Chapter 3

Introduction

The complete plastid DNA sequence has been reported from *Plasmodium falciparum* (Genbank Acc. No.: X95275 & X95276). The DNA was found to be circular and 35kb in size. About one third region of the circle is covered by inverted repeats namely IR_A and IR_B which includes various *tRNA* and *rRNA* genes. The rest of the circle includes genes for ribosomal proteins, RNA polymerase and various open reading frames with regions having identified as well as unidentified functions. The complete organization of circular DNA gives a clear indication of a plant origin. The genes show about 15% to 60% of identity at sequence level with those of blue – green or red algae genomes (Wilson *et. al.,* 1996; Williamson *et. al.,* 1997). The main conserved regions of this circular DNA are as follows:–

The *tufA* gene

The *tufA* gene has a high A+T content as any nuclear gene of *Plasmodium falciparum.* Despite this, it encodes one of the best conserved proteins of Plastid DNA. The predicted *Plasmodium falciparum* TufA peptide shares about 45% identity with the *tuf* genes of *E. coli.* As in *E. coli* peptide and other similar proteins, several highly conserved functional domains are evident, including the four clusters of residues in domain I involved in GTP binding. These segments in *E. coli* carry the consensus elements, G19HVDHGK25, D83CPG86, N138KCD141, and S176AL178, involved in binding the phosphoryl, Mg^{2+}, and guanine residues of GTP, respectively. In the reported *P. falciparum* sequence there is only one substitution, C140E. The residues defining the GDP binding pocket also are conserved (in *E. coli* G24, N138, K139, D141, S176, L178). In a less well conserved region, topologically close to the GTP binding domain (amino acid residues 183 to 192), the malarial sequence has a specific insertion like other plastid versions of EF-Tu. These residues form a loop barely

discernable in the mitochondrial equivalent (*TufM*) of *Saccharomyces cerevisiae* and *Homo sapiens* and absent from *E. coli* (Wilson *et. al.*, 1996).

The *ORF470/ sufB* gene

ORF470 encodes a well-conserved protein recorded from the plastids of three red algae, as well as the diatom *Odontella sinensis*, and *Mycobacterium leprae* (Wilson *et. al.*, 1996). The gene is reported to be an orthologue of *ycf24* and *sufB* protein genes. SufB protein is known to work in complex with other members of family, SufC and SufD. The three members make up a complex SufBCD which in presence of SufE can stimulate the desulphurase activity of SufS some 30-fold in the *suf* system. *Suf* system helps in maintenance and regulation of Iron – sulphur [Fe – S] protein cluster. Iron – sulphur proteins play important roles in electron transfer as well as in redox and non-redox catalysis. The [Fe-S] prosthetic group is also essential for several enzymes and proteins that participate in anabolic pathways in apicomplexan plastids. Moreover, in higher plants, [Fe-S] proteins are components of the plastid protein import complex at the level of the inner membrane (Wilson, 2005).

The *clp* gene

A member belonging to Caseinolytic protease family of molecular chaperons, present downstream to the *tuf* gene along with *rRNA*, *tRNA*, RNA polymerase subunits and ribosomal protein genes. The Clp system was first identified as a heat-shock-inducible, multicomponent, ATP-dependent protease complex capable of hydrolyzing casein. ClpC possesses a conserved AAA domain (ATPases associated with a variety of cellular activities) that belongs to the Walker super family of ATPases and GTPases. It has been predicted on the basis of sequence similarity that ClpC belongs to the double nucleotide binding form of Clp, although only the second of the two ATP-binding domains is conserved in the malarial protein. In malarial parasites its actual function remains unexplored experimentally (Rathore *et. al.*, 1999).

The *rpo* genes

The *rpoB* gene and part of *rpoC* have been described as indicators of the plastid origin of the 35kb circle. The complete sequence of *rpoC* shows that it lacks the intron typical of higher plants. Furthermore, the malarial circle gene homologous to *E. coli rpoC* is split into *rpoC*1 and *rpoC*2 as in other plastid and cyanobacterial genomes. The open reading frame of *rpoC*2 has a frameshift in a poorly conserved central section (Wilson *et. al.*, 1996).

Inverted repeats (Large and Small Subunit rRNAs, tRNAs)

The inverted repeat (IR) has a total length of approximately 10.5kb and encodes duplicated genes for small subunit (*SSU*) and large-subunit (*LSU*) *rRNAs* and nine different *tRNAs*. *tRNA* genes flank both forms of the *rRNA* gene in the IR. The disposition of the *rRNA* and *tRNA* genes in the IR differs in several respects from that of other recorded plastids. However, this order is conserved in *T. gondii*, suggesting that it may be an ancestral apicomplexan character derived from numerous rearrangements and deletions in an ancient progenitor of the apicomplexan clade. All these tRNAs and rRNAs are believed to be sufficient to provide a minimal but complete set for translation of the protein-encoding genes on the circle (Wilson *et. al.*, 1996).

The *rp* genes

A total of 17 ribosomal proteins encoding genes are spread throughout the circular DNA, mainly clustered between the *tufA* gene and 3' region of IR_B *LSU rRNA* gene. Among these 10 are *rps* and 7 are *rpl* encoding genes. Some of these genes (*rps2, rps12, rpl14, rpl16, rpl36*) are strongly conserved among plastid DNA reported from other apicomplexans while remaining (*rps4, rps5, rpl2, rpl23*) are poorly or non – conserved (*rpl4*). *rps2* gene is present downstream of the RNA polymerase genes *rpoB/C1/C2* , as in other plastid genomes (Wilson *et. al.*, 1996).

Results and Discussions

The *Plasmodium falciparum* apicoplast genome sequence was downloaded from the NCBI database. To amplify genes from plastid DNA of *Plasmodium vivax*, primers were designed with the help of downloaded sequence and using Generunner (Hastings Software Inc., USA) software. A normal as well as nested PCR approach was followed. Table 3.1 gives a list of all the primers that gave successful amplifications of plastid DNA genes from *P. vivax* DNA. Each pair of primers was first standardized for PCR using pure *P. falciparum* DNA. These reactions were then tried as such or with slight manipulations, using *P. vivax* DNA isolated from field samples. The amplified products were purified either with gel extraction method (QIAGEN) or with direct PCR purification method (QIAGEN). The purified amplicons were sequenced and obtained sequences were analysed. A general search for sequence match was done using BLAST N (Altschul *et. al.*, 1990), followed by alignment methods such as 'Gene Tool' (www.BioTools.com) or 'Clustal W' (Thompson *et. al.*, 1994). Most of the sequences showed a good match (approximately 80 – 85%) with the *P. falciparum* plastid DNA sequence or with sequences from other apicomplexan organisms. The sequences were then aligned against the *P. falciparum* plastid DNA sequence and variations were noted down.

The finalized nucleotide sequences were translated to amino acid. These were aligned against the amino acid sequences of other organisms. The differences were noted down mainly against sequences from *P. falciparum* and some from those of other apicomplexans. For TufA and RpoC protein, variations were also studied against *Escherichia coli*, *Thermus aquaticus* and *Thermus thermophilus* protein sequences. These variations formed a base in carrying out the further modeling or phylogeny analysis of genes identified from the apicoplast genome of *P. vivax.*

