RESEARCH ARTICLE

Fine-scale genetic characterization of *Plasmodium falciparum* chromosome 7 encompassing the antigenic *var* and the drug-resistant *pfcrt* genes

RUCHI BAJAJ[1], SUJATA MOHANTY[2], A. P. DASH[1] and APARUP DAS[1,*]

[1]*Evolutionary Genomics and Bioinformatics Laboratory, National Institute of Malaria Research, 22 Sham Nath Marg, New Delhi 110 054, India*
[2]*Department of Biotechnology, Jaypee Institute of Information Technology University, A-10, Sector 62, Noida 201 301, India*

Abstract

The fact that malaria is still an uncontrolled disease is reflected by the genetic organization of the parasite genome. Efforts to curb malaria should begin with proper understanding of the mechanism by which the parasites evade human immune system and evolve resistance to different antimalarial drugs. We have initiated such a study and presented herewith the results from the *in silico* understanding of a seventh chromosomal region of the malarial parasite *Plasmodium falciparum* encompassing the antigenic *var* genes (coding *pfemp1*) and the drug-resistant gene *pfcrt* located at a specified region of the chromosome 7. We found 60 genes of various functions and lengths, majority (61.67%) of them were performing known functions. Almost all the genes have orthologs in other four species of *Plasmodium*, of which *P. chabaudi* seems to be the closest to *P. falciparum*. However, only two genes were found to be paralogous. Interestingly, the drug-resistant gene, *pfcrt* was found to be surrounded by seven genes coding for several CG proteins out of which six were reported to be responsible for providing drug resistance to *P. vivax*. The intergenic regions, in this specified region were generally large in size, majority (73%) of them were of more than 500 nucleotide bp length. We also designed primers for amplification of 21 noncoding DNA fragments in the whole region for estimating genetic diversity and inferring the evolutionary history of this region of *P. falciparum* genome.

[Bajaj R., Mohanty S., Dash A. P. and Das A. 2008 Fine-scale genetic characterization of *Plasmodium falciparum* chromosome 7 encompassing the antigenic *var* and the drug-resistant *pfcrt* genes. *J. Genet.* **87**, 59–64]

Introduction

Malaria, one of the deadliest, infectious diseases is caused by protozoan parasites of *Plasmodium* species, affecting about 300 million individuals around the globe, mainly tropical regions of the world causing about one million deaths annually. The parasites have two complex life cycles: one in the host (human) and the other in the vector (mosquito), thus making an excellent model to understand the host-parasite-vector interactions. Since genes carry footprints of individual characters and the whole-genome sequencing of all the three taxa have been reported, research directing towards understanding the above interactions is the need of an hour.

Efforts of controlling malaria has been drastically affected by two very important factors: (i) the development and spread of drug resistant parasites, and (ii) wide variations in the antigenic genes, making the parasite able to evade the host immune system fairly easily. Of the several reported drug-resistant genes in the malarial parasite, *Plasmodium falciparum*, the gene *pfcrt*, located in the middle part of the chromosome 7 has been reported to be highly correlated with the clinical episode of drug resistance (Sidhu *et al.* 2002). The gene product is a predicted transporter that localizes to the digestive vacuole membrane and may be involved in drug flux and/or pH regulation (Wellems and Plowe 2001). The red blood cell surface adhesive proteins (called *pfemp1*, for *P. falciparum* erythrocyte membrane protein 1) encoded

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

Keywords. malaria; *Plasmodium falciparum*; drug resistance; *var* gene; *pfcrt*; evolutionary bioinformatics.

