Genomic Characterization of Chromosome 1 of Plasmodium falciparum by Computational Methods

C Kumar, C Anuradha, K Venkateswara Swamy

Citation

C Kumar, C Anuradha, K Venkateswara Swamy. *Genomic Characterization of Chromosome 1 of Plasmodium falciparum by Computational Methods*. The Internet Journal of Microbiology. 2004 Volume 1 Number 2.

Abstract

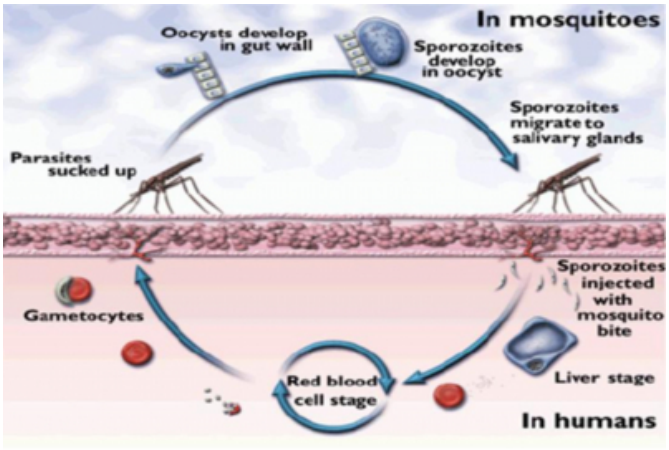
Malaria inflicts serious health and economic burdens on many countries throughout the world. The continued emergence of drug-resistant parasites, particularly in Plasmodium falciparum, underscores the need for new therapies and consequently the identification of novel targets for drug development. Increasing worldwide resistance of P. falciparum to traditional chemotherapy strategies such as chloroquine and mefloquine demonstrates the urgent need for the discovery of novel chemotherapeutic agents in the fight against malaria. While the stages of parasitic infection are well documented, until recently, little has been known concerning the regulation of the parasitic life cycle. However, the identity of family of enzymes such as kinases, etc., in P. falciparum with a high degree of sequence conservation to the mammalian targets has allowed researchers to begin investigation of the mechanisms that control passage through the parasitic life cycle. In present study we have conducted gene ontology studies on chromosome 1 of P. falciparum which is source of potential drug targets to fight against malaria to create new drugs or vaccines.

INTRODUCTION

In spite of early progress, malaria is still one of the most serious health problems facing humanity. It affects 300-500 million people causing over 2.5 million deaths annually, mostly in children (WHO Report, 1997). Malaria also poses a particular danger to pregnant woman and may lead to miscarriage or low birth weight of the child. In endemic regions, the disease is recognized as serous impediment to economic and social development (Bowman, et al .1999 and Ohashi et al., 2002). It is well know that malaria is caused by protozoan parasites of the genus Plasmodium. In humans, four species are responsible for malaria: P. falciparum, P. vivax, P. ovale, and P. malariae. The first one is the most dangerous. Two aspects have currently stimulated new efforts regarding medicinal and molecular studies about malaria: the rapid emergence of P. falciparum strains resistant to currently available antimalarial drugs (Newton and White, 1999) and the inefficacy of malarial vaccines (Soares and Rodrigues, 1998). The life cycle of Plasmodium is very complex (Fig.1). While stages of parasitic infection are well documented, until recently, little has been known concerning the regulation of the parasitic life cycle. Infections are usually transmitted by the bites of infected female Anopheles mosquitoes. The sporozoites are the infective stage, which migrate to the salivary glands of the mosquito where they are injected into the blood of the next host when the female mosquito bites. Because of the complex life cycle of these parasites it is difficult to develop a vaccine.

Figure 1

Figure 1: Life cycle of in mosquitoes and humans.



The complete genome sequence of P. falciparum is available at genome database, which can be downloaded through Internet. This information also help us to understand the biology of P. falciparum: how it invades the liver and the red cells, how it survives inside them and how it makes its way out. Knowledge of the pathogenesis of the disease is

improving. Recrudescing variants in chronic malaria infections are antigenically distinct from those of the parental parasite, indicating that P. falciparum has devised means to vary polypeptides exported to the infected erythrocyte surface, thus reflecting a fundamental element of parasitism: antigenic variation. We now know that this variation is brought about by a family of highly changeable adhesive polypeptides, such as P. falciparum erythrocyte membrane protein, which is localised at the infected erythrocyte surface, and the RIFIINS which were recently identified with the help of the genome sequence project. Yet others remain to be elucidate, including clag (cytoadherence-linked asexual gene) and CTP (conserved telomeric protein).