Primers from Apicoplast Genome

Table 3.1

GENE	PRIMER Id.	SEQUENCE
A) IR_A REGIONS		
Large Subunit rRNA gene (*LSU rRNA*)	VSAKD2AU2119	5' GCG GGA TCC AAG ACG TTA TAT ATA TTA AC 3'
	VSAKD2AD2119	5' GCG GAA TTC ATT TAT CTC TGT TTA ATT TGT A 3'
	VSAKD9d	5' CGC ACT CTT TAA AAG ATA ACT GCT TCT AA 3'
	VSAKD9d2	5' AAG TAT CTT TTA TCC CTA AGT TAC AAG 3'
	VSAKD19U	5' TAG GGT TAG TCG AAT CTT AAA ATG A 3'
	VSAKD19D	5' TTC AAC TTA TTA GGA ATT ATA CAC TA 3'
ORF470/ sufB gene	VSAKD8u	5' GAG TTC AAA TCT CAC CAT TAG CTT TTA TTA TT 3'
	VSAKD8d	5' GAT ACA TTT TAC TAC CAG TAT CAG CTA TTT GC 3'
	VSAKD8d2	5' GTA TTA ATT ACA TTT TTA TTA TAT TTC ATA TTT AA 3'
RNA polymerase C gene (*rpoC*)	VSAKD12U	5' ATG ATA ATA CAT AAT AAT ATA AAT TTT ATA GG 3'
	VSAKD12C	5' CCT AAT GGA TAA AAT TTT AAT GAA TAA CC 3'
	VSAKD12D	5' TTA ATA TAT ATA CAT ATA TAA ATT TAT AAT TAT TC 3'
RNA polymerase D/ C2 gene (*rpoD/ C2*)	VSAKD13U1	5' CTT CTA TTA ATA GAA TAA TTA TAA ATT TAT ATA TG 3'
	VSAKD13D1	5' GGA TAT TTT TTA TTA ATA ACT GTA TAT TAC T 3'
	VSAKD13U2	5' GTA ATA TAC AGT TAT TAA TAA AAA ATA TCC A 3'
	VSAKD13D2	5' TTA ATT TAC TAA ATA TCT ATA CCA TCC ATT AC 3'

Table 3.2

GENE	PRIMER Id.	SEQUENCE
B) IR_B REGIONS		
Elongation Factor gene (*tufA*)	VSAKD1BU1778	5' GCG GGA TCC CCT ATT ATA ATA TCG TAT G 3'
	VSAKD1BD1778	5' GCG GAA TTC AGA GCA ATG GAT TGA AGA T 3'
	VSAD3B1778U	5' CAG GTA GAG GTA CAG TAG TAA CAG GTA 3'
	VSAD3B1778D	5' TAC CTG TTA CTA CTG TAC CTC TAC CTG 3'
	VSAKD6u	5' ATG AAT AAT AAA TTA TTT TTA AGA AAT AAA CAA C 3'
	VSAKD6d	5' TTA ATT TTT TAT TTC TGT TAT AAT ACC TGC TCC 3'
Caseinolytic Protease C gene (*clpC*)	VSAKD7u	5' CAG AAA TAT GGG TAT TAA ATG ATT TAT TA 3'
	VSAKD2BD	5' GCG GAA TTC ACC GAT TAA TCT ATA TTT GTT A 3'
	VSAKD10U	5' ATG ATA ATT TTA AAT AAT CTT TAT TGT ACA AAA G 3'
	VSAKD10C	5' GAT AAT ATT TTT GCT AAT TCA GTT TTA CCA G 3'
	VSAKD17D	5' CTG TTA ATC TAC CTT CAT CTA ATA TTT G 3'
	VSAKD18D	5' TTA AGA ATG TTT GTA ATA TTG ATA AAT C 3'
IR_B Regions – *tRNA* genes, *rpl4* and *rpl23* genes	VSAKDIB3	5' TCG AAT CCC TTT TTC TCT ATT ATT AGA 3'
	VSAKD1B4	5' TAT CAT ACC AAC TAT CTA TTA ATT TAT AA 3'

Amplifications of *P. vivax* Plastid DNA genes

(A) Large Subunit rRNA (*LSU rRNA*)

The gene is 2695bp in size which expands from 2335 – 5029 position in the *P. falciparum* apicoplast genome IR_A (X95275) sequence. The primers were designed based upon the normal and nested PCR approach. The primers that gave amplification of 990bp (VSAKD2AU2119 and VSAKD9d), 1505bp (VSAKD2AU2119 and VSAKD9d2), 1392bp (VSAKD19U and VSAKD19D) and 880bp (VSAKD19U and VSAKD2AD2119) (Figure 3.1 A – D) are listed in the table 3.1.

The reaction conditions were -

Reaction Steps	**Primers used**	
	VSAKD2AU2119 and VSAKD9d	VSAKD2AU2119 and VSAKD9d2
Pre Denaturation	94^{o}C for 2 min	94^{o}C for 3 min
Denaturation	94^{o}C for 1 min	94^{o}C for 1.5 min
Annealing	47.5^{o}C for 2 min	50.1^{o}C for 2 min
Extension	72^{o}C for 2 min	72^{o}C for 2 min
Post Extension	72^{o}C for 4 min	72^{o}C for 4 min
Reaction Steps	**Primers used**	
	VSAKD19U and VSAKD19D	VSAKD19U and VSAKD2AD2119
Pre Denaturation	94^{o}C for 3 min	94^{o}C for 3 min
Denaturation	94^{o}C for 1 min	94^{o}C for 45 secs
Annealing	46.5^{o}C for 1.5 min	49^{o}C for 1 min
Extension	72^{o}C for 2 min	72^{o}C for 1.5 min
Post Extension	72^{o}C for 4 min	72^{o}C for 4 min

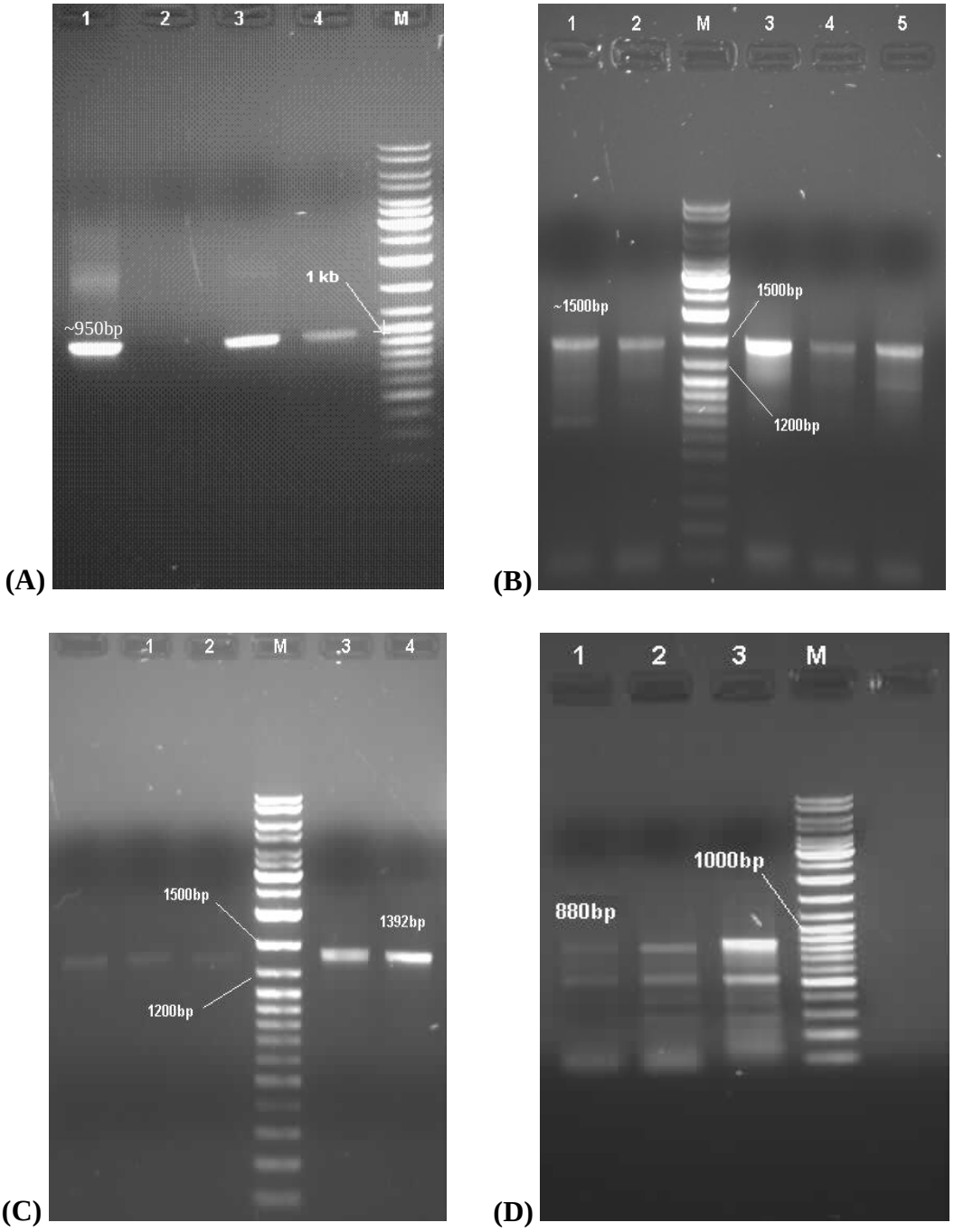
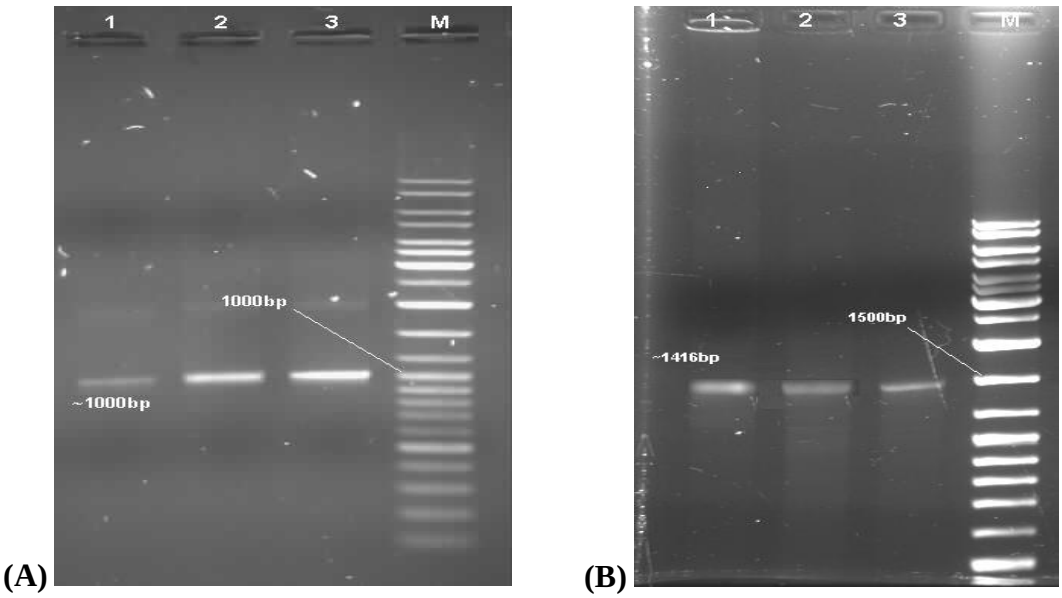