by the *var* genes are spread across the whole genome and are exposed to the immune system. There are at least 60 variations of the protein within a single parasite and perhaps unlimited versions within parasite populations. The parasite switches between broad repertoires of pfemp1 surface proteins, thus staying one step ahead of the pursuing immune system (Chen *et al*. 2000). The *Plasmodium* is the first of the kingdom Protista to be fully sequenced, the first of an extraordinarily diverse kingdom of 30 phyla. Sequencing the genome of *P. falciparum* is a major technical achievement — not least because its assembly was so tough because the genome is 80.6% A+T rich. The hope is that this sequence will offer new targets for both anti-malarial drugs and vaccines (Ashburner 2002). The *P. falciparum* nuclear genome comprises of 14 linear chromosomes ranging in size from 0.64 to 3.3 Mb. The genome of the 3D7 strain of *P. falciparum* was the first parasite genome to be sequenced completly (Taco *et al*. 2006). Almost 60% of the genes in *P. falciparum* have unknown functions which poses a great challenge in identifying the genes for drug resistance and vaccine targets in the genome. Thus the utilization of the whole genome sequence information must be targeted towards identifying the hitherto-unknown vaccine and drug targets. Further, considering the puzzling distribution of genetic diversity in different genes spanning across the genome of this species, a genomewide distribution of the genes which are possible targets to drugs and vaccines should be characterized first by bioinformatics methods (Das *et al*. 2007).

Here we report the results from a fine-scale scan of the DNA fragment of about 270 kb in chromosome 7 here after called as the specified region (figure 1) of the published 3D7 culture isolate (http://www.plasmodb.org) encompassing the centrally-located antigenic *var* genes and the drug-resistant *pfcrt* gene. This initial study is to closely understand this chromosomal fragment that might be a store house for very important functions of the parasite in relation to the high antigenic mimicry and evolution of drug resistance varieties. This is especially important in the light of three recent findings: (i) a strong genetic linkage between these two gene loci (Fowler *et al*. 2006), (ii) elevated polymorphisms between internal *var* locus, and the *pfcrt* gene extending over 200 kb (Mu *et al*. 2007) and (iii) involvement of gene(s) other than the *pfcrt* in providing drug resistance in Indian *P. falciparum* (Das *et al*. 2007).

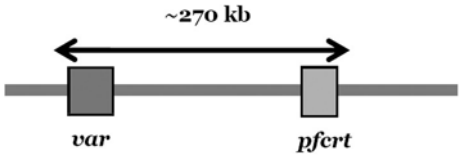

Figure 1. The ~270 kilobase specified region of *Plasmodium falciparum* chromosome 7.

Materials and methods

The DNA sequences of the specified region of *P. falciparum* 3D7 isolate were downloaded from the public domain (http://www.plasmodb.org) and scanning of the genes was performed in its protein databases and the gene IDs of the genes flanking *var* (central) and upto *pfcrt* on chromosome 7 were noted. The gene boundaries (according to PlasmoDB) were calculated manually after determining the order and location of genes (as per the PlasmoDB data resources, available at http://www.plasmodb.org). By this, the location of the intergenic regions was determined and accordingly, sequences retrieved from the sequence retrieval tool of PlasmoDB. Coding sequences (CDS) of each gene were retrieved with the help of the gene IDs from nucleotide database (http://www.plasmobd.org), through a method known as 'mundane'. The other properties of the specified region, viz. length of the intergenic regions, length of introns and exons, and length of each gene were determined manually with the help of the location given for each in the CDS. Prediction of orthologs and paralogs of genes were conducted using the database known as orthoMCL, version 1.2 (http://orthomcl.cbil.upenn.edu). The CDS of the orthologs and paralogs were retrieved with the help of blast results (http://www.plasmodb.org/plasmo/processQuestion.do). Finally, to have an inference on the genetic diversity in the neutral genetic regions of this specified region by single nucleotide polymorphism (SNP) detection through PCR amplifications and sequencing, different primers (both forward and reverse) were designed with the help of PrimerSelect, a computer software integrated as a part of the DNAstar (DNA Star, Madison, USA) version 5.0.

