RESEARCH NOTE

Development of nuclear DNA markers for evolutionary studies in *Plasmodium falciparum*

CELIA THOMAS, SNEH SHALINI, N. RAGHAVENDRA, MEENAKSHI CHOUDHARY, ANJU VERMA, HEMA JOSHI, A. P. DASH and APARUP DAS*

National Institute of Malaria Research (ICMR), 22 Sham Nath Marg, Delhi 110 054, India

Introduction

Estimation of genetic diversity in species populations and inferring the evolutionary dynamics of different genes are important in biomedical research, especially in finding new drug target genes and developing new effective drugs and vaccines. Recent researches in evolutionary genetics have revealed that estimation of genetic diversity is strongly dependent on the genetic markers used, thus making appropriate evolutionary inference at species and gene levels difficult. Considering these facts, we have used published whole-genome sequence information to develop nuclear DNA markers in the human malaria parasite *Plasmodium falciparum* that would help in understanding the precise roles of demography and natural selection in the evolution of *P. falciparum*.

Selection of genetic loci in evolutionary studies has important bearings on the estimation of genetic diversity, which is the baseline to infer the evolutionary history of the genes of interest and of species populations. This is primarily because unequal extents of genetic diversity have been detected in different genes along the length of chromosomes in an array of organisms, and this variation has been interpreted in terms of several evolutionary forces acting differentially on the genome (Shen *et al*. 2000; Wilder *et al*. 2004). Recombination has been shown to have a greater role in generating genetic variation in the chromosome and there seems to be a high positive correlation between the rate of recombination and genetic diversity (Begun and Aquadro 1992; Glinka *et al*. 2003; Haddrill *et al*. 2005). This has made it necessary for evolutionary biologists to choose genetic markers in appropriate genomic regions to answer specific questions at hand (e.g. distinguishing between demography and natural selection). As shown recently with *Drosophila*, it is now possible to disentangle the effects of different evolutionary forces with careful choice of DNA fragments in the genome (Baines *et al*. 2004; Das *et al*. 2004; see also Das 2005 for a review).

The parasite *P. falciparum*, responsible for most human malaria, is among the most studied pathogens of all time, probably surpassed only by the human immunodeficiency virus and the tuberculosis bacterium *Mycobacterium tuberculosis* (Doolittle 2002). Malaria is one of the most important diseases of humans over the world. It is heavily endemic in most of the tropical and subtropical countries and is becoming endemic in some temperate countries. Fighting against this dreadful disease requires correct estimation of genetic diversity and evolutionary history across the whole distribution range of both the parasites and the vectors. This is especially important since reports of field resistance in the parasite against antimalarial drugs and in the insect vectors against insecticides are pouring in. Understanding the genetic mechanism(s) of drug resistance and generating strategies for development of antimalarial drugs and vaccines are thus the need of the hour. In view of this, correct estimation of genetic diversity in the malaria parasite, and knowledge of the evolutionary forces (demography and/or natural selection) that govern it, should be the first step in the process.

Material and methods

The *P. falciparum* nuclear genome comprises 14 linear chromosomes, ranging in size from 0.64 Mb (chromosome 1) to 3.3 Mb (chromosome 14), and two nonnuclear genomes: a compact mitochondrial genome of 6 kb and a plastid-like, 35-kb circular genome that resides in an organelle known as the apicoplast (Waller and McFadden 2004). First-pass annotation of the *P. falciparum* genome in late 2002 irrevocably changed the face and practice of malaria research (Kooij *et al*. 2006). This comprehensive data set, combined with the availability of substantial sequence tracts from the rodent

*For correspondence. E-mail: aparup@mrcindia.org.

Keywords. nuclear DNA markers; genetic diversity; *pfcrt*; demography; natural selection; *Plasmodium falciparum*.

malaria parasite *P. yoelii* (Carlton *et al*. 2002), along with the soon to be published *P. vivax* genome, have made it possible to embrace the latest global-genome-survey technologies.