Figure 3.1: Amplification of regions from *P. vivax* plastid *LSU rRNA* gene - (A) VSAKD2AU2119 and VSAKD9d – Upper 950bp region; (B) VSAKD2AU2119 and VSAKD9d2 – Upper ~1500bp region; (C) VSAKD19u – VSAKD19d – Lower 1392bp region; (D) VSAKD19U and VSAKD 2AD2119 – Central 880bp region. M = 1kb DNA Ladder Mix (MBI Fermentas, #SM0331)

(B) *ORF470/ sufB*

The gene is 1413bp in size which expands from 5142 – 6554 position in the *P. falciparum* apicoplast genome IR_A sequence. The primers that gave amplification of approximately 1000bp (VSAKD8u and VSAKD8d) (Figure 3.2 A) and of complete 1413bp (VSAKD8u and VSAKD8d2) (Figure 3.2B) are listed in the table 3.1. The reaction conditions were –

Reaction conditions for approx. 1.0Kb (VSAKD8u and VSAKD8d**)**		**Reaction conditions for complete 1.413Kb (**VSAKD8u and VSAKD8d2**)**	
Pre Denaturation	- 94^{o}C for 2 min	Pre Denaturation	- 94^{o}C for 2 min
Denaturation	- 94^{o}C for 45 secs	Denaturation	- 94^{o}C for 1 min
Annealing	- 52^{o}C for 1 min	Annealing	- 47.5^{o}C for 2 min
Extension	- 72^{o}C for 1.5 min	Extension	- 72^{o}C for 2 min
Post Extension	- 72^{o}C for 4 min	Post Extension	- 72^{o}C for 4 min



(A) **(B)**

Figure 3.2: Amplification of *P. vivax* plastid *sufB/ ORF470* gene - (A) VSAKD8u and VSAKD8d – Upper 1.0kb region; (B) VSAKD8u and VSAKD8d2 – Complete 1.4kb gene; M = 1kb DNA Ladder Mix (MBI Fermentas #SM0331)

(C) Elongation Factor (*tufA)*

The gene is 1233bp in size and expands from 8755 – 9987 position in the *P. falciparum* apicoplast genome IR_B region. The primers were designed based upon the normal and nested PCR pattern. Initially primers were designed from the flanking region of the gene taking into consideration the high proportion of A/T present at the start and the end of the gene. These primers listed in Table 3.1, VSAKD1BU1778 and VSAKD1BD1778 would amplify a region of 1778bp (Figure 3.3). These primers carry a restriction site Bam H1 (upstream) and EcoR1 (downstream) at 5' end. The reaction conditions used were -

Primers: VSAKD1BU1778 and VSAKD1BD1778: 1778bp

Reaction Conditions:

Pre Denaturation - 94^{o}C for 2 min

Denaturation - 94^{o}C for 1 min

Annealing - 48^{o}C for 1.5 min

Extension - 72^{o}C for 2 min

Post Extension - 72^{o}C for 4 min

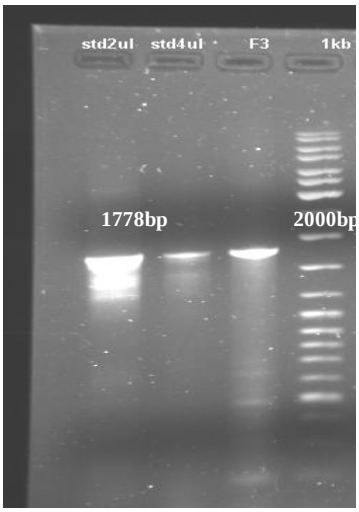


Figure 3.3: Amplification trial of 1778bp region for *tufA* gene

After PCR a band at approximately 1778bp position was visible on gel electrophoresis. This DNA band was eluted and sequenced. On sequence analysis

match of this 1778bp region was found with the human DNA which proved that there was amplification of human DNA and not the required gene.

Another pair of primers (VSAKD6u and VSAKD6d) was designed (Table 3.1), from the start and the end of the gene respectively. This would amplify the *tufA* gene region of 1233bp. Amplification of the desired region was obtained (Figure 3.4) and the gene amplification was confirmed by preliminary analysis using nested primers, VSAD3B1778U and VSAD3B1778D (Figure 3.5 A, B). The reaction conditions were:

Primers: VSAKD6u and VSAKD6d = approximately **1.23kb**

Reaction Conditions:

