Malaria is one of the most serious infectious diseases in the world, inflicting acute illness on more than 300 million people and leading to at least one million deaths annually. Of the four Plasmodium species that use human as their vertebrate host, Plasmodium falciparum is the most lethal and P. vivax causes severe malaria with less lethality but high incidence of relapse. Thus, the pathogenesis and disease severity are quite different in malaria infection for both the cases. The malaria parasites are apicomplexan protozoan that contain a multi-membranous plastid-like organelle termed as ‘apicoplast’.

In silico characterization of genetic homology in nuclear-encoded apicoplast-targeted genes between Plasmodium falciparum & P. vivax

Sujata Mohanty, Shilpa Saklani & Mukesh Mahajan

Department of Biotechnology, Jaypee Institute of Information Technology University, NOIDA, India

Background & objectives: Resistance to anti-malarial drugs by the parasites is one of the major obstacles to malaria control. The primary objective of this work was to find specific nuclear-encoded-apicoplast-targeted genes that are conserved between two different human malaria parasite species, Plasmodium falciparum and P. vivax to find a common drug/vaccine targets for both the species.

Methods: Using computational genomics, possible nuclear-encoded-apicoplast-targeted genes were identified in P. falciparum genome. With comparative genomic approaches, homologous genes were identified between the two different human malaria species, P. falciparum and P. vivax.

Results: Of the total 545 reported nuclear-encoded-apicoplast-targeted genes in P. falciparum, we could narrow down to as less as five genes that were found to have highly conserved nucleotide stretches in P. vivax. However, two such genes were of importance, as the majority of the protein coding regions (exons) of these genes were found to be highly conserved between them.

Interpretation & conclusion: This preliminary study shows that nuclear-encoded-apicoplast-targeted genes were conserved between the two human malaria parasites and these could be targeted for developing a common drug to cure both forms of malaria.

Key words: Apicoplast - comparative genomics - malaria - Plasmodium falciparum - Plasmodium vivax

Malaria is one of the most serious infectious diseases in the world, inflicting acute illness on more than 300 million people and leading to at least one million deaths annually. Of the four Plasmodium species that use human as their vertebrate host, Plasmodium falciparum is the most lethal and P. vivax causes severe malaria with less lethality but high incidence of relapse. Thus, the pathogenesis and disease severity are quite different in malaria infection for both the cases. The malaria parasites are apicomplexan protozoan that contain a multi-membranous plastid-like organelle termed as ‘apicoplast’. Although the function of the apicoplast was highly debated after its discovery, later experimental evidences proved that this organelle performs varieties of important metabolic activities that are necessary for the survival of the parasites. The size of the apicomplexan plastid genome is extremely small (35 kilobase) with a very limited number of genes encoding 30 proteins serving many essential functions in P. falciparum. Thus the significance of apicoplast lies in its indispensable nature to the survival of parasite. With the availability
of the whole genome sequences of *P. falciparum* and readily available unpublished genome sequences for *P. vivax* in the public domain, new hopes have been generated for development of suitable and effective disease control measures targeting apicoplast.

Comparative genomics helps in understanding how genomes have evolved in time and moreover, in determining functions of genes, regulatory networks, and non-coding areas of genomes. Characterization of genes that are conserved across different taxa, comparison of intron-exon length, total gene length, etc., help in understanding how genes are affected by different evolutionary forces. Comparative genomics also allows inferences to be drawn about the coding potential of related genomes and the evolutionary forces that have influenced genome organization. It is thus anticipated that comparative genomics between these *Plasmodium* species can prove useful in unraveling functional features of *Plasmodium* species-specific genes, thereby providing information that will advance the mechanism of host-parasite interactions and malaria pathogenesis.

We therefore performed a comparative genomics study using the published genome sequence information of *P. falciparum* for the 545 nuclear-encoded apicoplast targeted genes and looked for their genetic homology with *P. vivax*. We also looked for the DNA sequence homology between the introns and exons for each homologous gene which are of importance not only to find out the similarity/difference between the two species but also to use the information for development of new drugs/vaccines for malaria control.