To initiate genetic diversity estimation and population structure of Indian *P. falciparum*, we developed nuclear DNA markers for detection of single-nucleotide polymorphisms (SNPs). For designing the putatively neutral DNA fragments (that bear no or very weak signals of past selection events) which would help in deciphering the demographic history of *P. falciparum*, we scanned the whole genome of *P. falciparum* available in the public domain and isolated introns in every kind of gene (known, putative and hypothetical). We followed the approach described by Das *et al*. (2004) to choose these putatively neutral DNA fragments, except in distinguishing genes according to the recombination environment, as the rate of recombination for each chromosome is incompletely known in *P. falciparum*. Specifically, we considered introns 450–850 bp long (see electronic supplementary material at http://www.ias.ac.in/jgenet) and designed primers in the exons flanking these introns (exon priming intron crossing, or EPIC fragments). The lengths of the fragments have been kept relatively bigger than suggested by Das *et al*. (2004) in view of the reported rarity of SNPs in *P. falciparum* introns (T. J. C. Anderson, personal communication). We scanned the whole genome of *P. falciparum* through the NCBI 'sequence viewer', which provides a clear distinction between the coding and noncoding region with the coded amino acid. This helps in detecting the start and end of the introns fairly easily. Since the nucleotide positions are also shown in 'sequence viewer', we could get the sequence of each intron of interest from the sequence download page just by providing the start and end positions. For additional clarification, we used the PlasmoDB (http://www.plasmodb.org) database. We tried several online tools, e.g. Gene Fisher (http://bibiserv.techfak.uni-bielefeld.de/genefisher/), Oligo Analyzer (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/), PrimerSelect (a unit of DNASTAR) and Primer 3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) to design primer sequences in the exons to amplify the selected flanking intron fragments. Of all the tools, we found the programs Primer 3 and PrimerSelect to be very simple and effective for our work and designed all the primers using these two tools (see electronic supplementary material).

Results and discussion

The total numbers of introns across all the 14 chromosomes of *P. falciparum* and the selected introns that are considered to be putatively neutral are shown in figure 1. The total number of introns is positively correlated with the size of the chromosome and the correlation was found to be statistically significant (Pearson's $r = 0.92$, $P < 0.001$). Similarly, the size of the chromosome and the number of genes present also have a statistically significant positive correlation ($r = 0.99$, $P < 0.001$). However, these were not true for the selected

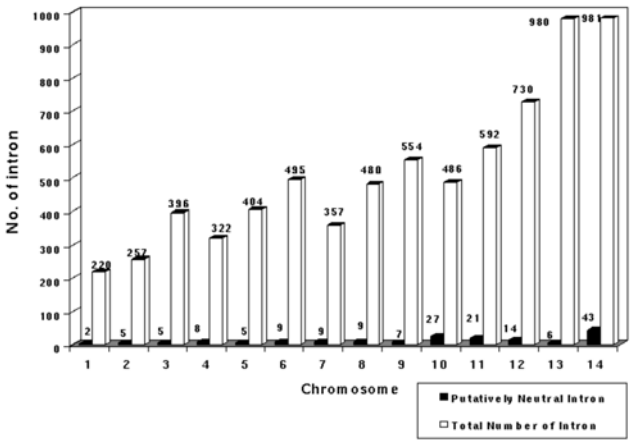

Figure 1. Distribution of total number of introns and selected putatively neutral introns in *P. falciparum* chromosomes.

putatively neutral fragments; these fragments were found to be distributed randomly across different chromosomes, without corresponding to either the size of the chromosome, the number of genes present, or the number of introns. Although a minimum number of two putatively neutral loci could be detected in the smallest chromosome (chromosome 1) and a maximum of 43 in the biggest (chromosome 14), it is clear that these DNA markers are randomly distributed in chromosomes (figure 1). Surprisingly, we could only detect six putatively neutral fragments in chromosome 13, which is the second largest chromosome in *P. falciparum*. It is to be noted that the total number of introns is almost the same in chromosomes 13 and 14 (980 in 13, 981 in 14) and there is not much difference in the sizes of these chromosomes (2732 kb in chromosome 13, 3292 kb in chromosome 14). Thus, it is certain that there is an enormous variation in the intron size distribution across the chromosomes, and chromosome 13 is biased towards introns of extreme sizes (either less than 450 bp or higher than 850 bp length). The same might also be true for chromosomes 5, 7 and 9. In total, we found 170 putatively neutral intron fragments across the whole genome and are at present amplifying them using PCR in Indian *P. falciparum*. The best amplified fragments with about five SNPs in a population would be the criteria (for details, see Das 2005) to further consider them in population survey in Indian *P. falciparum* for estimating genetic diversity governed by genetic drift.