Pre Denaturation - 94^{o}C for 2 min

Denaturation - 94^{o}C for 1 min

Annealing - 48^{o}C for 1.5 min

Extension - 72^{o}C for 2 min

Post Extension - 72^{o}C for 4 min

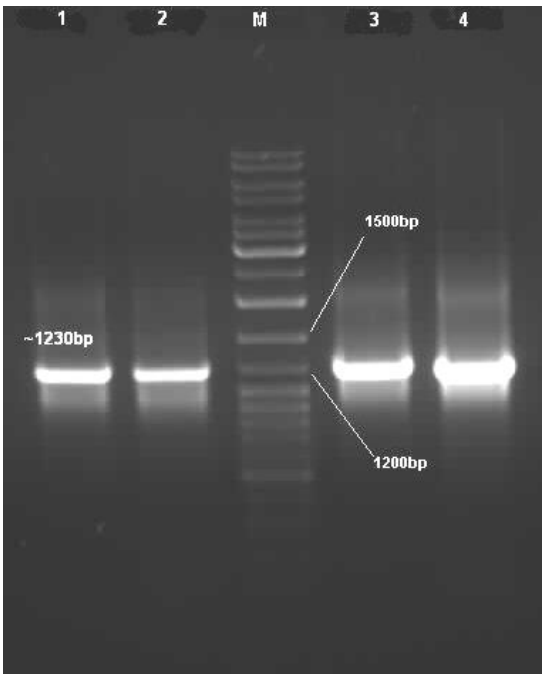
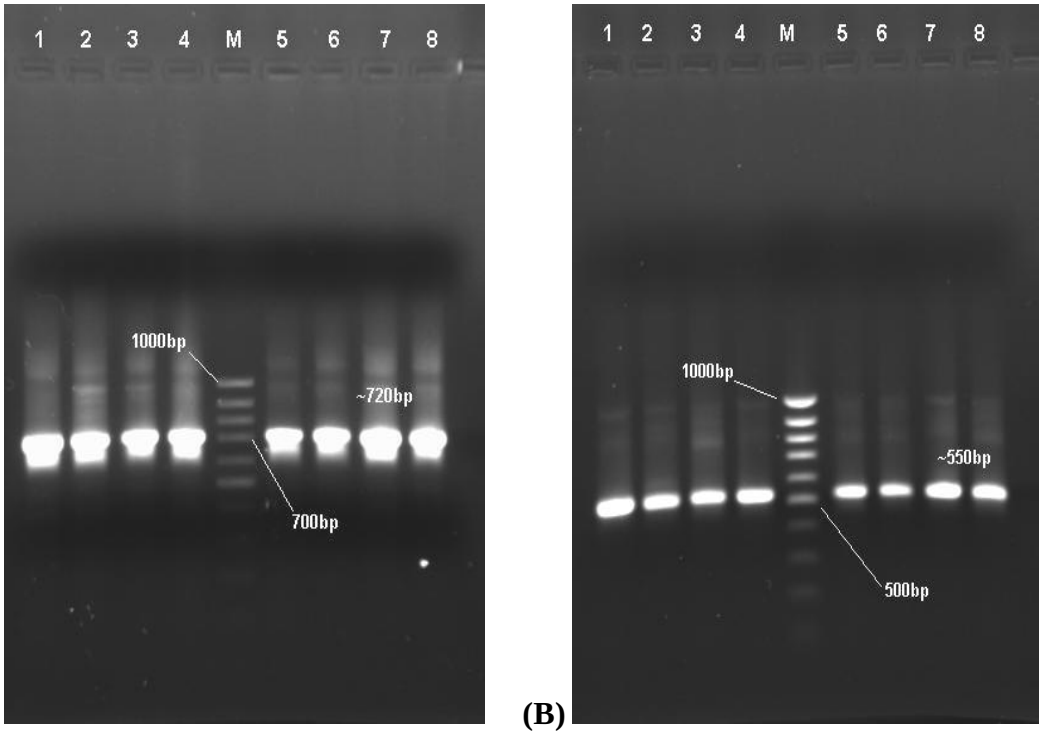


Figure 3.4: Amplification of complete *tufA* gene from *P. vivax* plastid DNA of Indian isolates using primers VSAKD6u AND VSAKD6d. Lanes 1 – 4= *P. vivax* isolates; M= 1kb DNA Ladder Mix (MBI Fermentas #SM0331)

Primers for First half of *tufA*: **720 bp** VSAKD6U and VSAD3b1778D	Primers for Second half of *tufA*: **550 bp** VSAD3b1778U and VSAKD6D
Reaction Conditions: Pre Denaturation - 94^{o}C for 2 min Denaturation - 94^{o}C for 1 min Annealing - 50^{o}C for 1.5 min Extension - 72^{o}C for 1.5 min Post Extension - 72^{o}C for 4 min	Reaction Conditions: Pre Denaturation - 94^{o}C for 2 min Denaturation - 94^{o}C for 1 min Annealing - 53^{o}C for 1 min Extension - 72^{o}C for 1min 15 sec Post Extension - 72^{o}C for 4 min



(A) **(B)**

Figure 3.5: Amplification of *P. vivax* plastid *tufA* gene regions – (A) VSAKD6U and VSAD3b1778D – Upper 720bp region; (B) VSAKD6D and VSAD3b1778U – Lower 550bp region; Lanes 1-8- *P. vivax* isolates; M= 1kb DNA Ladder Mix (MBI Fermentas #SM0331)

(D) Caseinolytic Protease C (*clpC*)

The gene is 2331bp in size which expands from 10934 to 13234 positions in the *P. falciparum* apicoplast genome IR_B sequence. The primers were designed for the complete and partial regions of the gene. Using the designed primers, amplifications of ~1500bp and 1112bp regions (Figure 3.6 A, B) were obtained while the complete gene could not be amplified. The primers that gave amplifications are listed in the table 3.1.

The reaction conditions were –

Reaction Steps	**Primers used**		
	VSAKD7u & VSAKD2BD	VSAKD10U & VSAKD10C	VSAKD7U & VSAKD17D
Pre Denaturation	94^{o}C for 3 min	94^{o}C for 3 min	94^{o}C for 3 min
Denaturation	94^{o}C for 1.5 min	94^{o}C for 1.5 min	94^{o}C for 1 min
Annealing	46^{o}C for 3 min	49^{o}C for 2.5 min	48.1^{o}C for 1.5 min
Extension	72^{o}C for 2.5 min	72^{o}C for 3min	72^{o}C for 2 min
Post Extension	72^{o}C for 4 min	72^{o}C for 4 min	72^{o}C for 4 min

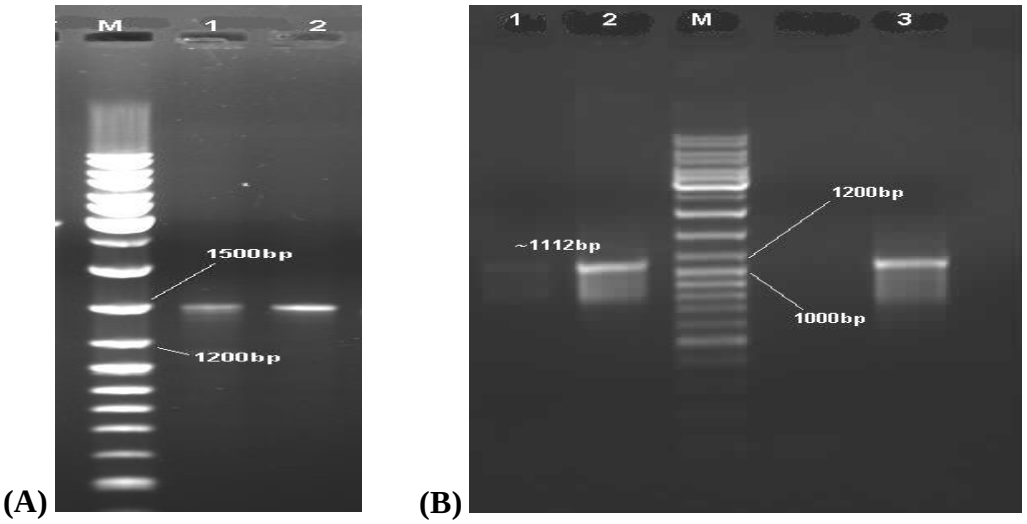


Figure 3.6: Amplification of *P. vivax* plastid *clpC* gene regions - (A) VSAKD7u – VSAKD2BD – Lower 1.5kb region; (B) VSAKD7u–VSAKD 17d – Central 1.112kb region; Lanes 1–3= *P. vivax* isolates; M= 1kb DNA Ladder Mix (MBI Fermentas #SM0331)

(E) RNA Polymerase C (*rpoC*)

The gene is 1728bp in size which expands from 10108 to 11835 positions in the *P. falciparum* apicoplast genome IR_A sequence. The primers were designed based upon the normal PCR approach. The primers that gave amplification of complete 1728bp (VSAKD12U and VSAKD12D) (Figure 3.7) are listed in the table 3.1. The reaction conditions were –

Primers: VSAKD12U and VSAKD12D: 1728bp

Reaction Conditions:

