J Vect Borne Dis 40, March–June 2003, pp 9–15

Malaria genome project and its impact on the disease

Poonam Verma & Y.D. Sharma*

Department of Biotechnology, All India Institute of Medical Sciences, New Delhi, India , ydsharma@hotmail.com.

Malaria remains uncontrolled to-date due to lack of effective parasite and vector control strategies. With the completion of the host, parasite and vector genome projects more suitable and effective disease control measures can be achieved. Here we have reviewed the *Plasmodium falciparum* genome project and its impact on malaria research in future. The parasite genome project has revealed certain metabolic pathways which can be targeted to develop antimalarial drugs. It has also identified large number of potential antigens for the future potential vaccines. Now the researchers in the malaria field can plan to take up the studies, which can yield more fruitful results within the limited financial resources using bioinformatics, proteomics, structural, functional and comparative genomics, etc.

Key words Apicomplexon – *P. falciparum* genome – mitochnodrial 6 kb element

Malaria is the most serious and wide spread parasitic disease of humans. Around 40% of the world's population reside in malaria affected areas. Each year, there are approximately 400 million cases with 2.5 to 3 million deaths due to malaria[1]. The greatest malaria mortality rates have been estimated in Sub-Saharan Africa where children account for 90% of all deaths[2]. Of the four *Plasmodium* species, which use human as their vertebrate host, *Plasmodium falciparum* is the most lethal and probably the third most studied pathogen after HIV and TB[1,3].

P. falciparum has a complex life cycle involving two different hosts—human and the female *Anopheles* mosquito. In each host, the parasite passes through several stages to grow and multiply further. Large number of stage specific genes are expressed by the parasite for its survival in its hostile host. With the development of *in vitro* culture (for the clinically relevant intra-erythrocytic stages of the parasite) and the molecular tools (such as transfection, microarray, etc.) the *P. falciparum* biology has evolved very rapidly over the last 25 years.

It was the year 1996, when an International Consortium of Scientists from various institutions set a goal to determine the sequence of *P. falciparum* genome because the human genome project was already progressing at a satisfying speed[4]. After seven years, in October 2002, the project was completed which provided a vast genetic information. Also to delight of malariologists, the genome of the anopheline vector was simultaneously completed. Now, we have all the genetic information of all the three partners involved in the game — the human host, the mosquito vector, and *Plasmodium* (the parasite). This should expedite research efforts in the field of malaria. Here, we have attempted to review the information

**Corresponding author*

emerged from the parasite genome project and its utility in future malaria research.

Genome in nutshell

Nuclear genome: The 22, 853, 764 bases of the genome are organised in the form of 14 chromosomes ranging in size from 0.643 to 3.24 Mb where chromosome 1 is the smallest and 14 the largest. Sequencing of the *P. falciparum* chromosomes 1, 3, 4, 5, 6, 7, 8, 9 and 13 has been performed by the Sanger Institute, UK, whereas chromosomes 2,10,11 and 14 by TIGR, USA, and the chromosome 12 by the Stanford Genome Technology Centre. The highly A+T rich (80.6%) genome has around 5,268 genes. Seventy per cent of these gene products have been detected by proteomic analysis and have also matched the expressed sequence tags (ESTs). Analysis has shown that 52.6% of the genome constitutes coding region. On an average there are 2.39 exons per gene with 23.7% G+C content. Roughly 54% of the genes have been predicted to contain introns. A total of 7,406 introns have been identified with an average length of 178.7 bp with G+C content of 13.5%. As expected introns and other non-coding regions are more A+T-rich than the exons. Sequencing has revealed 43 tRNA genes binding to all codons except the two coding for cysteine (TGT and TGC). There are three 5S rRNA genes and seven genes for other rRNAs.

35 kb circular DNA: The 35 kb circular apicoplast encodes for only 30 proteins[5]. Rest of the proteins are imported from the cytoplasm and thus are nuclear encoded. Laboratory and computational based methods have estimated approximately 550 proteins which are targeted to the apicoplast[6]. These include important enzymes catalysing DNA replication, repair, transcription, translation and post-translational modifications, cargo proteins and proteins for other activities. No photosynthetic genes have been found.

