, clinical cases and treatment failures, etc. However, it is possible to envisage situations where intervention only marginally affects such measurable outcomes over the period of observation. Parasite genetic diversity could then be exploited to detect subtle changes in the population. Changes in the overall genetic composition of the population or the multiplicity of infection observed in individuals can provide an indication of an effect resulting from an intervention such as the deployment of bednets or vaccination. Knowledge on the evolutionary history of the parasite populations as inferred from genomic outlook with rigorous statistical analyses with population genetic model testing would definitely contribute to these efforts. Though *P. falciparum* and *P. vivax* express differential disease patterns, molecular analysis of mitochondrial and nuclear genomes revealed almost a similar level of diversity in both the species. However, it seems that *P. vivax* possesses a slightly higher genome diversity than that of *P. falciparum*. The analyses of the genome diversity patterns supported the hypothesis that both the parasites underwent ancient population expansions and the age of the MRCA of the mitochondrial genomes is close to the previous estimates of the time of the human mitochondrial MRCA and the origin of modern *Homo sapiens*. This is consistent with the hypothesis that both these *Plasmodium* species were parasites of the hominid lineage before the origin of modern *H. sapiens* and that their population expansion coincided with the population expansion of their host.

Knowledge of the evolutionary history of malarial parasite populations is essential for detecting genetic adaptive variation such as that expected in potential antigens and for understanding patterns of linkage disequilibrium that could enable the identification of genes associated with, for example, drug resistance and vaccine candidates. An increasing number of *Plasmodium* antigens is now under development as vaccine candidates, many of which are polymorphic and the induced immune responses might be more effective against parasites carrying particular allelic variants. Early detection and monitoring of resistant parasites is needed to avoid the disastrous humanitarian problems that have resulted from the origin and spread of drug-resistant parasites. Direct monitoring requires knowledge of the genetic basis and evolutionary pattern of resistance[64]. Therefore, understanding the geographic population structure is also important in epidemiological surveillance studies because the source of parasites could be established more easily in cases of imported malaria[65]. Also, population genetic studies of genes related to drug resistance and vaccine candidates would help in formulating new drugs and vaccines taking care of the net genetic diversity and evolutionary trend of the parasite populations.

1. Breman, J. G., The ears of the hippopotamus: Manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.*, 2001, **64**, S1–S11.
2. Sachs, J. D., A new global effort to control malaria. *Science*, 2002, **298**, 122–124.
3. Hastings, I. M., Bray, P. G. and Ward, S. A., A requiem for chloroquine. *Science*, 2002, **298**, 74–75.
4. Wootton, J. C. *et al.*, Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature*, 2002, **418**, 320–323.
5. Hemingway, J. and Ranson, H., Insecticide resistance in insect vectors of human disease. *Annu. Rev. Entomol.*, 2000, **45**, 369–389.
6. Hemingway, J., Field, L. and Vontas, J., An overview of insecticide resistance. *Science*, 2002, **298**, 96–97.
7. Gallup, J. L. and Sachs, J. D., The economic burden of malaria. *Am. J. Trop. Med. Hyg.*, 2001, **64**, S85–S96.
8. Hartl, D., The origin of malaria: Mixed messages from genetic diversity. *Nature Rev. Microbiol.*, 2004, **2**, 15–22.
9. Snow, R. W. *et al.*, The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, 2005, **434**, 214–217.
10. Garnham, P. C. C., *Malaria Parasites and Other Haemosporidia*, Blackwell Scientific Publications, Oxford, 1966.
11. Carlton, J. M. *et al.*, Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.*, 1999, **101**, 23–32.
12. Carlton, J. M., Profiling the malaria genome: a gene survey of three species of malaria parasite with comparison to other apicomplexan species. *Mol. Biochem. Parasitol.*, 2001, **118**, 201–210.
13. Gardner, M. J. *et al.*, Genome sequences of the human malaria parasite *Plasmodium falciparum*. *Nature*, 2002, **419**, 531–534.
