Chapter 7

Introduction

The genetic organization of the *Plasmodium* species is diverse. This genetic diversity has allowed malarial parasites to adapt to environmental insults such as host immunity and chemotherapy. Molecular mechanisms which dictate this genetic diversity include genetic crossing in the definitive host followed by a huge asexual expansion under the selective pressures of the intermediate host. Furthermore, individual parasitic species are likely to have had long evolutionary relationships with their definitive and intermediate hosts. It has been suggested that much of the intra – species polymorphisms have been as a result of random point mutation and recombination among subtelomeric as well as intragenic repeated sequences. As a result, *Plasmodium* species display a high degree of inter and intra – species divergence (McIntosh *et. al.*, 1998). This genetic diversity and the presence of a plastid organelle with a complete circular genome have kept the malariologists into a great puzzle over the years in regards to the evolutionary origin of malaria and different species of *Plasmodium.*

Phylogenetic studies carried out using various nuclear genes (*CSP*, *DHFR*, etc.) indicate towards a close relation between the human parasite *P. falciparum* and the chimpanzee parasite *P. reichenowi*. While the other human parasites including *P. vivax, P. malariae* and *P. vivax* – like parasites seem to have close relation with *P. knowlesi*, a common parasite of primates *Macaca fascicularis* and *M. mulatta* in Southeast Asia, and *P. cynomolgi*. This divergence between the major human malaria parasites seems to have started approximately 165 million years ago where *P. falciparum* got diverged from the clade of *P. vivax*, *P. malariae* and *P. vivax* – like parasites. The later three diverged from themselves around 100 million years ago. This corresponds to the time when there was divergence between the reptiles, birds and mammals (Escalante *et. al.*, 1994; Escalante *et. al.*, 1995a, b). Even the phylogenetic data derived from the mitochondrial genes (Escalante *et. al.*, 2005;

Rathore *et. al.*, 2001) indicates the divergence of *P. falciparum* from other human malaria parasites.

The apicoplast genome carries some of the highly conserved genes. A phylogenetic study related with these genes will provide a lead into the quest for the evolutionary origin of malaria and its parasites. The conserved genes encoding some major proteins and an organization that shows its independent functionality, indicates clearly towards the monophyletic origin of the organelle and its genome (Kohler *et. al*., 1997). The genome content shows high level of similarity with that of chloroplast genome (except for the missing photosynthetic genes as they were not required). The genes within this circular DNA (*sufB*) seems to have some relation with the nuclear genes (*sufC*) that are known to encode plastid targeted proteins. This fact has raised the concern at various time points over the functional properties of the plastid circular DNA. Since its discovery, the genes from this genome have been studied from various *Plasmodium* species such as rodent parasites *P. berghei* (whole genome) and *P. yoelii,* avian parasite *P. gallinaceum*, human parasites *P. falciparum* (whole genome) and *P. vivax* (only two partial genes from Salvador and Vietnam strains) and monkey parasites *P. knowlesi* and *P. cynomolgi.* Based upon the differences studied in the apicoplast genes isolated from Indian *P. vivax* isolates, we have tried to study the relation and divergence of Indian *P. vivax* from rest of the world malaria parasites.

Results and Discussions

The phylogenetic analysis was performed for three genes of plastid DNA, viz, *tufA*, *clpC* and *sufB*. The sequences of other species of *Plasmodium* were obtained from the NCBI database. Since much work has not been done on the plastid DNA genes, very less number of sequences were available from the database as compared to those of mitochondrial and nuclear genes.

tufA gene

Phylogeny using the *P. vivax* Salvador and other *Plasmodium* species *tufA* gene has been reported by Escalante (Escalante *et. al.*, 2005). We have used sequences of *tufA*

complete gene from three Indian isolates and analysed them against the *tufA* genes from *P. vivax* Salvador, Vietnam and other *Plasmodium* species. Complete analysis was done using MEGA 3.1 software. The sequences were first aligned using Clustal W (Thompson *et. al.*, 1994) and the alignment was subjected to various evolutionary distance calculation substitution models for nucleotide type data. The substitution model used for distance calculation for nucleotide substitution was Kimura 2 parameter (Table 7.1) which gave an average distance of 0.062 among all the sequences used. An average of 0.089 was obtained for synonymous substitution (Table 7.2) of codons by using Nei – Gojobori model. Phylogeny tree was built using neighbour – joining (NJ) method based upon the Kimura 2 distance data (Figure 7.1).