Mitochondrial genome: The mitochondrial genome, present in all malarial species, consists of a multicopy 6 kb tandemly repeated element. It is conserved across the phylum apicomplexan[7–9]. Electron microscopy detected primarily linear forms of varying length, perhaps reflecting the variance in number of tandemly repeated 6 kb element. Nucleotide sequence has been determined from various *Plasmodium* species which is highly conserved[10–13].

The 6 kb elements of all the *Plasmodium* species are known to have genes for cytochrome oxidase subunit I & III and cytochrome b. Majority of the mitochondrial proteins are, therefore, nuclear coded. This element is also found to contain large number of fragmented ribosomal RNA genes but no tRNA gene[14]. Malarial mitochondrial genome has a slightly different base composition (68% A+T) from that of the nuclear genome. It has been shown that the majority of the mitochondrial DNA consisted of polydispersed head to tail tandem arrays of the 6 kb element in the size range of 6–23 kb and circular forms account only for approximately 1–2%[15].

Proteome: The genomic data of *P. falciparum* provided a foundation to perform proteomic studies. Since, the life cycle of this organism consists of several distinct stages, each requiring specific expression of genes, the proteomic studies were carried out for different stages. Multi dimensional protein identification technology was applied to characterise the proteomes. The technology combines in-line, high resolution liquid chromatography and tandem mass spectrometry[16]. Total 5,268 proteins have been predicted out of which 3,208 are categorised into hypothetical proteins. A total of 551 proteins are targeted to the apicoplast and 246 to the mitochondria. About 1,631 proteins were found to contain transmembrane domains and 544 contained signal peptides[17].

Results of comparative proteomics from sporozoite, merozoite, trophozoite and gametocyte stages showed that they contained 1,049, 839, 1,036 and 1,147 proteins respectively[18]. Only 152 proteins, which constitute 6% of the total proteome, were found to be common to all stages. These common proteins include transcription factors, ribosomal proteins, histones and

cytoskeletal proteins. Protein analysis has also shown that the proteins expressed in the sporozoite stage are different from the other stages. On the contrary, merozoites, trophozoites and gametocytes have only 20% stage specific proteins[18].

Functional features of the genome

Enzymes: Out of 5,268 genes, the products of 733 genes have been identified as enzymes, playing an important role in the metabolism of the organism. Of them 435 proteins have been assigned the Enzyme Commission (EC) numbers. Enzymes required in glycolysis, TCA cycle, pentose phosphate pathways are present but a gene encoding fructose bis-phosphate could not be detected. Identified genes encoding important enzymes include ATP synthase, purine transporters and nucleoside interconversion enzymes, heme biosynthesis enzymes, enzymes involved in type II fatty acid biosynthesis (except thioesterase), three enzymes of the mevalonate-independent isoprenoid synthesis—1-deoxy-D-xylulose-5-phosphate synthase, 1-deoxy-D-xylulose-5-phosphate reductoisomerase and 2C-methyl-D-erythritol 2,4-cyclo diphosphate synthase[19, 20]. *P. falciparum* also have genes encoding enzymes for chorismate synthesis from erythrose 4-phosphate and phosphophenol pyruvate through shikimate pathway[21–24].

Transporter and secretary proteins: Genomic analysis shows a very limited number of transporter proteins as compared to other eukaryotes correlating with the lower percentage of multi spanning membrane proteins. Identified transporters include water/glycerol channel proteins, glucose/proton symporters, one expected sugar transporter, three carboxylate transporters, two nucleoside/nucleobase transporters, nine members of the mitochondria carrier family including an ATP/ADP exchanger[25] and a di/tri carboxylate exchanger which is perhaps involved in the transport of TCA cycle intermediates across mitochondrial membrane, phosphoenol pyruvate/phosphate and sugar phosphate/phosphate antiporters, transporters of inorganic ions, drugs and hydrophobic compounds, Na^+/proton, Ca^{++}/proton exchangers, metal cation transporters, all units of V-type ATPases, proton pumping pyrophosphatases and a number of other transporters. During the analysis no amino acid transporter was identified[1].

Homologues of important components of SRPs, translocons, signal peptidase complex, vesicle assembly components and proteins for docking and fusion like Cop I, Cop II, clathrin, adaptin, GTP binding proteins, etc. have been found.