14. Carlton, J. M. *et al.*, Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelli yoelli*. *Nature*, 2002, **419**, 512–519.
15. Das, A., Population genomic and bioinformatic studies reveal evolutionary history of *Drosophila ananassae*. *Curr. Sci.*, 2005, **89**, 1316–1321.
16. Cui, L. *et al.*, The genetic diversity of *Plasmodium vivax* populations. *Trends Parasitol.*, 2003, **19**, 220–226.
17. Livingstone, F. B., Anthropological implications of sickle-cell gene distribution in West Africa. *Am. Anthropol.*, 1958, **60**, 533–562.
18. Wiesenfeld, S. L., Sickle-cell trait in human biological and cultural evolution. *Science*, 1967, **157**, 1134–1140.

19. Olago, D. O., Vegetation changes over palaeo-time scales in Africa. *Climate Res.*, 2001, **17**, 105–121.
20. Salamini, F. *et al.*, Genetics and geography of wild cereal domestication in the Near East. *Nature Rev. Genet.*, 2002, **3**, 429–441.
21. Coluzzi, M., Petrarca, V. and Di Deco, M. A., Chromosomal inversion intergradations and incipient speciation in *Anopheles gambiae*. *Boll. Zool.*, 1985, **52**, 45–63.
22. Coluzzi, M., The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. *Parassitologia*, 1999, **41**, 277–283.
23. Hey, J., Parasite populations: The puzzle of *Plasmodium*. *Curr. Biol.*, 1999, **9**, R5652–R5657.
24. Rich, S. M. *et al.*, Malaria's eve: evidence of a recent population bottleneck throughout the world population of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, 1998, **95**, 4425–4430.
25. Su, X.-Z. *et al.*, A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science*, 1999, **286**, 1351–1353.
26. Anderson, T. J. C. *et al.*, Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.*, 2000, **17**, 1467–1482.
27. Volkman, S. *et al.*, Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. *Science*, 2002, **298**, 216–218.
28. Gardner, M. J. *et al.*, Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science*, 1998, **282**, 1126–1132.
29. Bowman, S. *et al.*, The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature*, 1999, **400**, 532–538.
30. Volkman, S. K. *et al.*, Recent origin of *Plasmodium falciparum* from a single progenitor. *Science*, 2001, **293**, 482–484.
31. Mu J. *et al.*, Polymorphism in a *Plasmodium falciparum* erythrocyte-binding ligand changes its receptor specificity. *J. Exp. Med.*, 2002, **196**, 1523–1528.
32. Anderson, T. J. C., Su, X. Z., Roddam, A. and Day, K. P., Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. *Mol. Ecol.*, 2000, **9**, 1599–1608.
33. Conway, D. J. *et al.*, Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.*, 2000, **111**, 163–171.
34. Joy, D.A. *et al.*, Early origin and recent expansion of *Plasmodium falciparum*. *Science*, 2003, **300**, 318–321.
35. Watts, S., *Epidemics and History: Disease, Power and Imperialism,* Yale Univ. Press, New Haven, 1997.
36. Vaidya, A. B. *et al.*, Structural features of *Plasmodium* cytochrome-b that may underlie susceptibility to 8-aminoquinolines and hydroxyl-naphthoquinones. *Mol. Biochem. Parasitol.*, 1993, **58**, 33–42.
37. Polley, S. D. and Conway, D. J., Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics*, 2001, **158**, 1505–1512.
38. Roper, C. *et al.*, Intercontinental spread of pyrimethamine resistant malaria. *Science*, 2004, **305**, 1124.
39. Volkman, S. *et al.*, A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genet.*, 2006, **39**, 113–119.
40. Mu, J, *et al.*, Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biol.*, 2005, **3**, e335.
41. Jeffares, D. C. *et al.*, Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nature Genet.*, 2006, **39,** 120–125.
42. Mu, J. *et al.*, Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nature Genet.*, 2006, **39**, 126–130.
43. Nielsen, R., Molecular signatures of natural selection. *Annu. Rev. Genet.*, 2005, **39**, 218.
44. Trimnell, A. R. *et al.*, Global genetic diversity and evolution of *var* genes associated with placental and severe childhood malaria. *Mol. Biochem. Parasitol.*, 2006, **148**, 169–180.