**Material & Methods**

Earlier work using bioinformatics approaches to identify the apicoplast-related genes discovered 17,679 genes, and of these only 545 high confidence apicoplast proteins were retrieved. In this study, the sequence information (kindly provided by Dr Geoff McFadden, Melbourne University) of these 545 genes was used with the help of the PlasmoDB web database (www.plasmodb.org) version 4.4 (Date of accession April, 2007). All these nuclear-encoded-apicoplast-targeted gene sequences were re-screened to detect complete sequence information. Many genes in the database were with incomplete information, which were not considered in further study. After such re-screening, only 420 total nuclear-encoded-apicoplast-targeted sequences of *P. falciparum* could be considered for the comparative genomics study with *P. vivax*. Sequence homology of these apicoplast targeted genes of *P. falciparum* was then searched in the *P. vivax* database using BLAST program as indicated in the PlasmoDB website. Homologous sequences having 50 per cent or more identity in both the species were considered for further analyses. For homology identifications, the sequences were aligned using CLUSTAL X algorithm to find conserved regions in both the coding (exons) and non-coding (introns) regions of each individual genes. A minimum of eight consecutive conserved nuclear bases between these two species were considered as a ‘homologous stretch’. Distribution of these stretches was further characterized based on their position and functional status (coding vs. non-coding) in each of these genes. The detail procedure adopted in the whole study including the method followed is represented in a flowchart (Fig. 1).

---

**Fig. 1.** Description of detail step-by-step method followed in the study.
Table. Details of *Plasmodium falciparum* genes, and corresponding homologs of *P. vivax* showing maximum conserved stretches, lengths, per cent similarities, molecular weight and reported functions

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene_ID (P. falciparum)</th>
<th>Chromosome</th>
<th>Gene_ID (P. vivax)</th>
<th>No. of stretch</th>
<th>Max. length of nucleotides</th>
<th>Per cent similarity in exons</th>
<th>Molecular weight* (Function)</th>
<th>Encoded proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>PF14_0664</td>
<td>14</td>
<td>Pv_6838.phat_204</td>
<td>100</td>
<td>20</td>
<td>59.0</td>
<td>391.0 KDa, (Fatty acid synthesis)</td>
<td>Biotin carboxylase subunit of acetyl CoA carboxylase, putative</td>
</tr>
<tr>
<td>2.</td>
<td>PF10_0149</td>
<td>10</td>
<td>Pv_6865.phat_768</td>
<td>103</td>
<td>26</td>
<td>0.0</td>
<td>69.4 KDa, (Protein biosynthesis)</td>
<td>Cysteine – tRNA ligase, putative</td>
</tr>
<tr>
<td>3.</td>
<td>PF10_0369</td>
<td>10</td>
<td>Pv_6571.phat_95</td>
<td>184</td>
<td>30</td>
<td>0.0</td>
<td>102.9 KDa, (Nucleotide excision repair)</td>
<td>Helicase, putative</td>
</tr>
<tr>
<td>4.</td>
<td>PF11_0493</td>
<td>11</td>
<td>Pv_6835.phat_168</td>
<td>120</td>
<td>277</td>
<td>0.0</td>
<td>10.3 KDa, (Not reported)</td>
<td>DNA repair helicase, putative</td>
</tr>
<tr>
<td>5.</td>
<td>PF07_0115</td>
<td>7</td>
<td>Pv_6843.phat_54</td>
<td>68</td>
<td>29</td>
<td>88.2</td>
<td>225.0 KDa, (Transporter proteins)</td>
<td>Cation transporting ATPase, cation transporter</td>
</tr>
</tbody>
</table>