Nucleic acid related proteins: The core proteins involved in nucleotide excision repair (XPB/Rad 25, XPG/Rad 2, XPF/Rad 1, XPD/Rad 3, ERCCI) and for homologous recombinational repair (Rad 50, Rad 51, MRE II and DMC I) are present in *P. falciparum* genome but their accessory proteins could not be found. The ability of *P. falciparum* to perform post-replication mismatch repair proves with the presence of Mut L and Mut S homologues. Genes or their homologous encoding enzymes for non homologous end joining were absent explaining some of the unusual properties of the telomeres in this species[1].

Proteins for immunological avoidance: 59 *var*, 149 *rif* and 28 *stevor* genes coding for *P. falciparum* erythrocyte membrane proteins (pfEMP1), repetitive interspersed family (rifin) and subtelomeric variable ORF (stevor) are present. The products of the *var* genes are transported to the membrane and infected RBCs. The proteins mediate adherence to host endothelial receptors[26]. *Var* gene products and rifins are responsible for antigenic variation leading to the immune evasion allowing chronic infection and transmission thus var proteins are principally involved in the induction of protective immunity. Rifins are also expressed on the surface of infected red cells. The exact function of rifins and stevors is unknown and needs to be investigated.

Possible harvestings from the genome

Facilitating basic malaria research: This is certain that *P. falciparum* genome is going to boost the basic

research of this parasite, manifolds. The genome will help to understand the *P. falciparum* biology more clearly as well as it will provide a firm foundation for starting new research. For example with the help of the genes involved in protein transport, probes can be made which will help in establishing the mechanisms as well as defining the components of protein transport systems. In the genome a number of genes with unknown function, have been predicted. With the help of their sequence, their protein functions will be determined.

The availability of *P. falciparum* genome will clear the way for more *in silico* analysis of stage specific transcription patterns. The comprehensive study of its biology is needed in order to search for novel drug targets and vaccine developments.

Drug targeting: With the help of the malaria genome data, various biochemical pathways important for the survival of the malaria parasite can be targeted for drug development. The study of a number of genes encoding different transcription factors, proteins involved in signal transduction, enzymes, etc. will help in the screening for new antimalarial drugs. For example, even after the years of study the exact mode of action of quinoline antimalarial and the mechanisms of parasite resistance to these drugs are still not completely understood. If the molecular targets of the quinolines could be identified and the molecular basis of resistance defined, it might be possible to develop new drugs that target the same metabolic processes but somehow evade the resistance mechanisms. Malaria parasite induces new permeation pathways, in part, by increased activity of endogenous transporters thereby increasing the permeability of malaria infected erythrocytes to a wide range of solutes. With the help of genome data, genes encoding transporter proteins as well as proteins themselves can be exploited as drug targets.

Serial analysis of gene expression (SAGE), combined with the information available through the malaria genome project, could increase our knowledge and point the way to new drug targets. The technology can be used to identify genes (or gene clusters) whose expression levels are affected when the parasite is treated with cytotoxic drug. Different responses in parasites that are resistant to a particular drug can also be monitored.

Vaccine development: The most important use of genome data would be to develop an effective vaccine against this disease. Effective vaccine aims to stimulate the humoral and/or cellular immune response. Antibodies, produced by B-lymphocytes are effective in the protection against extracellular stages of viruses and parasites, whereas cell mediated immunity reflected by cytolytic lymphocytes and production of cytokines by T-helper cells of Th1 subset (particularly the macrophage-activating lymphokine interferon gamma IFN-γ), is essential to defend the host against intracellular stages of the parasites.

The *P. falciparum* genome will enhance the vaccine development by the detection of the potential antigens, that could be screened for derived characteristics like surface expression or limited antigenic diversity by stage-specific gene expression and proteomic analysis.

On one hand the proteome of the *P. falciparum* will facilitate the discovery of novel subunit vaccines composed of purified proteins and polysaccharide antigens, on the other hand the genomic sequence will facilitate the discovery of new DNA vaccines in which an antigen coding gene is inserted into suitable expression vector (plasmid) and the purified recombinant vector encoding the immunogen is injected into the host.

Comparative genomics: Comparative studies are very useful for the annotation of any unannotated genome. The genome sequence of *P. falciparum* will provide a mean to the other ongoing sequencing projects of the other species of *Plasmodium* for their annotation. The function of a gene can be predicted by comparing the homology between its nucleotide sequence with the sequences of other genes of known function. Once the significant homology is established,

experiments can be performed in the right direction to confirm them.