Results and discussions

Scanning of the specified region of ~270 kb revealed a total of 60 genes (figure 2) of which only 38.33% have been assigned some function (either putative or known) and rest 61.6% are hypothetical, indicating that the function of a large portion of this specified region is yet to be revealed. It may harbour certain genes that might be playing a very important role in the drug resistance or some other task. The probability of occurrence of genes that might be playing role in microevolution of the parasite for overcoming the human imposed control measures (either drug or human immunity) is high as the region has been shown to be probably under the natural selection (Mu *et al*. 2007). We divided these genes into eight classes based on their total length in nucleotide base pairs (figures 3 and 4). It is evident that majority of the genes within this specified region are relatively bigger in size (52 out of 60 genes are of considerably longer in size i.e., more than 500 base pairs). Further, more than 50% of the small sized (less than 500 base pair) genes are of known function and the function of the genes decreases as the gene size increases. Hence, there seems to be a negative correlation between the gene size and function. Genes responsible for translation, transcription and replication are

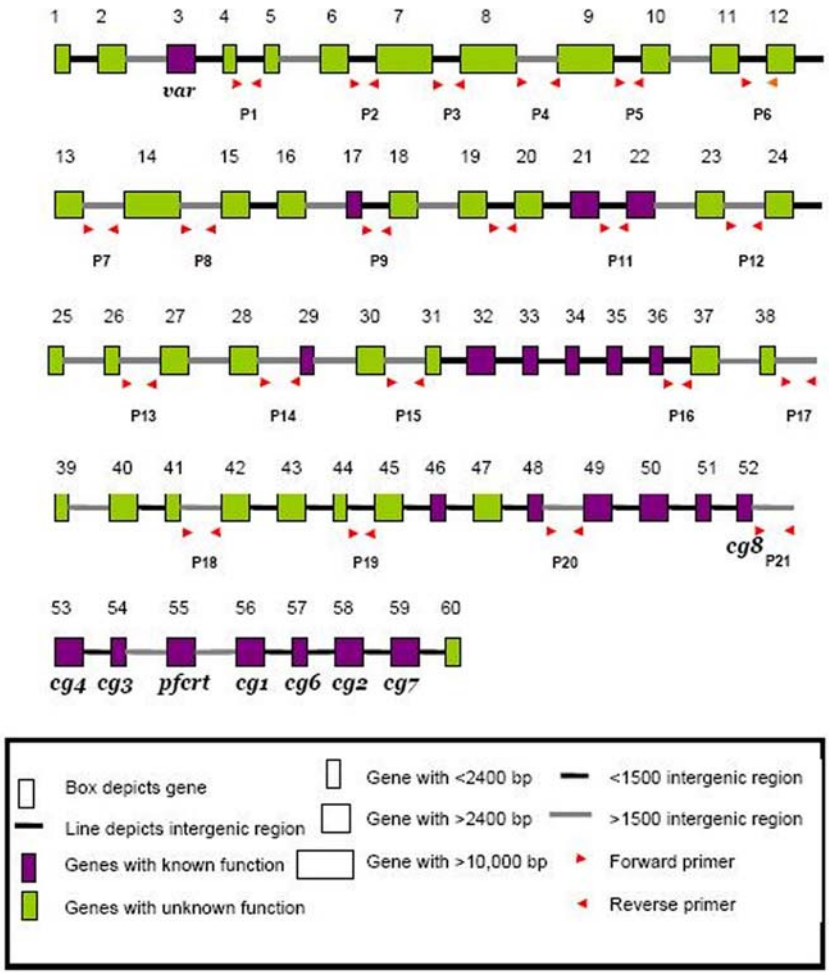

Figure 2. Fine details of the specified region showing the genes, intergenic regions and locations of the primers designed for amplification of different regions