To disentangle effects of genetic drift and natural selection in shaping genetic diversity in Indian *P. falciparum* populations, we considered a single gene named *pfcrt*, present on chromosome 7, along with the putatively neutral loci. Numerous studies have shown that mutations in this gene confer resistance to the antimalarial drug chloroquine (Cooper *et al*. 2005). It has been suggested that resistant varieties of *P. falciparum* have multiply originated in South-East Asia and South America, and then spread all over the world where

malaria is endemic (Anderson and Roper 2005), providing an impression that the drug-resistance form (haplotype) of the *pfcrt* gene has been swept by natural selection. Although different mutations and some haplotypes have been suggested to be responsible for providing resistance against the drug (e.g. *K76T*, *A220S*) (Cooper *et al*. 2005), it is important to determine which specific haplotype(s) of the gene (considering the whole gene) is responsible for resistance. Thus, with the goal of studying the distribution of different haplotypes with respect to the *pfcrt* gene, and their diversity and evolutionary history in India, we developed different fragments to sequence the whole gene. Figure 2 shows the structure of the whole *pfcrt* gene and the fragments we have designed in this study. As depicted in the figure, we have divided the whole *pfcrt* gene into three different fragments for amplification by PCR and five internal fragments (within the three main fragments) for sequencing purposes. The list, and the nucleotide sequences, of the primers used are shown in table 1.

Although there are 5268 predicted protein-encoding genes, at an average frequency of one gene every 4.3 kb reported, in the *P. falciparum* genome (Kooij *et al*. 2006), we could only predict 170 introns that could be considered as partially putatively neutral. This might be due to the fact that the *P. falciparum* genome contains a large number of genes without introns, and the distribution of length of introns is highly skewed towards small introns. There are relatively very few introns whose size is more than 450 bp, and genes with each introns seem to be distributed randomly across the whole genome. Hence we could not find any correlation between the introns of our interest either with the number of genes in a chromosome or with the total number of introns in a chromosome.

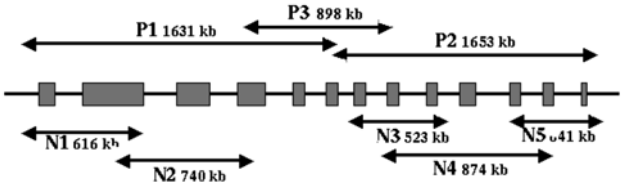

Figure 2. *pfcrt* gene in *P. falciparum*. P1–P3 are fragments for amplification of the whole gene and N1–N5 are fragments used as internal fragments for sequencing within P1 and P2. The list of the primers used is provided in table 1.

Table 1. Description of primers for amplification of the whole *pfcrt* gene in *Plasmodium falciparum*.

Primer type	Primer sequence	Product length (bp)
Primary		
P1	F 5′ CCGTTAATAATAAATACACGCAG 3′ * R 5′ GTTTTAAAAATGGAAGGGTGTATAC 3′ *	1631
P2	F 5′ ACCTGTATACACCCTTCCATTTTT 3′ * R 5′ CCTTATAAAGTGTAATGCGATAGC 3′ *	1653
P3	F 5′ TCGGAGCAGTTATTATTGTTG 3′ R 5′ TTACCCATGCTCCGTCAC 3′	898
Nested		
N1	F 5′ TAATAATAAATACACGCAGTCATA 3′ R 5′ TGAATTTCCCTTTTTATTTCCAAA 3′ **	616
N2	F 5′ GTAACATCCGAAACTCACAA 3′ R 5′ AAGGCACTAATTAAGACAAGA 3′	740
N3	F 5′ CCCTTTTTAGTTCATTTACCATA 3′ R 5′ ATTCCCATATTTATTTCCTCTTG 3′	523
N4	F 5′ GGTATGGCTAAGTTATGTGATGA 3′ R 5′ CAAATAGGTAGCCAAACTGTAAA 3′	874
N5	F 5′ TATAGGGTGATGTTGTAAGAG 3′ R 5′ ATAAAGTGTAATGCGATAGC 3′	641

*Reported earlier (Dorsey *et al*. 2001); **reported earlier (Vathsala *et al*. 2000).