CHROMOSOME STRUCTURE

P. falciparum chromosomes vary considerably in length, with most of the variation occurring in the subtelomeric regions. Field isolates, even those from individuals residing in a single village, exhibit extensive size polymorphism that is thought to be due to recombination events between different parasite clones during meiosis in the mosquito. Chromosome size variation is also observed in cultures of erythrocytic parasites, but is due to chromosome breakage and healing events and not to meiotic recombination. Subtelomeric deletions often extend well into the chromosome, and in some cases alter the cell adhesion properties of the parasite owing to the loss of the gene(s) encoding adhesion molecules. Because many genes involved in antigenic variation are located in the subtelomeric regions, an understanding of subtelomere structure and functional properties is essential for the elucidation of the mechanisms underlying the generation of antigenic diversity (Greenwood and Mutabingwa, 2002). Subtelomeric exchanges occur in other eukaryotes, but the regions involved are much smaller (2.5–3.0 kb) in S. cerevisiae. Previous studies of P. falciparum telomeres suggested that they contained six blocks of repetitive sequences that were designated telomere-associated repetitive elements. Whole genome analysis reveals a larger (up to 120 kb), more complex, subtelomeric repeat structure than was observed previously. The conserved regions fall into five large subtelomeric blocks. Centromeres have not been identified experimentally in malaria parasites. However, putative centromeres were identified by comparison of the sequences of chromosomes 2 and 3. Eleven of the 14 chromosomes contained a single region of 2–3 kb with extremely high (A/T) content (97%) and imperfect short tandem repeats, features resembling the regional S. pombe centromeres. Unlike many other eukaryotes, the malaria parasite genome does not contain long tandemly repeated arrays of ribosomal RNA (rRNA) genes. Instead, Plasmodium parasites contain several single 18S-5.8S-28S rRNA units distributed on different chromosomes (Duret, 2000).

GENOME STRUCTURE

The P. falciparum 3D7 nuclear genome is composed of 22.8 mega-bases (Mb) distributed among 14 chromosomes ranging in size from approximately 0.643 to 3.29 Mb. The overall (A/T) composition is 80.6%, and rises to 90% in introns and intergenic regions. The structures of protein-encoding genes were predicted using several gene-finding programs and manually curated. Approximately 5,300 protein-encoding genes were identified. This suggests an average gene density in P. falciparum of 1 gene per 4,338 base pairs (bp), slightly higher than was found previously with chromosomes 2 and 3 (1 per 4,500 bp and 1 per 4,800 bp, respectively). The higher gene density reported here is probably the result of improved gene-finding software and larger training sets that enabled the detection of genes overlooked previously 8 . Introns were predicted in 54% of P. falciparum genes, a proportion roughly similar to that in S. pombe and Dictyostelium discoideum, but much higher than observed in S. cerevisiae where only 5% of genes contain introns. Excluding introns, the mean length of P. falciparum genes was 2.3 kb, substantially larger than in the other organisms in which the average gene lengths range from 1.3 to 1.6 kb. P. falciparum genes showed a markedly greater proportion of genes (15.5%) longer than 4 kb compared to S. pombe and S. cerevisiae (3.0% and 3.6%, respectively). The explanation for the increased gene length in P. falciparum is not clear. Many of these large genes encode uncharacterized proteins that may be cytosolic proteins, as they do not possess recognizable signal peptides. No transposable elements or retrotransposons were identified (Hernandez, et al .1996).

Since the sequencing of the first two chromosomes of the malaria parasite, P. falciparum, there has been a concerted effort to sequence and assemble the entire genome of this organism here we report the sequence of chromosome 1 of P. falciparum 3D7. We describe the methods used to map, sequence and annotate this chromosome 1 in the present paper. During annotation, we assign Gene Ontology terms to the predicted gene products, and observe clustering of some malaria-specific terms to specific chromosomes. We identify a highly conserved sequence element found in the intergenic region of internal var genes that is not associated with their telomeric counterparts (Gardner, 1998). Berriman, et al.