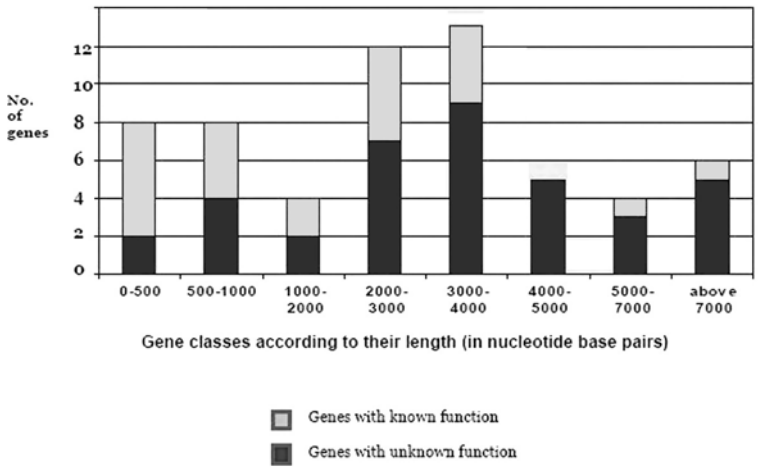

Figure 3. Distribution of genes with known and unknown function in the specified region.

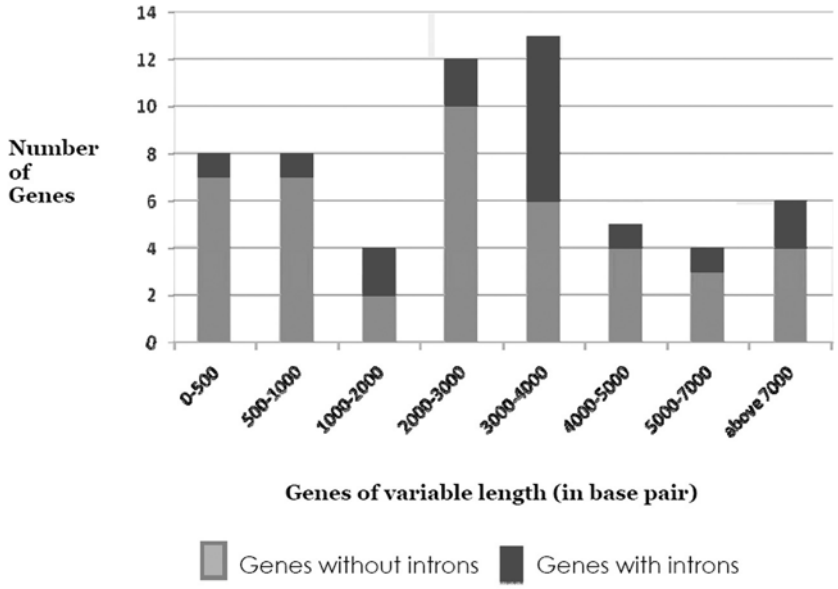

Figure 4. Distribution of gene with and without introns in the specified region.

much more prevalent than the other types, and small genes seem to be highly functional. This situation possibly provides the parasite to save energy being used in transcription (splicing mechanisms) and utilizing in some other more important tasks such as high rate of reproduction, host immunity evasion mechanisms and many more (Hilleman 2004). This prediction is further substantiated by the finding of a relatively high proportion (about 71.67%) of intronless genes in the specified region (figure 5). It was also evident that genes of relatively bigger size with introns and smaller genes without introns seem to perform important functions in the parasite. Parasites, in general are known to have reduced genome size probably due to preferential loss of genes underlying the biosynthesis of compounds that can be easily taken up from the host, such as amino acids, nucleotides and vitamins (Konstantinidis and Tiedje 2004). There is currently an

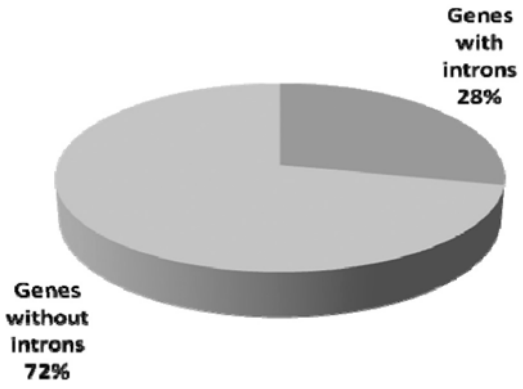

Figure 5. Percentage of genes with and without introns in the specified region.

