Acta Tropica 122 (2012) 138–149

Contents lists available at SciVerse ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

Plasmodium vivax apicoplast genome: A comparative analysis of major genes from Indian field isolates☆

Vishal Saxena[a,*], Shilpi Garg[a], Jyotsna Tripathi[a], Sonal Sharma[a], Deepak Pakalapati[a], Amit K. Subudhi[a], P.A. Boopathi[a], Gagandeep S. Saggu[a], Sanjay K. Kochar[b], Dhanpat K. Kochar[c], Ashis Das[a]

[a] Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan, India
[b] Department of Medicine, S.P. Medical College, Bikaner, Rajasthan, India
[c] Department of Neurology, Kothari Medical and Research Institute, Bikaner, Rajasthan, India

ARTICLE INFO

Article history:
Received 24 October 2010
Received in revised form
24 December 2011
Accepted 2 January 2012
Available online 10 January 2012

Keywords:
Plasmodium vivax
Apicoplast
Plastid DNA
Codon usage
Phylogeny

ABSTRACT

The apicomplexan parasite *Plasmodium vivax* is responsible for causing more than 70% of human malaria cases in Central and South America, Southeastern Asia and the Indian subcontinent. The rising severity of the disease and the increasing incidences of resistance shown by this parasite towards usual therapeutic regimens have necessitated investigation of putative novel drug targets to combat this disease. The apicoplast, an organelle of procaryotic origin, and its circular genome carrying genes of possible functional importance, are being looked upon as potential drug targets. The genes on this circular genome are believed to be highly conserved among all *Plasmodium* species. Till date, the plastid genome of *P. falciparum*, *P. berghei* and *P. chabaudi* have been detailed while partial sequences of some genes from other parasites including *P. vivax* have been studied for identifying evolutionary positions of these parasites. The functional aspects and significance of most of these genes are still hypothetical. In one of our previous reports, we have detailed the complete sequence, as well as structural and functional characteristics of the Elongation factor encoding *tufA* gene from the plastid genome of *P. vivax*. We present here the sequences of large and small subunit rRNA (*lsu* and *ssu rRNA*) genes, *sufB* (ORF470) gene, RNA polymerase (*rpo B, C*) subunit genes and *clpC* (casienolytic protease) gene from the plastid genome of *P. vivax*. A comparative analysis of these genes between *P. vivax* and *P. falciparum* reveals approximately 5–16% differences. A codon usage analysis of major plastid genes has shown a high frequency of codons rich in A/T at any or all of the three positions in all the species. TTA, AAT, AAA, TAT, and ATA are the major preferred codons. The sequences, functional domains and structural analysis of respective proteins do not show any variations in the active sites. A comparative analysis of these Indian *P. vivax* plastid genome encoded genes has also been done to understand the evolutionary position of the Indian parasite in comparison to other *Plasmodium* species.

1. Introduction

Malaria caused by the *Plasmodium* species, poses major health concerns in many regions of the world. Reports on severe manifestations caused by *Plasmodium vivax* (Kochar et al., 2005, 2007, 2009; Alexandre et al., 2010), one of the major human malaria parasites, has highlighted the need to further study and identify novel and more effective therapeutic drug targets against this less studied parasite. Targets for drug design are conventionally nuclear genome encoded products. The apicomplexans, are known to harbour two additional extrachromosomal genomes, viz. mitochondrial and plastid, carrying genes, which encode proteins with potentially crucial functions. The mitochondrial DNA in *P. vivax* is a 6 kb linear molecule (Sharma et al., 1998) also conserved in *P. falciparum*, another major human malaria parasite. The plastid genome of *P. vivax* has not been investigated and except reports on partial sequences of a few select genes no information is available in literature. In one of our previous reports we had detailed the comparative genomics and functional aspects of Elongation factor Tu A (*tufA*) gene and protein, respectively, from the apicoplast genome of *P. vivax* (Saxena et al., 2007).