Comparative genomics is also useful to unfold the evolutionary history among genes, proteins or the whole organisms. Comparative studies of *P. falciparum* genome with the other available genomes have shown that it is more similar to *Arabidopsis thaliana* in terms of whole genome content[1].

Another application of comparative genomics is to search for any suitable species, different from the pathogen, that can be modelled to study that disease. Model species should have a number of features depending on the type of study. In order to study a disease, the model organism should have conserved immunologic responses as well as the homologous genes, involved in pathogenicity. Comparative studies can prove or disapprove any organism to be a model. For example, *P. yoelii*, the rodent malaria parasite, has long been used as model for malaria research as it shows similarity in many biological responses to that of the human malaria parasite *P. falciparum*[27]. Comparative analysis has shown that a number of features are common between the two species. Synteny is highly conserved between the two species in region of house keeping genes but low in the genes involved in antigenic variation[27].

The *Yir* gene family is the largest family of genes in *P. yoelii yoelii* genome having 838 *Yir* genes in totality. The *Yir* gene family is homologous to the *Yir* multi gene family in *P. vivax* which encodes proteins showing immunovariation in natural infection[27]. The *Yir* family in *P. yoelii yoelii* makes it a good model system for studying antigenic variation in *P. vivax*.

Diagnostics: Accurate diagnosis is the key step to cure any disease effectively. Apart from the characteristic symptoms of malaria, the correct identification of the causing species is required to confirm the disease and to begin the treatment. Now a days a number of diagnostic approaches are applied of which PCR based approaches have special attentions. The high accuracy of the technique, giving results in very short time, has made it a versatile diagnostic approach particularly the drug resistant cases. The *P. falciparum* genome is undoubtedly going to provide the correct and specific target sequences which can be identified through PCR from the blood samples.

Conclusion

The availability of the genome of *P. falciparum* will provide numerous opportunities for research on this organism. The information which is generated from the sequencing of the genome is extremely useful in developing effective drugs and vaccines to combat this disease. No doubt that the malaria genome in conjunction with the genome information on its host and vector will energise the investigators to develop better and newer control strategies.

References

1. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002; *419* : 498–511.

2. Breman JG. The ears of the hippopotamus: Manifestation, determinants, and estimates of the malaria burden. *Amer J Trop Med Hyg* 2001; *64* : 1–11.

3. Russell F Doolittle. The parasite genome: The grant assault. *Nature* 2002; *419* : 493–4.

4. Hoffman SL, Bancroft WH, Gottlieb M, James SL, Burroughs EC, Stephenson JR, Morgan MJ. Funding for malaria genome sequencing. *Nature* 1997; 387 : 647.

5. Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, Williamson DH. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* 1996; *261* : 155–72.

6. Zuegge J, Ralph S, Sehmuker MC, Fadden GI, Schneider G. Deciphering apicoplast targeting signals-feature extraction from nuclear encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* 2001; *280* : 19–26.

7. Gardner MJ, Bates PA, Ling IT, Moore DJ, Mc Cready S. Gunasekera MBR, Wilson RJM, Williamson DH. Mitochondrial DNA of the human malaria parasite *Plasmodium falciparum*. *Mol Biochem Parasitol* 1998; *31* : 11–8.

8. Joseph JT, Aldritt SM, Unnasch T, Puijalon O, Wirth DF. Characterization of a conserved extrachromosomal element isolated from the avian malarial parasite *Plasmodium gallinaceum*. *Mol Cell Biol* 1989; *9* : 3621–9.

9. Feagin JE. The extrachromosomal DNAs of apicomplexan parasites. *Ann Rev Microbiol* 1994; *48* : 81–104.

10. Viadya AB, Akella R, Suplick K. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase pair DNA of a malarial parasite. *Mol Biochem Parasitol* 1989; *35* : 97–107.

11. Sharma I, Rawat DS, Pasha ST, Biswas S, Sharma YD. Complete nucleotide sequence of the 6 kb element and conserved cytochrome b gene sequences among Indian isolates of *Plasmodium falciparum*. *Intl J Parasitol* 2001; *31* : 1107–13.

12. Aldritt SM, Joseph JT, Wirth DF. Sequence identification of cytochrome b and *Plasmodium gallinaceum*. *Mol Cell Biol* 1989; *9* : 3614–20.

13. Suplick K, Akella R, Saul A, Vaidya AB. Molecular cloning and partial sequence of a 56 kb repetitive DNA from *P. falciparum*. *Mol Biochem Parasitol* 1988; *30* : 289–90.