increasing number of evidence that favours the existence of universal trends between functional gene content and genome size. Jordan *et al*. (2001) analysed 21 genomes and showed that lineage specific gene expansion is positively correlated with genome size and may account for up to 33% of the coding capacities in the genome. In a comparative genomic study involving a larger number of species, Konstantinidis and Tiedje (2004) found in *Psuedomonas aeruginosa* and *Streptomyces coelicolor*, that a disproportionate increase in regulatory and transport genes involved in secondary metabolism relative to other smaller genome-sized species (Jordan *et al*. 2001; Konstantinidis and Tiedje 2004). The same criteria are observed in *P. falciparum* genome, as seen that a large number of genes are involved in host immunity evasion and in drug resistance. The Plasmodium genome, being relatively small in size than the two above taxa, should be disproportionately enriched in protein translation, DNA replication, cell division, nucleotide metabolism genes and depleted in regulation and secondary metabolism genes (Konstantinidis and Tiedje 2004). Accordingly, the fine scale scanning of the specified region in *P. falciparum* reported here reveale that 43.47% of the known functional genes are involved in one of the process: transcription, translation or replication. Since characterization of orthologs and paralogs is critical for the construction of a robust evolutionary classification and reliable functional annotation of newly sequenced genomes (Koonin 2005), we determined both orthologs and paralogs of all the genes in the specified region across four other species of *Plasmodium*. Only two genes (Pf07_0036 and Mal7p1.12), were found to have paralogs and almost all the other genes have orthologs across all the four species (detail data not shown). In general, the genes in the specified region of

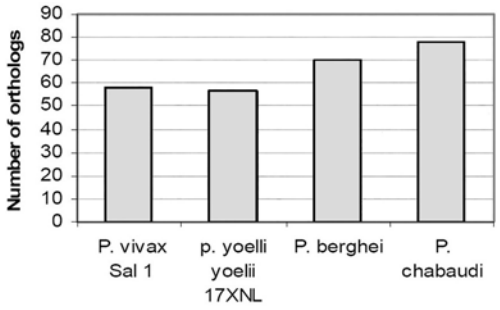

Figure 6. Number of genes in the specified region that are orthologous to other species of the genus *Plasmodium*.

P. falciparum shows more homology with *P. chabaudi* (figure 6). Such a clear and wide distinction between the number of orthologs and paralogs gives an insight of Plasmodium evolution, showing that the Plasmodium ancestors were highly evolved at the specified region under scan. A large number of orthologs in comparison to paralogs signifies that the specified region is fairly old and evolves in different lineages. From the study it also became clear that *P. chabaudi* might have shared a common ancestor with *P. falciparum*, at least at the specified region. This conclusion, however, cannot be made general because of two reasons: firstly, the *P. reichenowi* (otherwise closely related to *P. falciparum* genetically) data was not available for homology analysis and secondly, the present work involves only the specified region of about 270 kb of *P. falciparum*. The four genes coding for tRNAs in the specified region do not have any homologs, hence they can be called as unique genes. These are homologous sequences that diverge beyond recognition such that the similarity between two orthologs is not higher to be called as orthologs (Koonin 2005).