The development of nuclear DNA markers has far-reaching significance in malaria research. While, on one hand, the putatively neutral fragments would help in estimating genetic diversity of the parasite populations that is due to genetic drift (demographic events), on the other they would also help in deciphering population genetic structure, historical population expansion, and population bottlenecks. Likewise, the fragments we report here for sequencing the whole *pfcrt* gene would not only help in discovering new population-specific endemic mutations, but also in finding existing associations between mutations and different *pfcrt* haplotypes. This is of immense importance because, except for detection and reporting of specific mutations, a population-genetic survey of the whole *pfcrt* gene has remarkably been overlooked.

Acknowledgements

This investigation received financial support from the UNICEF / UNDP / World Bank / WHO Special Assistance Program for Research and Training in Tropical Diseases (TDR), Geneva. We are thankful to O. P. Singh and P. K. Mallick for their help at several stages of primer designing. Special thanks to N. Nanda for his generous help and allowing use of his computer facilities.

References

Anderson T. J. C. and Roper C. 2005 The origins and spread of antimalarial drug resistance: lessons for policy makers. *Acta Trop*. **94**, 269–280.

Baines J. F., Das A., Mousset S. and Stephan W. 2004 The role of natural selection in genetic differentiation of worldwide populations of *Drosophila ananassae*. *Genetics* **168**, 1987–1998.

Begun D. J. and Aquadro C. F. 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**, 519–520.

Carlton J. M., Angiuoli S. V., Suh B. B., Kooij T. W., Pertea M., Silva J. C. *et al*. 2002 Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519.

Cooper R. A., Hartwig C. L. and Ferdig T. M. 2005 *pfcrt* is more than the *Plasmodium falciparum* chloroquine resistance gene: a functional and evolutionary perspective. *Acta Trop*. **94**, 170–180.

Das A. 2005 Population genomic and bioinformatic studies reveal evolutionary history of *Drosophila ananassae*. *Curr. Sci*. **89**, 1316–1321.

Das A., Mohanty S. and Stephan W. 2004 Inferring the population structure and demography of *Drosophila ananassae* from multilocus data. *Genetics* **168**, 1975–1985.

Doolittle R. F. 2002 The grand assault. *Nature* **419**, 493–494.

Dorsey G., Kamya M. R., Singh A. and Rosenthal P. J. 2001 Polymorphisms in the *Plasmodium falciparum pfcrt* and *pfmdr-1* genes and clinical response to chloroquine in Kampala, Uganda. *J. Infect. Dis*. **183**, 1417–1420.

Glinka S., Ometto L., Mousset S., Stephan W. and De Lorenzo D. 2003 Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: A multi-locus approach. *Genetics* **165**, 1269–1278.

Haddrill P. R., Thornton K. R., Charlesworth B. and Andolfatto P. 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res*. **15**, 790–799.

Kooij T. W. A., Janse C. J. and Waters A. P. 2006 Plasmodium postgenomics: better the bug you know? *Nat. Rev. Microbiol*. **4**, 344–357.

Shen P., Wang F., Underhill P. A., Franco C., Yang W.-H., Roxas A. *et al*. 2000 Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* **70**, 7353–7359.

Vathsala P. G., Pramanik A., Dhanasekaran S., Usha Devi C., Pillai C. R., Subbarao S. K. *et al*. 2000 Widespread occurrence of the *Plasmodium falciparum* chloroquine resistance transporter (*Pfcrt*) gene haplotype SVMNT in *P. falciparum* malaria in India. *Am. J. Trop. Med. Hyg*. **70**, 256–259.

Waller R. F. and McFadden G.I. 2004 The apicoplast. In *Genomes and the molecular cell biology of malaria parasites* (ed. A. Waters and C. Janse). Horizon Scientific Press, Norfolk.

Wilder J. A., Mobasher Z. and Hammer M. F. 2004 Genetic evidence for unequal effective population sizes of human females and males. *Mol. Biol. Evol*. **21**, 2047–2057.

Received 21 June 2006