(2001) have reported that Gene Ontology (GO) was used to classify genes across the entire genome and as GO had not been previously applied for annotating an intracellular parasite, new parasite-specific GO terms were created (Ashburner and Lobo, 2000). The proportion of genes associated with parasite-specific processes or localized in parasite-specific compartments varies between chromosomes. Whereas most 'housekeeping' genes appear to be evenly distributed across the chromosomes and chromosome 1 appears to have the lowest proportion of genes annotated with apicoplast localization. The potential of the (G/C) rich sequences to form DNA secondary structures supports a possible function as regulatory elements in var-related genetic processes in P. falciparum.

MATERIALS AND METHODS

Here we present the methodology used for analysis and characterization of chromosome 1 of Plasmodium falciparum 3D7.

NUCLEOTIDE SEQUENCE OF CHROMOSOME 1

The complete nucleotide sequence of Plasmodium falciparum 3D7 chromosome1 was downloaded from http://www.ncbi.nlm.nih.gov/. It has four sub regions with total region of 0-643 Kbp. These four sub regions are designated by Accession Numbers: AL_031747, AL_031745, AL_031746, AL_031744.

CHARACTERIZATION OF CHROMOSOME 1

DNA sequence of all sub regions were analyzed for the presence of open reading frame (ORF) by submitting the nucleic sequence details to ORF finder online tool available with National Center for Biotechnology Information (http://www.ncbi.nim.nih.gov/orf). The characterization was also made with Vector NTI suite-V.9 (http://www.informaxinc.com/solutions.vectornti) to translate the DNA sequence to protein sequence. Vector NTI is an integrated sequence analysis and data management software package, which allows molecular biologists to analyze, manipulate, construct, store and manage complex biological molecules. Annotation of genes were studied by manual curation of the output of the software which finds genes in microbial DNA, especially the genomes of bacteria and archaea (Salzberg, et al .1999).

CHARACTERIZATION OF PROTEIN TRANSLATED FROM CHROMOSOME 1

Pfam and Prosite tools available at http://www.expasy.org were used for characterizing protein sequences. Pfam (Bateman, et al .2004) is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. Prosite (Hulo et al., 2004) is a database of protein families and domains. It is based on the observation that, while there is a huge number of different proteins, most of them can be grouped, on the basis of similarities in their sequences, into a limited number of families. Proteins or protein domains belonging to a particular family generally share functional attributes and are derived from a common ancestor.