Pre Denaturation - 94^{o}C for 3 min

Denaturation - 94^{o}C for 1.5 min

Annealing - 46^{o}C for 3 min

Extension - 72^{o}C for 2.5min

Post Extension - 72^{o}C for 4 min

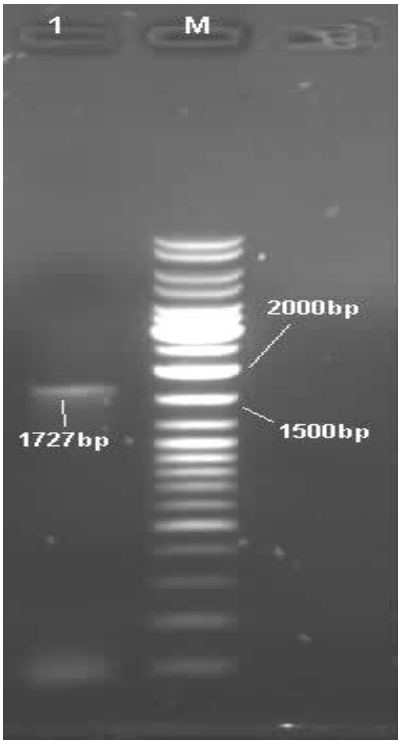


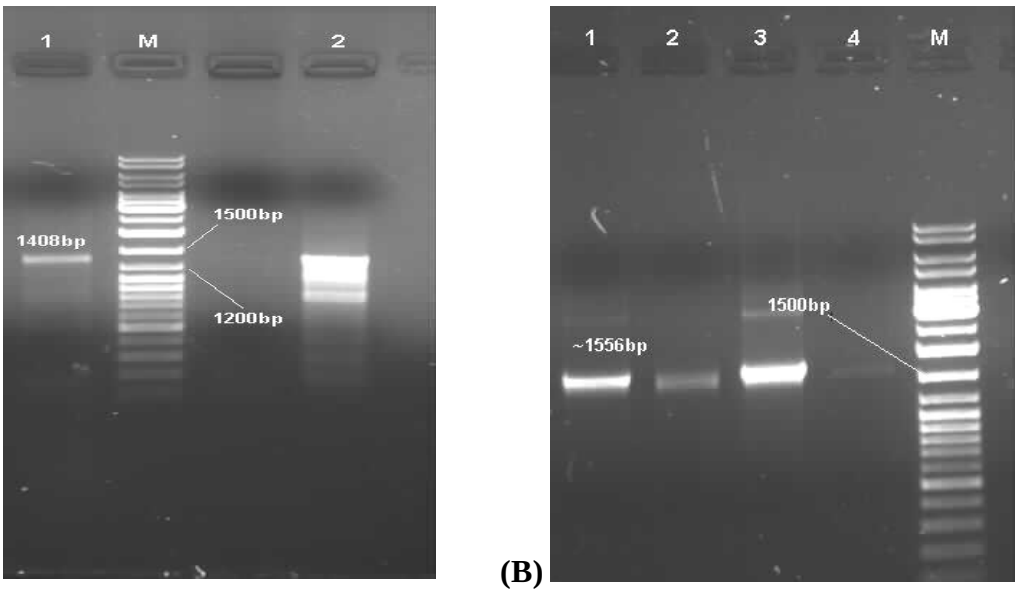
Figure 3.7: Amplification of *P. vivax* plastid *rpoC* gene (1728bp);
M= 1 kb DNA Ladder Mix (MBI Fermentas #SM0331)

(F) RNA Polymerase D/ C2 (*rpo D/ C2*)

The gene is 2882bp in size which expands from 11844 to 13418 and 13418 to 14725 positions in the *P. falciparum* apicoplast genome IR_A sequence. The primers were designed based upon the normal and nested PCR approach. The primers that gave amplification of 1408bp (VSAKD13U1 and VSAKD13D1) (Figure 3.8A) and of 1556bp (VSAKD13U2 and VSAKD13D2) (Figure 3.8B) are listed in the table 3.1.

The reaction conditions were –

Reaction Steps	**Primers used**	
	VSAKD13U1 and VSAKD13D1	VSAKD13U2 and VSAKD13D2
Pre Denaturation	94^{o}C for 3 min	94^{o}C for 3 min
Denaturation	94^{o}C for 1.5 min	94^{o}C for 1.5 min
Annealing	45.8^{o}C for 2 min	46.2^{o}C for 2 min
Extension	72^{o}C for 2.5 min	72^{o}C for 2.5 min
Post Extension	72^{o}C for 4 min	72^{o}C for 4 min



(A) **(B)**

Figure 3.8: Amplification of *P. vivax* plastid *rpoD* gene - (A) VSAKD13U1- VSAKD13D1 – Upper 1.4kb region; (B) VSAKD13U2 - VSAKD13D2 – Lower 1.5kb region; Lanes 1 – 4=*P. vivax* isolates; M = 1kb DNA Ladder Mix (MBI Fermentas #SM0331)

(G) IR_B Regions – *tRNA* genes, *rpl4* and *rpl23* genes

The IR_B region of *P. falciparum* plastid DNA includes a number of ribosomal protein genes, *tRNA* and *rRNA* genes. All these genes vary in length from 200bp to 3000bp and thus amplification of individual gene is difficult and unfeasible. Thus, primers were designed to amplify two or more smaller genes together from *P. vivax* isolates. A region of 1312bp, from position 1636 to 2948, including tRNA genes (*D, K, E, P*) and ribosomal proteins (4, 23) was amplified (Figure 3.9) using primers VSAKDIB3 and VSAKDIB4 (Table 3.2). Following reaction conditions were used for the amplification –

Primers: VSAKDIB3 and VSAKDIB4: 1312bp

Reaction Conditions:

Pre Denaturation - 94°C for 3 min

Denaturation - 94°C for 1 min

Annealing - 45.8°C for 2 min

Extension - 72°C for 2 min

Post Extension - 72°C for 4 min

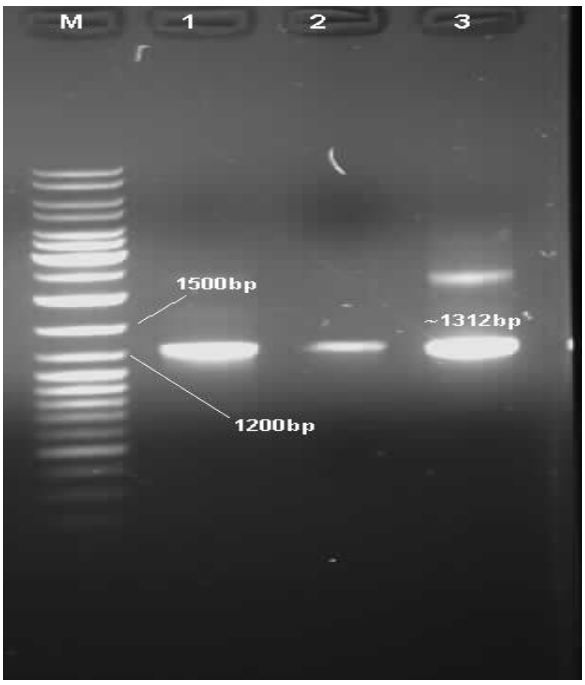


Figure: 3.9: Amplification of *P. vivax* plastid DNA IR_B region including genes for *tRNAs*, *rpl4* and *rpl23***.** Lanes 1-4 = *P. vivax* isolates; M = 1kb DNA Ladder Mix (MBI Fermentas #SM0331)

Sequence Analysis

The parasites *P. falciparum* and *P. vivax* have been divided into separate classes based on their genome A/T content. The later having more G/C with an A/T content of approximately 70 – 75 %. This fact is likely to cause difference in their codon usage pattern. The plastid DNA molecule has its own origin of replication (Williamson *et. al*., 2002; Singh *et. al*., 2003; Singh *et. al*., 2005; Williamson *et. al*., 1996) and replicates essentially during the organelle division. To understand whether the A/T content and codon usage variation has any effect on the plastid DNA gene usage, we have focused on the sequence comparison of the genes of plastid DNA of the two parasites. The amplified fragments were gel extracted and sequenced. The obtained sequences were compared with the available NCBI database using BLAST N (Altschul *et. al*., 1990) followed by specific comparison with reported *P. falciparum* plastid DNA sequence using Clustal W (Thompson *et. al*., 1994). The obtained sequences were analysed based upon two parameters. The nucleic acid changes were noted at first followed by changes at amino acid level. The changes at nucleic acid level showed a varying pattern for different genes. These changes affected the amino acid sequence to a certain extent while deletion or addition of amino acids was also noted.