Most of the intergenic regions in the specified region spacing two genes are considerably larger, only 27% are less than 500 nucleotide base pair length and the rest are from 500–4000 base pair length (figure 2). Considering most of the intergenic regions are fairly large, and the exact roles of these regions are still unknown, it might be predicted that certain portions might be playing some important roles. Moreover, the probability of occurrence of genes that might be playing role in microevolution of the parasite for overcoming the human imposed control measures (either drug or human immunity) is high as the specified region has been shown to be probably under natural selection (Mu *et al*. 2007). Another interesting finding is the occurrence of seven membrane transporters in the vicinity of *pfcrt* gene (figure 2). These transporters are called as CG proteins and are encoded by *cg1, cg2, cg3, cg4, cg6, cg7* and *cg8* genes. It is interesting to note that six out of seven CG protein homologs are identified to be involved in providing drug resistance in *P. vivax* (http://orthomcl.cbil.upenn.edu). Moreover, Wellems *et al*. (1971) showed experimentally that all these proteins (except *cg6*) are transcribed in erythrocytic stages, suggesting their role in erythrocytic stage. The authors also surveyed parasites from Indochina and found a strong pattern of linkage disequilibrium for *cg1* and *cg2* in the chloroquine-resistant parasites. Thus, as predicted, the specified region, especially surrounding the drug-resistant gene, *pfcrt* is of great interest for understanding the genetic basis of drug resistance in Plasmodium. To estimate

Table 1. List of primers designed for amplification of different DNA fragments (spacing every ~10 Kb) of specified region of *Plasmodium falciparum*.

Primer no.	Forward primer	Reverse primer
P1	5′ATTTTGTTTGATTTGTTCTTAT3′	5′TTGATTTCCCCCTTTTT3′
P2	5′ATAGGTAATGATAACAATAAAC3′	5′ATAAAAAGGGATAAAGTAGA3′
P3	5′TCAATTTATTCCCATCCAT3′	5′ATTTGTATTTTTAAGGTGTTGT3′
P4	5′TTGCCATCATAAGAATAAT3′	5′ATAATAAAATGATAACTA3′
P5	5′ACGGGCTTCATACATTACAA3′	5′GGGAAGAAGCAAAAGGAA3′
P6	5′CAGAATGAAACCAACAAGA3′	5′GACAATTTTAAGAGAACACATAC3′
P7	5′TGCGTTTCTTTCTTTTTAC3′	5′TTTGGTTTTTATTTTGTTTTAGT3′
P8	5′TATGAGATGCCCTAAAT3′	5′AAAAATCAGAAATAAGAGG3′
P9	5′AGCTTCTGTTCCACCTG3′	5′AATATTAATTTATCCAACTTTT3′
P10	5′GAACACAAATGGAAGAATAAA3′	5′CAAAACATAAAGGAGCACAAT3′
P11	5′TACGGAGTTGAAAAAGAATC3′	5′TAGTAACGTAAATGAAAAACAAA3′
P12	5′ATCAATAAATATCTACCGTCTG3′	5′AAAATAATAAAAATAAAATACAAG3′
P13	5′AAAAACTTATTCATCATT3′	5′TATTTTATTATTTTCATTTTA3′
P14	5′GTGTACTACCGATGAAGG3′	5′ATATAACGGTTGTGATGAG3′
P15	5′TCCGCCCTATTCTCATT3′	5′TTGCTTCTCCCACATTCT3′
P16	5′GGCGCTCTACCAACTGA3′	5′TATTATTTATTTTTCCTTATCTTC3′
P17	5′CCGCTATGACTGCTGAAG3′	5′AGGGTGACCATAAGGAATACT3′
P18	5′AAAAGAGGATTAGCCAAAAC3′	5′TACAAAAGCCATACAAATCAC3′
P19	5′GAAAAGTTATATGGGCAAAAT3′	5′TACGAGCAGGAAGTGACC3′
P20	5′GTTAAACAAAGGGGAAG3′	5′ATATATCAAGGCAACATT3′
P21	5′ACCCTCCTTGCGTAAAT3′	5′TAAGGGAAATGAATAAGGAA3′

genetic diversity in the specified region, 21 primer pairs were designed for the amplification of 21 different fragments of noncoding DNA. The primer sequences (both forward and reverse) have been presented in table 1 and figure 2. Since it is apparent that the specified region is of high importance as far as the essential functions of the parasite *P. falciparum* is concerned, the genetic diversity estimates would provide an evolutionary perspective of the region which ultimately would specify the role of natural selection on the maintenance of unusual amount of genome diversity that *P. falciparum* possesses.