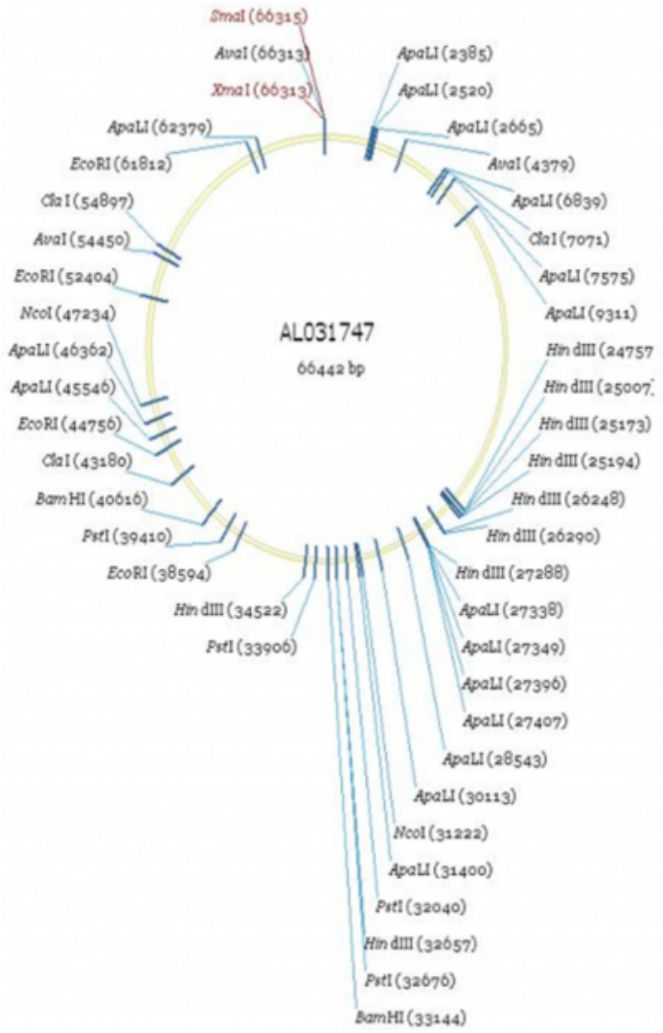
RESULTS AND DISCUSSION

Characterization of nucleotide sequences (AL031744, AL031745, AL031746 and AL031747) of chromosome1 of Plasmodium falciparum was done by Vector NTI 9.0 package and the results are presented in Fig. 2-5. The results of the complete sequence analysis of (AL031744, AL031745, AL031746 and AL031747) chromosome 1 after submitting to Vector NTI software package has been shown as: number of predicted genes, length of entire nucleotide region and name as well as number of restriction sites (Fig. 2-5).

In case of AL031747 nucleotide sequence, the number of predicted genes was 14, the number of restriction sites was approximately 44 and the length of the entire sequence was 66442 base pairs. Based on the restriction site analysis of AL031747 nucleotide sequence has revealed the occurrence of maximum HindIII restriction sites, which can be used in cloning technology (Fig. 2). Sequence analysis using Pfam has also revealed that this particular sequence codes for protein which belongs to the erythrocyte membrane protein PfEMP1 which has been identified as the rosetting ligand of the malaria parasite P. falciparum hence suggests that the predicted protein has glycosaminoglycan binding function.

Figure 2

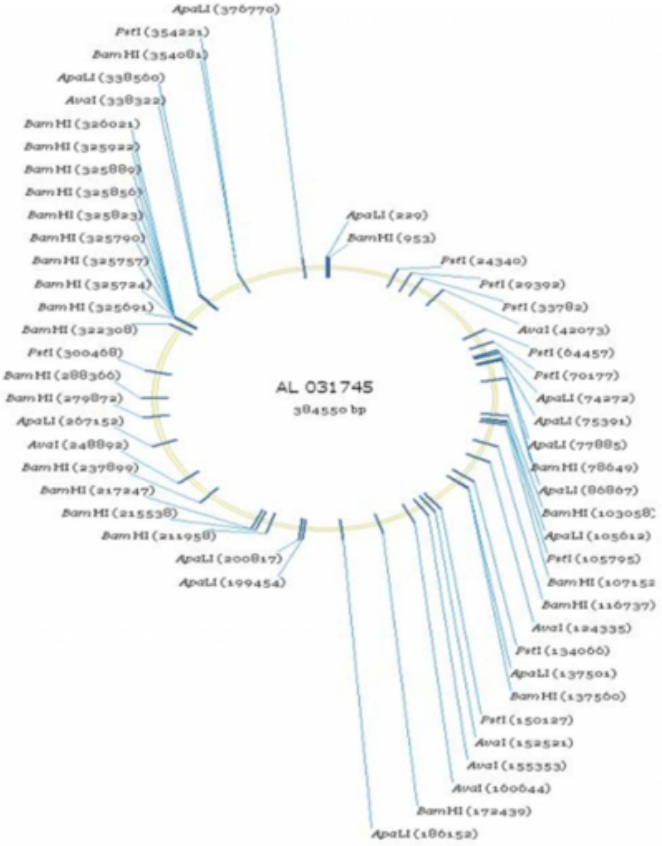
Figure 2: Genome analysis of AL031747 region of by Vector NTI 9.0 package.



In case of AL031745 nucleotide sequence, the number of predicted genes was 136, the number of restriction sites was approximately 54 and the length of the entire sequence was 384550 base pairs. Based on the restriction site analysis of AL031745 nucleotide sequence, it is possible to infer that this sequence has maximum of Bam HI restriction sites (Fig. 3). The Pfam result of predicted ORF from AL_031745 has revealed the presence of bromodomains that are 110 amino acid long domains which are found in many chromatin associated proteins. Bromodomains are found in a variety of mammalian, invertebrate and yeast DNA-binding proteins. In some proteins, the classical bromodomain has diverged to such an extent that parts of the region are either missing or contain an insertion (e.g., mammalian protein HRX, Caenorhabditis elegans hypothetical protein ZK783.4, yeast protein YTA7) (Chua et al., 2005).