45. Carlton, J. M., Towards a malaria haplotype map. *Nature Genet.*, 2007, **39**, 5–6.
46. Mendis, K. *et al.*, The neglected burden of *Plasmodium vivax* malaria. *Am. J. Trop. Med. Hyg.*, 2001, **64**, 97–106.
47. Baird, J. K., Chloroquine resistance in *Plasmodium vivax*. *Antimicrob. Agents Chemother.*, 2004, **48**, 4075–4083.
48. Feng, X. *et al.*, Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 8502–8507.
49. Escalante, A. A. *et al.*, The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 8124–8129.
50. Escalante, A. A. *et al.*, A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 1980–1985.
51. Mu, J. *et al.*, Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol. Biol. Evol.*, 2005, **22**, 1686–1693.
52. Jongwutiwes, S. *et al.*, Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. *Mol. Biol. Evol.*, 2005, **22,** 1733–1739.
53. Cornejo, A. E. and Escalante A. A., The origin and age of *Plasmodium vivax*. *Trends Parasitol.*, 2006, **22,** 558–563.
54. Gomez, J. C. *et al.*, Identification of a polymorphic *Plasmodium vivax* microsatellite marker. *Am. J. Trop. Med. Hyg.*, 2003, **69,** 377–379.
55. Leclerc, M. C. *et al.*, Meager genetic variability of the human malaria agent *Plasmodium vivax*. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 14455–14460.
56. Imwong, M. *et al.*, Microsatellite variation, repeats array length, and population history of *Plasmodium vivax*. *Mol. Biol. Evol.*, 2006, **23**, 1016–1018.
57. Russel, B., Suwanarusk, R. and Lek-Uthai, U., *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends Parasitol.*, 2006, **22**, 399–401.
58. Del Portillo, H. A. *et al.*, Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different *Plasmodium* species. *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 4030–4034.
59. Galinski, M. R. *et al.*, *Plasmodium vivax* merozoite surface protein-3 contains coiled-coil motifs in an alanine-rich central domain. *Mol. Biochem. Parasitol.*, 1999, **101**, 131–147.
60. Cheng, Q. and Saul, A. J., Sequence analysis of the apical membrane antigen I (AMA-1) of *Plasmodium vivax*. *Mol. Biochem. Parasitol.*, 1994, **65**, 183–187.
61. Figtree, M. *et al.*, *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. *Mol. Biochem. Parasitol.*, 2000, **108**, 53–66.
62. Snewin, V. A. *et al.*, A new polymorphic marker for PCR typing of *Plasmodium vivax* parasites. *Mol. Biochem. Parasitol.*, 1995, **71**, 135–138.
63. Prajapati, S. K. *et al.*, Allelic dimorphism of *Plasmodium vivax gam-1* in the Indian subcontinent. *Malaria J.*, 2006, **5**, 90.
64. Das, A. and Dash, A. P., Evolutionary paradigm of chloroquine resistant malaria in India. *Trends Parasitol.*, 2007, **23**, 132–135.
65. Rodriguez-Morales, A. J. *et al.*, Impact of imported malaria on the burden of disease in northeastern Venezuela. *J. Travel Med.*, 2006, **13**, 15–20.

ACKNOWLEDGEMENTS. We thank Prof. A. P. Dash, Director, NIMR, Delhi for providing facilities and encouragements and Priti Kumari and Surbhi Verma for assistance in literature surveys and drawing of the figures. SM thanks the Vice-Chancellor, JIITU for facilities and encouragement. The intramural funding from the Indian Council of Medical Research is thankfully acknowledged.