The plastid genome in *Plasmodium* species is circular and appears to be similar to the chloroplast DNA except for the missing photosynthetic genes. It is housed in a 3–4 membrane

☆ Nucleotide sequence data reported in this paper are available in the GenBank database under the accession number(s): DQ503575, DQ499659, HQ110105, HQ166839, and HQ398241–HQ398248.

* Corresponding author. Tel.: +91 9414069231; fax: +91 1596 244183.
E-mail addresses: vishalsaxena@bits-pilani.ac.in, vishalsaxena12@gmail.com (V. Saxena).

0001-706X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.actatropica.2012.01.007

layered organelle called apicoplast (Wilson et al., 1996; Kohler et al., 1997) whose origin has been debated for long. It is hypothesized that a cyanobacterium – like procaryotic organism was engulfed by a phagotropic eukaryote which was further engulfed by photoautotrophic protists and algal groups. The cyanobacterium-like organism in its primary host was not destroyed but was retained as an organelle with double membranes. This hypothesis is well supported by the fact that these membranes resemble the two membranes of Gram-negative ancestral cyanobacterium-like endosymbiont. The presence of two additional membranes outside the above two in apicoplast, outermost being similar to phagosomal membrane of secondary host cell and inner layer resembling plasma membrane of engulfed algal cells indicates to secondary endosymbiotic origin of apicoplast (Lim and McFadden, 2010). The secondary host cells were later termed as chimaeric 'meta-algal' cells (includes groups like heterokonts, cryptophytes, chlorarachniophytes and euglenoids) and apicomplexans are also believed to have evolved through similar evolutionary incidences (Wilson, 2005; McFadden and Waller, 1997; Foth and McFadden, 2003). The sequence analysis of apicoplast genomes from various apicomplexans has also shown their resemblance to the red algal plastid genomes. The genome of plastid was retained to some extent while many genes were transferred to the host nuclear genome during evolutionary stages. Thus the procaryotic origin of the apicoplast makes the organelle and its protein products putative drug targets (Fichera and Roos, 1997; McFadden and Roos, 1999; Ralph et al., 2001).

The circular plastid genome has been fully characterized as a 35 kb DNA molecule from *Plasmodium falciparum* (Wilson et al., 1996), 27 kb from *P. chabaudi* and 31 kb from *P. berghei* (Sato et al., 2000). It carries genes of functional importance and their protein products are believed to function within the organelle. Till date only *tufA* has been shown to be translationally active within the *P. falciparum* organelle (Chaubey et al., 2005; Biswas et al., 2011). In *P. falciparum*, two copies each of small subunit ribosomal RNA (*ssu rRNA)* and large subunit ribosomal RNA (*lsu rRNA)* genes are present in the approximately 10.5 kb long inverted repeat (IR) region of the plastid genome termed as IRA and IRB. These genes are flanked on both sides by a duplicate set of transfer RNA (*tRNA*) genes for 9 amino acid anticodons and are believed to be conserved throughout apicomplexans (Wilson et al., 1996). The IRA region is followed by the *orf470* gene (an orthologue of *E. coli sufB* gene and *ycf24* gene from red algae) which is believed to encode SufB protein, which works as a complex with other members of the Suf family, viz., SufC and SufD. The genes for both are present in nuclear genome of malaria parasite. The SufBCD complex acts in the *Suf* system helping in maintenance and regulation of the iron–sulphur [Fe–S] protein cluster. The latter (SufD) in turn, is involved in electron transfer as well as in redox and non-redox catalysis (Ellis et al., 2001; Wilson, 2005). A set of RNA polymerase (*rpo*) subunit genes viz. *rpo B*, and *C* are also present on *P. falciparum* plastid genome. These genes have been described as indicators of the plastid origin of the circular genome. The genes correspond to the β and β′ subunits of *E. coli* RNA polymerase. The complete sequence of *rpoC* is split into *rpoC1* and *rpoC2* as in other plastid and cyanobacterial genomes and lacks the intron typical of higher plants (Gardner et al., 1991, 1994). Similar to other plastid DNAs, a putative ribosomal protein (*rp*) gene, *rps2*, maps downstream to the RNA polymerase subunit genes. The IRB region is followed by *tRNA* genes for 11 amino acids anticodons completing the set of all 20 amino acid anticodon genes. Downstream to these genes are spread 17 ribosomal protein (*rpl* and *rps*) genes, which is followed mainly by *tufA* and *clpC* genes. The *clpC* gene is believed to encode for a member of the caseinolytic protease family of molecular chaperones. This is believed to encode for heat shock inducible, ATP-dependent protease complex (Rathore et al., 2001). There are 7 unidentified small ORFs present at different regions in the genome.