Figure 3

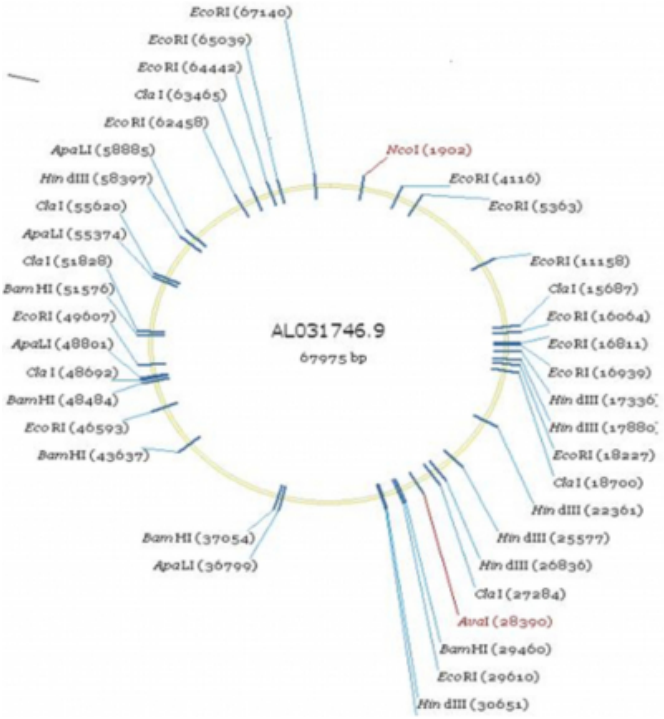
Figure 3: Genome analysis of AL031745 region of by Vector NTI 9.0 package.



Another case of AL031746 nucleotide sequence, the number of predicted genes was 17, the number of restriction sites was approximately 40 and the length of the entire sequence was 67975 base pairs. Based on the restriction site analysis of AL031746 nucleotide sequence, it is possible to infer that it has revealed the presence of maximum no of EcoR I restriction sites (Fig.4). Pfam results of the best predicted ORF from AL_031746 has revealed the presence of the ABC transporter transmembrane region of this family represents a unit of six transmembrane helices, which is predicted to function in ATP-binding.

Figure 4

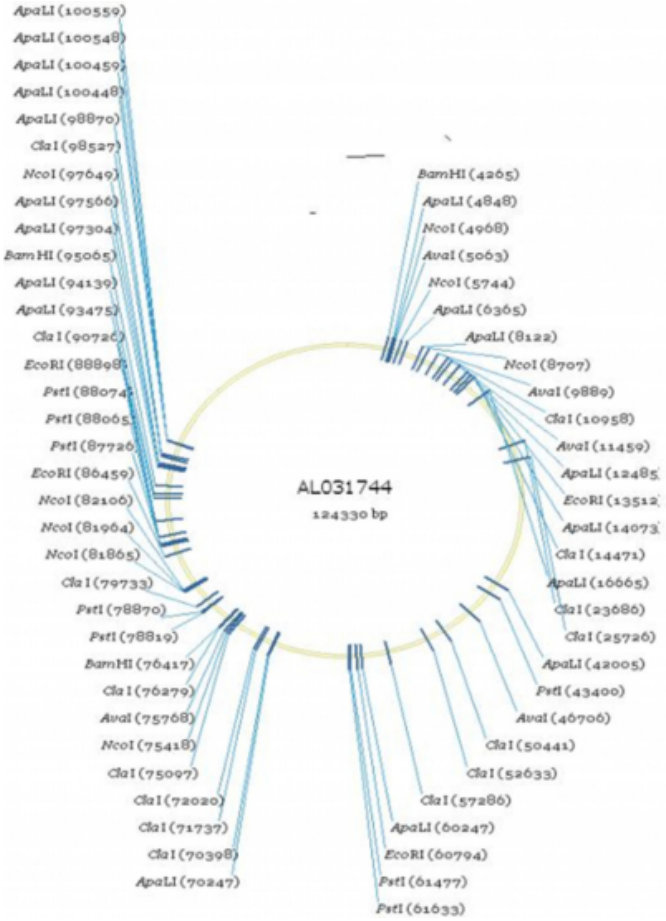
Figure 4: Genome analysis of AL031746 region of by Vector NTI 9.0 package.