The specified region is important for bearing the two most important genetic components by the most lethal form of human malarial parasite *P. falciparum*: (i) the chloroquine-resistance gene, and (ii) the antigenic gene. Most of the genes in this region were found to be of unknown function and small genes were found to be highly functional. Most of the genes are relatively bigger in size and majority were without introns. Interestingly, almost all the genes have orthologs across other *Plasmodium* species but only two genes have paralogs. *P. chabaudi* seems to be the most closest among all, although very marginally higher than others. The intergenic regions were found to cover large stretches of DNA and seven CG proteins are found to be flanked with the chloroquine–resistance gene, *pfcrt*. The present study thus would open up the field to further evolutionary understanding of the specified region, as a recent study indicated a high linkage disequilibrium between the *var* and the *pfcrt* genes (Fowler *et al*. 2006).

Acknowledgements

Intramural funding from the Indian Council of Medical Research (ICMR), New Delhi is gratefully acknowledged. SM thanks the Vice-chancellor, Jaypee Institute of Information Technology University for facilities and encouragements.

References

Ashburner M. 2002 A hat trick - *Plasmodium, Anopheles* and *Homo. Genome Biol*. **4**, 103.

Chen Q., Schilchtherle M. and Wahlgren M. 2000 Molecular aspects of severe malaria. *Clin. Microbiol. Rev*. **13**, 439–450.

Das A. and Dash A. P. 2007 Evolutionary paradigm of chloroquine-resistant malaria in India. *Trends Parasitol*. **23**, 132–135.

Das A., Bajaj R., Mohanty S. and Swain V. 2007 Genetic diversity and evolutionary history of *Plasmodium falciparum* and *P. vivax*. *Curr. Sci*. **92**, 1516–1524.

Fowler E. V., Chavchich M., Chen N., Peters J. M., Kyle D. E., Gatton M. L. and Cheng Q. 2006 Physical linkage to drug resistance genes results in conservation of *var* genes among west pacific *Plasmodium falciparum* isolates. *J. Infect. Dis*. **194**, 939–948.

Hilleman M. R. 2004 Strategies and mechanisms for host and pathogen survival in acute and persistent viral infections. *Proc. Natl. Acad. Sci. USA* **101**, 14560–14566.

Jordan K., Makarova K. S., Spouge J. L., Wolf Y. L. and Koonin E. V. 2001 Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res*. **11**, 555–565.

Konstantinidis T. K. and Tiedje M. J. 2004 Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. USA* **101**, 3160–3165.

Koonin E. V. 2005 Orthologs, paralogs and evolutionary genomics. *Annu. Rev. Genet*. **39**, 309–338.

Mu J., Awadalla P., Duan J., McGee K. M., Keebler J., Seydel K., McVean G. A. T. and Su X. Z. 2007 Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat. Genet*. **39**, 126–130.

Sidhu A. B. S., Verdier-Pinard D. and Fidock D. A. 2002 Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by *pfcrt* mutations. *Science* **298**, 210–213.

Taco W. A., Kooij J. C. J. and Waters A. P. 2006 *Plasmodium* post-genomics: better the bug you know? *Nat. Rev. Microbiol*. **4**, 344–357.

Wellems T. E. and Plowe C. V. 2001 Chloroquine-resistant malaria. *J. Infect. Dis*. **184**, 770–776.

Wellems T. E., Su X., Kirkman L. A., Fujioka H. and Wellems T. E. 1971 Complex Polymorphisms in a ~330 kDa protein are linked to chloroquine-resistant *Plasmodium falciparum* in Southeast Asia and Africa. *Cell* **91**, 593–603.

Received 12 August 2007; accepted 14 November 2007
Published on the Web: 12 March 2008