14. Sharma I, Pasha ST, Sharma YD. Complete nucleotide sequence of the *Plasmodium vivax* 6 kb element. *Mol Biochem Parasitol* 1998; *47* : 259–63.

15. Sharma I, Sharma YD. Malarial mitochondrial genome: The 6 kb element. *Indian J Malariol* 2001; *38* : 45–60.

16. Washburn MP, Walters D, Yates JR. Large-scale analysis of the yeast protease by multidimensional protein identification technology. *Nature Biotechnol* 2001; *19* : 242–7.

17. Gardner MJ, Shallom SJ, Carlton JM, Salzberg SL, Nene V, Shoaibi A, Ciecko A, Lynn J, Rizzo M, Weaver B, Jarrahi B, Brenner M, Parvizi B, Tallon L, Moazzez A, Granger D, Fujii C, Hansen C, Pederson J, Feldblyum T, Peterson J, Suh B, Angiuoli S, Pertea M, Allen J, Selengut J, White O, Cummings LM, Smith HO, Adams MD, Venter JC, Carucci DJ, Hoffman SL, Fraser CM. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* 2002; *419* : 531–4.

18. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, Witney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinden RE, Yates JR, Carucci DJ. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 2002; *419* : 520–6.

19. Rohdich F *et al*. Biosynthesis of terpenoides. 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ISP F) from *Plasmodium falciparum*. *European J Biochem* 2001; *268* : 3190–7.

20. Kemp LE, Bond CS, Hunter WN. Structure of 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase: An essential enzymes for isoprenoid biosynthesis and target for antimicrobial drug development. *Proc Natl Acad Sci* USA 2002; *99* : 65–96.

21. Dieckmann A, Jung A. Mechanisms of sulphadoxine resistance in *Plasmodium falciparum*. *Mol Biochem Parasitol* 1986; *19* : 143–7.

22. McConkey GA. Targeting the shikimate pathway in apicomplexan parasites. *Nature* 1998; *393* : 801–5.

23. Roberts CW, Roberts F, Lyons RE, Kirisits MJ, Mui EJ, Finnerty J, Johnson JJ, Ferguson DJ, Coggins JR, Krell T, Coombs GH, Milhous WK, Kyle DE, Tzipori S, Barnwell J, Dame JB, Carlton J, McLeod R. The shikimate pathway and its brances in apicomplexan parasite. *J Infect Dis* 2002; *185* (Suppl.): 525–36.

24. Dyer M, Wong IH, Jackson M, Huynh P, Mikkelsen R. Isolation and sequence analysis of a cDNA encoding an adenine nucleotide translocator from *Plasmodium falciparum*. *Biochim Biophys Acta* 1994; *1186* : 133–6.

25. Kyes SA, Horrocks P, Newbold CI. Antigenic variation at the infected red cell surface in malaria. *Ann Rev Microbiol* 2001; *55* : 673–707.

26. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Sedegah M, Shoaibi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii. Nature* 2002; *419* : 512–9.

27. Hall N, Pain A, Berriman M, Churcher C, Harris B, Harris D, Mungall K, Bowman S, Atkin R, Baker S, Barron A, Brooks K, Buckee CO, Burrows C, Cherevach I, Chillingworth C, Chillingworth T, Christodoulou Z, Clark L, Clark

R, Corton C, Cronin A, Davies R, Davis P, Dear P, Dearden F, Doggett J, Feltwell T, Goble A, Goodhead I, Gwilliam R, Hamlin N, Hance Z, Harper D, Hauser H, Hornsby T, Holroyd S, Horrocks P, Humphray S, Jagels K, James KD, Johnson D, Kerhornou A, Knights A, Konfortov B, Kyes S, Larke N, Lawson D, Lennard N, Line A, Maddison M, McLean J, Mooney P, Moule S, Murphy L, Oliver K, Ormond D, Price C, Quail MA, Rabbinowitsch E, Rajandream MA, Rutter S, Rutherford KM, Sanders M, Simmonds M, Seeger K, Sharp S, Smith R, Squares R, Squares S, Stevens K, Taylor K, Tivey A, Unwin L, Whitehead S, Woodward J, Sulston JE, Craig A, Newbold C, Barrell BG. Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature* 2002; *419* : 527–31.