Another case of AL031744 nucleotide sequence, the number of predicted genes was 33, the number of restriction sites was approximately 62and the length of the entire sequence was 124330 base pairs. The restriction site analysis of AL031744 nucleotide sequence has also revealed the presence of maximum of Hind III restriction sites (Fig.5). Best predicted ORF from AL_031744 has revealed that FF (Phenylalanine) domain has been predicted to be involved in protein-protein interaction. This domain was recently shown to bind the hyperphosphorylated C-terminal repeat domain of RNA polymerase II, confirming its role in protein-protein interactions. Moreover another predicted ORF from AL_031744 has revealed that Brevenin/esculentin/gaegurin/rugosin family contains a number of defense peptides secreted from the skin of amphibians, including the opiate-like dermorphins and deltorphins, and the antimicrobial dermoseptins and temporins whose function is well established in antimicrobial peptide activity (Chen et al., 2005).

Figure 5

Figure 5: Genome analysis of AL031744 region of by Vector NTI 9.0 package.



CONCLUSIONS

P. falciparum is the most virulent of the four species causing malaria and responsible for most malarial deaths. The particular virulence of P. falciparum is partly due to the ability of infected erythrocytes to adhere to a variety of host receptors and avoid spleenic clearance. The complete sequence analysis of chromosome 1 of P. falciparum has revealed numerous novel CDSs. This chromosome is predicted to contain approximately 200 genes of which most of them are known to be predicted as rifin, var, stervor genes and also a gene which encodes for rRNA units like 5.8s, 18s and 28srRNA. The predicted proteins were also known to belong Rifin/Stevor family, which are multi copy gene family of subtelomeric open reading frames and rif interspersed repetitive elements. Both families contain three predicted transmembrane segments. It has been proposed that stevor and rif are members of large super family that code for variant surface antigens.

The rif genes are know to encode proteins called RIFINs

exposed on the surface of infected erythrocytes. The functions of these proteins are not known and they have not been shown to mediate binding.

The var gene family in two exon units encodes PfEMP1 antigens. Exon I codes for the extracellular and variable part of the protein as well as a transmembrane region and Exon II encodes the intracellular and relatively conserved acidic terminal segment (ATS). The most variable part of the protein contains a N-terminal segment followed by segments composed of three domain types: Duffy binding-like domains (DBL-domains): Cysteine-rich inter-domain regions (CIDRs) and C2 (calcium binding domain).

The presence of the C2 domain, which is a Ca2+-dependent membrane-targeting module found in many cellular proteins involved in signal transduction or membrane trafficking, has revealed that M. falciparum is thought to be involved in calcium-dependent phospholipid binding and in membrane targeting processes such as subcellular localization. C2 domains are unique among membrane targeting domains in that they show wide range of lipid selectivity for the major components of cell membranes, including phosphatidylserine and phosphatidylcholine.

The presence of Duffy binding domain has revealed that this parasite invade human erythrocytes that express Duffy blood group surface determinants. Duffy is expressed on RBCs (in Duffy-positive individuals), endothelial cells of post capillary venules and Purkinje cells of the cerebellum. On RBCs, the Duffy antigen acts as a receptor for invasion by the malarial parasite; Duffy-negative individuals, whose RBCs do not express the receptor are resistant to such infection. The normal physiological function of Duffy remains unclear (Choe et al., 2005). It is believed to play a more important role on endothelial cells, since expression on these cell types is highly conserved, whereas the function on RBCs appears to be dispensable in order to confer resistance to malaria. The signaling pathways activated by Duffy are also unknown and the receptor has not been shown to act through a G protein. The variant surface antigen family Plasmodium falciparum erythrocyte membrane protein-1 (PfEMP1) is an important target for protective immunity and is implicated in the pathology of malaria through its ability to adhere to host endothelial receptors (Huber et al., 2005). Specific inhibitors can be designed further using drug design to prevent interaction of PfEMP1 with specific receptors on erythrocytes for effective inhibition of ‘rosetting' (clumping of parasitized RBCs involving PfEMP1 and receptors associated with severe malaria).

As the effects of traditional chemotherapy agents decrease as a result of growing parasitic resistance, the global threat of malaria remains a cause for concern. P. falciparum resistance to many chemotherapeutic agents such as fansidar and chloroquine has developed worldwide, creating dire social and health implications for developing countries (Sachs and Malaney, 2002). Drug resistance of this scale requires scientists to identify new mechanisms by which to inhibit P. falciparum and thus prevent the spread of malaria. Efforts are currently underway to further delineate the structural prerequisites for molecular recognition and tight binding inhibitors for key enzymes of parasite and thereby to guide the structure-based design of a new generation of antimalarial drugs for use with strains of P. falciparum resistant to traditional therapeutic agents. The genome sequence analysis of chromosome 1 of P. falciparum through the present study will provide the foundation for future research on this organism, and are exploited in the search for new drugs and vaccines to fight malaria.