Though a few of these genes (*lsu rRNA*, *ssu rRNA*, *tufA*, *clpc*) have been partially sequenced from the *P. vivax* Sal I strain, they have been used only for phylogenetic studies to understand the evolutionary position of *P. vivax* (Rathore et al., 2001; Escalante et al., 2005; Perkins et al., 2007; Mitsui et al., 2010). There is as yet no detailed information on the functional aspects of these genes in *P. vivax*. We were thus curious to know (a) the complete sequence of *P. vivax* apicoplast genome; (b) whether the functional domains in the proteins that might be encoded by apicoplast genome are completely conserved in *P. vivax*; and (c) whether the differences observed between the genetic contents of the *P. falciparum* nuclear genome (∼80% A/T rich) and *P. vivax* nuclear genome (∼58% A/T rich) and the isochore nature (Carlton et al., 2008) of *P. vivax* chromosome have introduced any codon biasness in the translation of apicoplast genome products in the two parasites.

Here we report the complete and partial sequences of major genes (*ssu* and *lsu rRNA*, *tRNA*, *orf470*/*sufB*, *rpoB*, *rpoC1*, *rps2*, and *clpC*) from the apicoplast genome of Indian *P. vivax* field isolates. A comparative genomic and functional analysis of these *P. vivax* sequences, with sequences for similar genes available from other *Plasmodium* species has also been performed. Sequences of *tufA* genes amplified from new field isolates have been compared with sequences from the previous report (Saxena et al., 2007) to analyze variations in field isolates. A number of phylogenetic studies have been reported using apicoplast genes (mainly *tufA* by Escalante et al. (2005) and Mitsui et al. (2010); *clpC* by Rathore et al. (2001); *ssu rRNA* by Nishimoto et al. (2008) and Mitsui et al. (2010)) to elucidate the evolutionary position of *P. vivax*. We have also performed a preliminary phylogenetic analysis using apicoplast sequences from Indian *P. vivax* isolates in comparison with the *P. vivax* Sal I and orthologues from other *Plasmodium* species.

2. Material and methods

2.1. Plasmodium vivax sample collection and DNA isolation

Blood samples of clinically proven *P. vivax* malaria cases were collected by trained clinicians, from Bikaner district, Rajasthan, India. A formal approval of participating Institutes' ethical committees and consent of patients were obtained for further studies. The blood samples were preserved in ACD (acid citrate dextrose) anticoagulant, shipped to our lab at BITS, Pilani in cold chain and stored at −20 °C. Complete DNA was isolated from these samples as described previously (Saxena et al., 2007) by treating them with lysis buffer (NaCl, Tris–HCl (pH 8.0), EDTA and SDS (1%)) and proteinase K (Sigma) followed by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and overnight ethanol precipitation at −20 °C. The DNA was precipitated and suspended in TE buffer (pH 8.0). The concentration and purity of samples were determined by conventional method of optical density (O.D.) measurement at A260 and ratio of A260/A280, respectively. Samples with high concentration of DNA (>∼75 ng/μl) and purity were used for PCR reactions. To verify presence of *P. vivax* and to exclude the presence of other human malaria parasites, a PCR based diagnostic procedure, using 18S rRNA gene based primers (Das et al., 1995; Kochar et al., 2005) was performed, using the whole blood DNA as template.

2.2. PCR primer designing and P. vivax plastid gene amplification

To amplify *P. vivax* plastid DNA genes, the primers (Table 1) were designed using *P. falciparum* plastid genome sequences obtained from NCBI database (Genbank accession nos. X95275 and X95276). All PCR protocols were first standardized using pure DNA isolated