tufA Gene and Protein Sequence

The *P. falciparum* plastid DNA *tufA* gene is 1233bp in size. The amplified *P. vivax tufA* gene products (Figure 3.4), when sequenced (Appendix A.1 a, b, c) and analysed were found to be 1230bp in length as compared to expected 1233bp sequence. *tufA* gene products from different isolates were sequenced to confirm this length of gene. The verified sequence was subjected to BLAST N (Altschul *et. al.*, 1990) where the sequence showed certain percent match with *P. falciparum* plastid DNA *tufA* gene sequence (89%) and with various other *tufA* sequences from different species present in the database. The so obtained *P. vivax tufA* sequence was then compared with the *P. falciparum* plastid DNA *tufA* gene sequence (X95276) using Clustal W (Thompson *et. al.*, 1994). There were about a total of 10 – 11% variations noted at the nucleic acid level. The difference of three bases (CCT) was present at 787^{th}-789^{th} positions

which were found absent in all the *P. vivax* sequences from our isolates. This could possibly introduce a gap at amino acid level in the sequence. The sequences from different *P. vivax* isolates showed a variation of only 2% among themselves but all had similar percentage variation with that of *P. falciparum* sequence. There was higher content of G+C in the *P. vivax* sequence, which may implicate that even the plastid DNA gene is not left out of the process of codon biasing. This can be confirmed only with analysis of other genes from the *P. vivax* plastid DNA.

The *P. vivax tufA* nucleic acid sequence was then translated to 410 amino acids and compared with 411 amino acid long *P. falciparum* sequence (Appendix C.1). There was 11 – 12% variation in the genes of the two parasites. Among all the differences noted, about half of the amino acid variations were conserved and the other half showed either no consensus or weak conservation of amino acids (Table 3.3). There was a gap at 263rd amino acid position in the *P. vivax* TufA sequence where a Proline (P263) present in *P. falciparum* sequence was missing. A Proline is known to produce a kink in the structure. Its absence may lead to certain change in the structure. Some other variations noted (Table 3.4) were analysed on the basis of amino acid properties.

Table 3.3: Major Amino Acid variations in EF – Tu protein sequence

a) Amino acid variation showing no consensus		
Amino acid position	***P. falciparum***	***P. vivax***
167	Tyrosine	Asparagine
196	Tyrosine	Glycine
216	Lysine	Isoleucine
223	Methionine	Arginine
240, 403	Glycine	Arginine
241	Lysine	Proline
246	Cysteine	Tyrosine
263	Proline	- (A deletion)
264	Asparagine	Isoleucine
279	Threonine	Isoleucine

b) Amino acid variation showing conservation of weak groups		
Amino acid position	***P. falciparum***	***P. vivax***
31	Alanine	Proline
142	Aspartate	Serine
208, 219	Aspartate	Glycine
213	Proline	Threonine
218	Asparagine	Glycine
239	Threonine	Lysine
367	Alanine	Glycine

Table 3.4: Amino Acid variations based on specific properties in EF – Tu

a) Hydrophilic to Hydrophobic			
A.A. Position	**Hydrophilic**	**Hydrophobic**	**From - To**
12	Arginine	Histidine	*P. vivax - P. fal*
29	Serine	Threonine	*P. vivax - P. fal*
31	Proline	Alanine	*P. vivax - P. fal*
167	Asparagine	Tyrosine	*P. vivax - P. fal*
196	Glycine	Tryptophan	*P. vivax - P. fal*
213	Proline	Threonine	*P. fal – P. vivax*
216	Lysine	Isoleucine	*P. fal – P. vivax*
223	Arginine	Methionine	*P. vivax - P. fal*
239	Lysine	Threonine	*P. vivax - P. fal*
240	Arginine	Glycine	*P. vivax - P. fal*
246	Cysteine	Tyrosine	*P. fal – P. vivax*
264	Asparagine	Isoleucine	*P. fal – P. vivax*
346	Arginine	Histidine	*P. fal – P. vivax*
363	Asparagine	Valine	*P. fal – P. vivax*

b) Aliphatic to Aromatic			
A.A. Position	**Aliphatic**	**Aromatic**	**From – To**
28	Leucine	Phenylalanine	*P. fal – P. vivax*
167	Asparagine	Tyrosine	*P. vivax - P. fal*
196	Glycine	Tryptophan	*P. vivax - P. fal*
246	Cysteine	Tyrosine	*P. fal - P. vivax*
265	Leucine	Phenylalanine	*P. fal - P. vivax*

To further verify our findings, that the obtained *P. vivax tufA* gene sequence is of prokaryotic origin, original database sequence for *Thermus aquaticus* (SwissProt accession no. **Q01698**) and *Thermus thermophilus* (SwissProt accession no. **P07157**) were aligned with *Escherichia coli* EF-*Tu* (SwissProt accession no. **P02990**). Next *P. falciparum* EF-Tu (SwissProt accession no.**Q25820**) and our *P. vivax* sequences were aligned to the above alignments to understand how this sequence is different from other reported EF-Tu. Alignment problems were resolved using an iterative procedure of model building alternatives. The sequence alignment was further processed and annotated using ALSCRIPT (Barton, 1993) (Figure 3.10).

Three gaps were observed in the *E. coli* sequence compared with *T. aquaticus*: one residue between Gly41 and Gly42 after a helix A; 10 residues between Glu178 and Gly179 deleting part of helix E, and one residue between Lys247 and Glu248 in the loop between β strands c2 and d2. Alignment of the *P. vivax* and the *P. falciparum* EF-Tu sequences to the above alignments showed that gaps or insertions occurred in the same three regions as in *E. coli* sequence, plus one additional Tyrosine deletion and one Serine insertion (Figure 3.10). Along with one deletion of Pro263 in the *Plasmodium vivax* sequence, the two *Plasmodium* sequences also differ at 51 positions, which include 19 conservative substitutions. On primary structure level most of these differences do not co localize with important functional sites except Arg222, Ala223 and a conservative Leu237 (Figure 3.10)