*Source: www.plasmodb.org*
Results & Discussion

Although whole nuclear genome of the malaria parasite *P. falciparum* has been fully sequenced\(^1\), the apicoplast genome and the nuclear genome are being constantly annotated with addition of new information. Consequently, the web databases are being continuously changed and updated with newer versions. In the present study characterization and understanding of such genes were undertaken in a better perspective and also to find evolutionarily conserved regions in the recently-sequenced *P. vivax* genome. We started with 545 different apicoplast-related gene sequences in *P. falciparum* genome. However, our search was narrowed down to only 420 nuclear-encoded-apicoplast targeted genes to be considered in the present study. It was evident that the nuclear-encoded, apicoplast-targeted genes were present all over the genome and found in all the 14 chromosomes except in the second chromosome (Fig. 2). However, the distribution of these genes was quite uneven. A minimum of nine genes were present on the first chromosome and a maximum of 70 genes in the 14\(^{th}\) chromosome. It seemed that the distribution of these genes was roughly directly proportional to the length of the chromosome, as the first chromosome was the smallest and the 14\(^{th}\) chromosome the longest chromosome in the *P. falciparum* genome.

Of the 420 genes considered to be nuclear-encoded-apicoplast-targeted in *P. falciparum*, only 386 genes showed some kind of homology (to our criteria, at least 8 consecutive nucleotides should be conserved between the two species and at least 50 per cent homology for a particular gene, see above). However, the per cent homology of different genes between *P. falciparum* and *P. vivax* was different. As many as 171 genes fell under the class of 51-60 per cent homology between the two species and only 17 under the category of 81-95 per cent similarities. Of the 386 genes scanned and compared between these two species, five genes located in different chromosomes were found to have maximum homology stretches (Table). The molecular weight and function of these genes in *P. falciparum* and the encoded proteins both in *P. falciparum* and *P. vivax* are presented in the Table as indicated in the PlasmoDB website. Two of these genes in *P. falciparum* (gene IDs PF14_0664 and PF07_0115) need special mention, since as many as 59 and 88 per cent of exons (coding region), respectively, were found to be conserved between the two species. However, the homology stretches were not that long in comparison to the other three genes. It is interesting to note that the three *P. falciparum* genes (gene IDs PF10_0149, PF10_0369 and PF11_0493) had homology only in introns with *P. vivax*. Further, two of these three genes present in the 10\(^{th}\) chromosome of *P. falciparum* (PF10_0149, PF10_0369) had more number of homologous stretches but with less number of consecutive nucleotides (Fig. 3), whereas, the gene (PF11_0493) present in the 11\(^{th}\) chromosome had...
both more number of homologous stretches, and also the maximum number of consecutive nucleotides (a maximum of 277 nucleotide bases) in a particular stretch (Fig. 3).

The present results are interesting in many aspects. First of all, with computational genomic approach we could narrow down to a few candidate nuclear-encoded-apicoplast-targeted-genes in the most devastating malaria parasite, P. falciparum. Considering the apicoplast as the most important organelle in the parasite as several important biochemical pathways are being mediated for the parasite’s survival\textsuperscript{13}, genes responsible for the apicoplast could be better drug/vaccine targets\textsuperscript{14,15}. Due to scant work in apicoplast genomics, the candidate genes determined in this study would be worth analyzing in detail in the future. Secondly, since P. vivax malaria is very much prevalent in India, and genome sequences of P. vivax are still fully un-annotated and unpublished, the best way to characterize candidate genes in P. vivax is to find homologous genes using fully-annotated genome information of P. falciparum. Of the five candidate genes we found here, two genes (gene IDs PF14\_0664 and PF07\_0115) showed maximum homology in the coding regions. Although the exact functions of these genes are not known, these all seem to serve important functions in the parasites\textsuperscript{1}. Thus these two genes appear to be the best suitable genes which could be targeted for a common drug/vaccine development for both the Plasmodium species, common in India. Further work is needed employing modern biological techniques before any valid conclusions are made.

**Acknowledgment**

Authors thank Dr Geoffrey I. McFadden, Plant Cell Biology Research Centre, University of Melbourne, Australia, for providing the apicoplast genome sequence information.

**References**


Reprint requests: Dr Sujata Mohanty, Senior Lecturer, Department of Biotechnology, Jaypee Institute of Information Technology University A-10, Sector-62, NOIDA 201 301, India e-mail: sujatam2005@yahoo.com; sujata.mohanty@jiit.ac.in