ACKNOWLEDGEMENTS

CSK thanks to DST, New Delhi (DST-FIST program) and UGC-New Delhi for financial support to establish computer networking Lab.

References

r-0. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Ashburner Aikawa .M, Lobo CA (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25(1):25-29.
r-1. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. (2004) The Pfam Protein Families Database. Nucleic Acids Res.,32: D138-D141.
r-2. Berriman, M., Aslett, M. & Ivens, A. (2001). Parasites are GO. Trends Parasitol. 17, 463–464.
r-3. Bowman and Bauer H. (1999). The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum. Nature 400: 532-538.
r-4. Chen T, Li L, Zhou M, Rao P, Walker B, Shaw C. (2005). Amphibian skin peptides and their corresponding cDNAs from single lyophilized secretion samples: Identification of novel brevinins from three species of Chinese frogs.
Peptides. (In Press).
r-5. Choe H, Moore MJ, Owens CM, Wright PL, Vasilieva N, Li W, Singh AP, Shakri R, Chitnis CE, Farzan M. (2005). Sulphated tyrosines mediate association of chemokines and Plasmodium vivax Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). Mol Microbiol. 55: 1413-22.
r-6. Chua YL, Channeliere S, Mott E, Gray JC. (2005). The bromodomain protein GTE6 controls leaf development in Arabidopsis by histone acetylation at ASYMMETRIC

LEAVES1. Genes Dev. 19:2245-54.
r-7. Duret, L. (2000). tRNA gene number and codon usage in the C. elegans genome are coadapted for optimal translation of highly expressed genes. Trends Genet.16: 287–289.
r-8. Gardner, M. J. (1998). Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum. Science 282: 1126–1132.
r-9. Greenwood, B. & Mutabingwa, T. (2002). Malaria in 2002. Nature 415: 670–672.
r-10. Hernandez, R. R., Hinterberg, K. & Scherf, A. (1996). Compartmentalization of genes coding for immunodominant antigens to fragile chromosome ends leads to dispersed subtelomeric gene families and rapid gene evolution in Plasmodium falciparum. Mol. Biochem. Parasitol. 78: 137–148.
r-11. Huber SM, Duranton C, Lang F.Patch-clamp analysis of the “new permeability pathways” in malaria-infected erythrocytes. (2005). Int Rev Cytol. 246:59-134
r-12. Hulo N., Sigrist C.J.A., Le Saux V., Langendijk-Genevaux P.S., Bordoli L., Gattiker A., De Castro E., Bucher P., Bairoch A. (2004) Recent improvements to the PROSITE database. Nucleic Acids. Res. 32:D134-D137.
r-13. Newton, P and White, N. (1999). Malaria: new developments in treatment and prevention. Ann. Rev. Med. 50:179-192.
r-14. Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Looareesuwan S, Tokunaga K. Lack of association between interleukin-10 gene promoter polymorphism, -1082G/A, and severe malaria in Thailand. (2002). Southeast Asian J Trop Med Public Health. 33 Suppl 3, 5-7.
r-15. Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria, Nature 415: 680–685.
r-16. Salzberg, S. L., Pertea, M., Delcher, A. L., Gardner, M. J. & Tettelin, H. (1999). Interpolated Markov models for eukaryotic gene finding. Genomics 59: 24–31.
r-17. Soares, IS and Rodrigues, MM. (1998). Malaria vaccine: road-blocks and possible solutions. Braz. J. Med. Biol. Res. 31:317-332.
r-18. World Health Organization Report, WHO Publications, Geneva, 1997.

Author Information

Chitta Suresh Kumar, M.D.
Associate Professor, Bioinformatics Center, Department of Biochemistry, Sri Krishnadevaraya University

C. M. Anuradha, M.Sc., M.Phil.
Bioinformatics Center, Department of Biochemistry, Sri Krishnadevaraya University

K. Venkateswara Swamy, M.Sc.
Bioinformatics Center, Department of Biochemistry, Sri Krishnadevaraya University