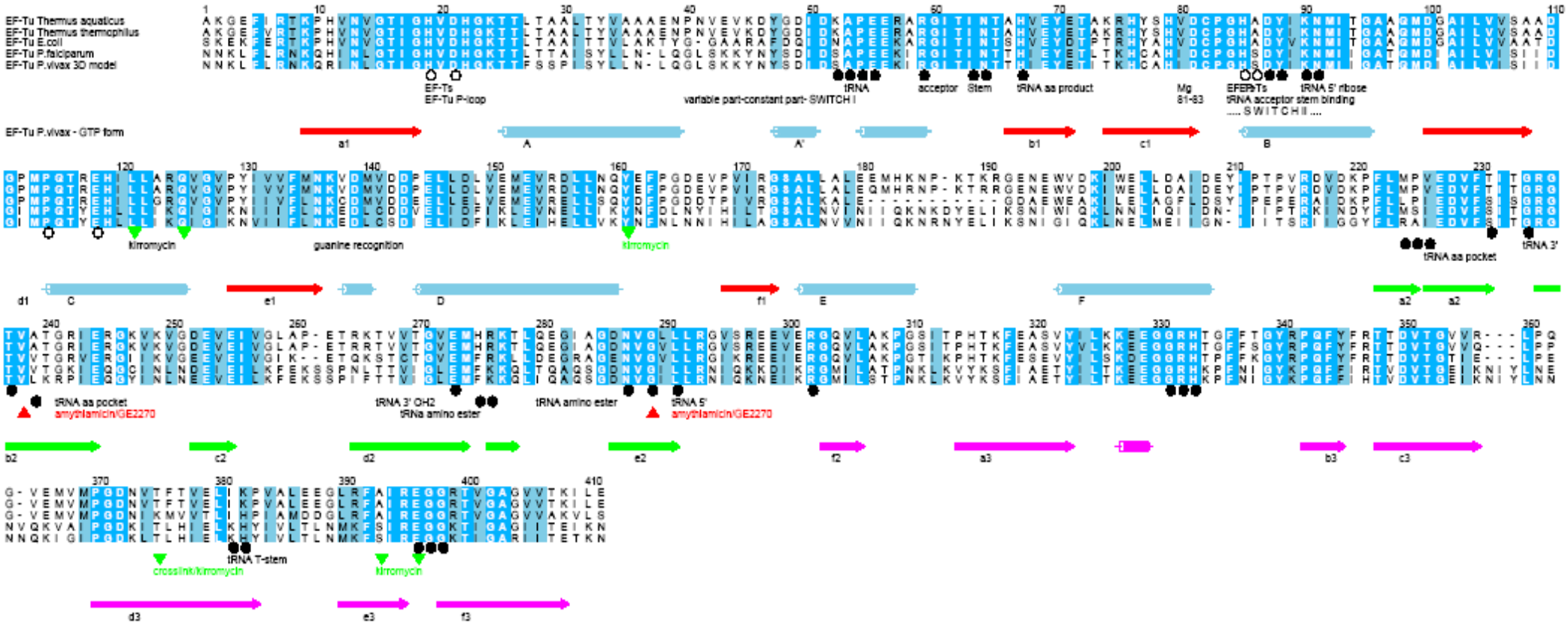


Figure 3.10: Amino acid alignment of *P.vivax* EF-Tu with homologous *P. falciparum* EF-Tu (Swissprot Q25820) sequence and homologous EF-Tu sequences for which the three- dimensional structure is known: *Thermus aquaticus* (Swissprot Q01698), *Thermus thermophilus* (Swissprot P07157) and *Escherichia coli* (Swissprot P02990). As par the accepted nomenclature for prokaryotes, residue 1 is the first after the start methionine. Secondary structure elements are shown above the alignment for the *P.vivax* TufA model where arrows- coloured for each domain as in b- represent β strands and cylinders α helices. Functional regions are indicated below the alignment, e.g. filled circles are conserved amino acids which bind to tRNA and open circles are conserved amino acids binding to EF-Ts. Green triangles and red triangles are sites of kirromycin and amythiamycin resistance respectively. Figure prepared using ALSCRIPT.

Further the obtained *P. vivax* TufA protein sequence was also analysed for variations in the active sites. The reported *P. falciparum* TufA amino acid sequence carries active sites for binding of phosphoryl (G19HVDHGK25), Mg^{2+} (D83CPG86), tRNA – GTP complex (N138KED141), GDP binding pocket (G24, N138, K139, D141, S176, L178) and antimicrobial agents like amythiamicin and kirromycin. The primary sequence shows amino acids conservation at higher level with the *E. coli* and *T. aquaticus* TufA sequences. Most conserved residues form the core of EF – Tu GTP binding. Mutations at V237, G287, G329, Q341, A387 and E390 in *P. falciparum* are responsible for resistance towards antimicrobials amythiamicin and kirromycin (Wilson *et. al.*, 1996; Rathore *et. al.*, 2001). The analysed *P. vivax* TufA sequence showed similar binding pockets and amino acids conservation.

A critical view of the *P. vivax* TufA amino acid sequence shows that, although all the above mentioned active sites are conserved, variations can be seen at just the neighboring amino acids. There are variations from aromatic to aliphatic, hydrophilic to hydrophobic nature and vice – versa. The hydrophobic pocket for binding of tRNA molecule is also conserved. The variations present near the active sites (Table 3.5) may affect the ligand (tRNA – GTP complex) binding properties of the protein.

Table 3.5: Variations appearing in close vicinity of active sites in *tufA* gene

Active Site	***P. falciparum***	***P. vivax***
G19HVDHGK25	L28TTA31	F28SSP31
N138KED141	I131, D142, V144	V131, S142, I144
	P213, T214, K216, N218, D219	T213, S214, I216, G218, G219
V237, G287	V238TGK241, C246	L238KRP241, Y246
G329, Q341, A387, E390	Q361, V363, V366, A367	N361, N363, I366, G367

An extended verification of obtained *P. vivax tufA* gene sequences was done by comparing this sequence with elongation factor gene sequences from

nuclear genome database of *P. falciparum*. This comparative analysis revealed 32 – 42% variations between the *P. vivax* and *P. falciparum* sequences (Table 3.6) proving the prokaryotic origin of *P. vivax* sequences. The sequences search from *P. vivax* nuclear genome database resulted in partial putative plastid *tufA* gene sequence only which showed 95% similarity with our sequence.

Table 3.6: Comparative match between *Plasmodium falciparum* nuclear *tufA* genes and Indian *P. vivax tufA*

S. No.	Database Sequence annotation	Percent match
1.	*P. falciparum* / CHR 11 / translation elongation factor EF-1, subunit alpha, putative	60.6%
2.	*P. falciparum* / CHR 6 / elongation factor G, putative	65%
3.	*P. falciparum* / CHR 3 / elongation factor (EF-TS), putative	60.5%
4.	*P. falciparum* / CHR 3 / elongation factor 1 (EF-1), putative	66.7%
5.	*P. falciparum* / CHR 6 / elongation factor G, putative	65%
6.	*P. falciparum* / CHR 7 / GTP-binding translation elongation factor tu family protein, putative	66.1%
7.	*P. falciparum* / CHR 11 / translation elongation factor EF-1, subunit alpha, putative	60.6%
8.	*P. falciparum* / CHR 12 / elongation factor g, putative	68.8%
9.	*P. falciparum* / CHR 13 / elongation factor tu, putative	65.2%
10.	*P. falciparum* / CHR 13 / elongation factor 1-gamma, putative	60.7%
11.	*P. falciparum* / CHR 13 / elongation factor Tu, putative	65.9%
12.	*P. falciparum* / CHR 13 / PF13_0304 elongation factor 1 α	59.7%
13.	*P. falciparum* / CHR 13 / PF13_0305 elongation factor 1 α	59.7%
14.	*P. falciparum* / CHR 14 / PF14_0486 elongation factor 2	64.9%
15.	*P. vivax* / Pv_5111.phat_1 90% identity to 94% of S72277: translation elongation factor EF-Tu - *P. falciparum* plastid	95.2%

ORF470/ sufB Gene and Protein Sequence

The *P. falciparum* plastid DNA *ORF470/ sufB* complete gene is 1413bp in length. As discussed above, initially two primers, one from start of gene and other from central region (~ 1059bp) were designed and used for amplification. The amplified and purified *P. vivax ORF470/ sufB* gene products, approximately 1059bp in length, were sequenced (Appendix B.1 a, b, c) and analysed. The verified sequence was subjected to BLAST N (Altschul *et. al.*, 1990) where the sequence showed certain percent match with *P. falciparum* plastid DNA *sufB* gene sequence and with sequences of *sufB* and *ycf24* (a red algae protein gene) homologues from different species present in the database.

The partial sequences were obtained from three *P. vivax* field isolates. These sequences were compared with *P. falciparum sufB* gene sequence using Clustal W (Thompson *et. al.*, 1994), and were found to have approximately 13 – 15% variations with *P. falciparum* gene and about 1 – 2% variations among themselves. Most of these variations were single base change variations. There were some deletion and insertions of single base as well. The sequences were translated to amino acids and analysed over BLAST P (Altschul *et. al.*, 1990). The sequence showed similarity with ABC transporter associated protein, common with all ORF470 genes. The sequence also showed matches with Fe – S assembly component SufB protein. Clustal W analysis of the amino acid sequences of these *P. vivax* SufB partial regions against *P. falciparum* SufB also showed about 13% variations.

To get amplification and sequence of complete *P. vivax sufB* gene, a third primer from end of gene (downstream) was designed and amplifications were tried for whole gene (1413bp) from *P. vivax* isolates. Amplification of the complete *ORF470/ sufB* gene could be obtained from only single isolate which was different from above three isolates. Many trials were carried out by varying PCR reaction conditions to get amplification from the above three and other isolates without any success. The only obtained 1413bp amplicon was

purified, sequenced (Appendix A.2) and analysed. To verify the obtained sequence as *P. vivax sufB* gene a BLAST N was carried out and similar hits were obtained as for the above three isolates. This verified complete *P. vivax sufB* gene sequence when compared with *P. falciparum sufB* gene sequence using Clustal W showed only 3% single base variations as compared to the above three isolates showing 13% variations. The sequence was then translated to amino acids and again compared with the *P. falciparum* sequence (Appendix C.2). The noted variations (Table 3.7) mainly had weak amino acid conservation or no consensus. At structural and functional level these may not cause any major variation to the protein.

Table 3.7: Major amino acid variations between *P. vivax* isolates and *P. falciparum* SufB

Position of Amino Acid in *P. falciparum*	**Amino Acid *P. falciparum***	**Amino Acid *P. vivax IndIso1***
161,198, 213, 257	Threonine	Proline
184	Cysteine	Glycine
192, 229	Cysteine	Tryptophan
211, 245	Glutamic Acid	Glycine
221	Serine	Proline
223	Valine	Glycine
262	Tyrosine	Aspartic Acid
265	Aspartic Acid	Glycine

Large Subunit ribosomal RNA gene (*LSU rRNA*)

The complete gene is approximately 2.7Kb in size and is present in duplicate copy in the inverted repeat region of the plastid DNA. Primers were first designed to amplify a 2119bp region from this gene. No amplifications could be obtained from these primers using all of the available *P. vivax* isolates. Following this, few more primers were designed for the partial regions of the

gene to get amplifications of 990bp from the upper half and 1392bp from the lower half. Amplification of above two desired regions were obtained from four *P. vivax* isolates and standard *P. falciparum* field isolate. The *P. vivax* products were purified and sequenced. Sequences could be obtained for the upper half region from all four isolates but from only one isolate for the lower half region (Appendix B.3 a-e). Following the similar approach as for *tufA* and *sufB*, the sequences were checked with BLAST N and were found to be similar to mainly *P. falciparum* plastid DNA *LSU rRNA* gene and various homologues from different organisms present in the NCBI database. Further, these *LSU rRNA* nucleotide sequences from all *P. vivax* isolates for both the regions were compared using Clustal W (Thompson *et. al*., 1994) among themselves and with the *P. falciparum* plastid *LSU rRNA* gene sequence (Appendix C.3 a, b).

Sequence analysis revealed a number of variations including base deletions and additions (Table 3.8) in *P. vivax* sequence as compared to *P. falciparum* sequence.

Table 3.8: Nucleotide base variation in LSU *rRNA* gene regions between *P. vivax* isolates and *P. falciparum*

Upper ~900bp region and Lower 1228bp region	
Position of base in *P. falciparum*	**Number of bases**
Addition of Nucleotide Bases in *P. vivax* gene from Indian Isolate	
82, 99, 342, 354, 1563	1
1689	2
283, 1434	3
Deletion of Nucleotide Bases in *P. vivax* gene from Indian Isolate	
433, 513, 1822	1
1397 – 1398, 2041- 2042, 2504 - 2505	2
64 – 66, 2017 - 2019	3
1407 - 1410	4
227 -231	5

The variations among the *P. vivax LSU rRNA* gene sequences from four isolates were between 2 – 3% in all but there were similar variations of all these four sequences with the *P. falciparum LSU rRNA* sequence which accounted for about 9% in total. The comparative analysis of the lower region sequence from *P. vivax* isolate with that of *P. falciparum* showed only 4.5% variations.

Casienolytic Protease C (c*lpC*) gene

The gene is 2.3kb in length and is present in the IR_B region of the *P. falciparum* plastid DNA circle. Primers designed to amplify the complete gene from *P. vivax* isolates could not give any amplification but showed amplification of approximately 2.3kb with the pure *P. falciparum* DNA. Thus, again a nested PCR approach was tried to amplify the parts of this gene. Amplifications were obtained for the lower 1500bp region of the gene from single isolate. The amplicon was sequenced (Appendix B.2 a, b) and analysed as for above genes. Several trials were made to amplify the remaining upper 800bp region of the gene without any success.

The obtained sequences could give 475bp and 687bp (position 730 – 1204 and 1614 – 2301 in *P. falciparum clpC* gene respectively) regions from the lower half of the gene. A comparison of these regions with the *P. falciparum* gene showed almost 8% overall variations. The 687bp region of Indian *P. vivax clpC* showed about 92% similarity with the Salvador *P. vivax clpC* 642bp region. The amino acid translated sequence of the *P. vivax* 687bp region (amino acid position 539 – 764 in *P. falciparum clpC*) also showed 8% variations (Table 3.8) from that of *P. falciparum,* with a deletion of Glycine at 552 position (Appendix C.4).

The region also contains a conserved domain for ATPase binding and showed similarity with 24% region of *clpA* gene. The BLAST N and BLAST P searches of this region showed varying but good (92 – 95%) match with similar genes and C – terminal protein region from other malaria parasites,

respectively. This might indicate that the gene is conserved even after many evolutionary changes. The N – terminal region of *P. falciparum* ClpC protein is known to be atypical (Wilson, 2005), thus the sequence of this region have to be obtained to state any conservational aspect of the protein.

Table 3.9: Major variations between *P. vivax* and *P. falciparum* 229 amino acids from ClpC sequences

Amino Acid Position in *P. falciparum* sequence	**Amino Acid in *P. falciparum***	**Amino Acid in *P. vivax***
539	Isoleucine	Asparagine
540, 544, 551	Glycine	Arginine
543	Proline	Serine
547	Glycine	Valine
549	Serine	Arginine
550	Glutamic Acid	Arginine
552	Glycine	- (A deletion)
553, 557	Glutamine	Proline
558	Valine	Glycine
756	Lysine	Threonine
759	Phenylalanine	Glycine

RNA Polymerase C (*rpoC*) gene

The gene is 1728bp in length. Based upon the *P. falciparum* plastid DNA sequence, three primers were designed to amplify and sequence the complete gene from *P. vivax* isolates. Amplification was obtained from different *P. vivax* isolates, but the complete gene could be sequenced (Appendix A.3) from only one isolate. The gene was first subjected to BLAST N where it showed 90 – 94% match with the similar gene from other *Plasmodium* species. These included *P. falciparum* (92%), *P. chabaudi* (92%), *P. yoelii* (90%) and *P. berghei* (93%).

The obtained *P. vivax* gene sequence was compared to the *P. falciparum rpoC* gene using Clustal W (Thompson *et. al.*, 1994) and about 8% variations were seen at the nucleic acid level while the percent score at amino acid level was 10% (Appendix C.5). These variations included deletion of three bases at position 522 – 524 in *P. vivax* sequences leading to a deletion of amino acid 'Lysine' from position 174 in *P. falciparum* sequence. The major amino acid variations are listed in the Table 3.9.

Table 3.10: Major amino acid variations between *P. vivax* and *P. falciparum* RpoC

Position in *P. falciparum* RpoC protein sequence	**Amino acid in *P. falciparum* sequence**	**Amino acid in *P. vivax* sequence**
68, 90, 213, 217, 436	Lysine	Isoleucine
86	Phenylalanine	Glycine
97	Isoleucine	Lysine
165	Serine	Leucine
174	Lysine	- (A deletion)
210	Isoleucine	Asparagine
214	Tyrosine	Proline
301	Threonine	Methionine
336	Threonine	Isoleucine
374	Asparagine	Tyrosine
388	Serine	Phenylalanine
405	Isoleucine	Serine
495	Lysine	Leucine
519	Glutamic Acid	Valine
541	Phenylalanine	Serine

There are two active sites reported in the RNA Polymerase C protein of *E. coli*. The amino acids present in these regions include five cysteine residues

responsible for binding of zinc motifs and two conserved regions, viz. G343KRV347D and N456ADFDG461D, for DNA template binding. These regions are reported to be conserved in *Plasmodium falciparum* RpoC (Gardner *et. al.*, 1991a). The obtained *P. vivax* sequence shows similar amino acid residues, indicating the conservation of active sites.