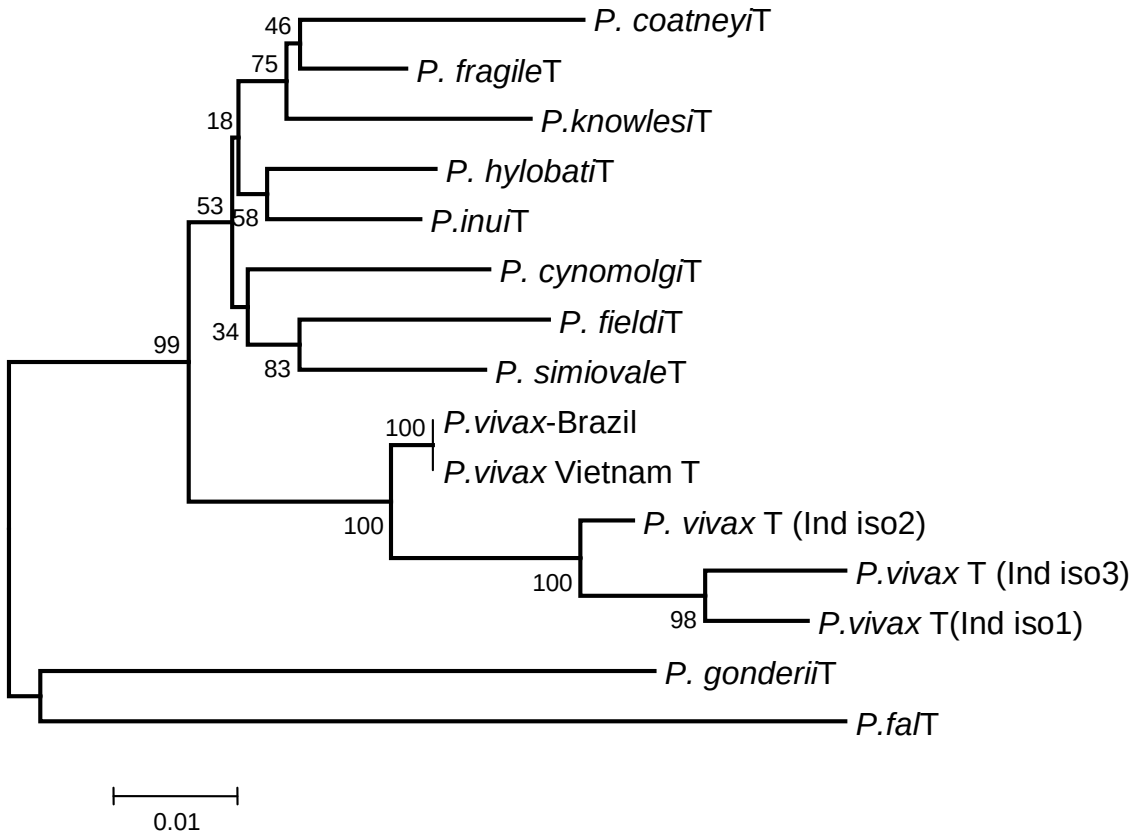

Figure 7.1: Molecular Phylogeny of *tufA* gene based on NJ method

Table 7.1: Distance calculation of Nucleotide Substitution in *tufA* gene using Kimura 2 parameter

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
[1]*P.coatneyi*T															
[2]*P.cynomolgi*T	0.049														
[3]*P.vivax*_T(Indiso3)	0.088	0.079													
[4]*P.vivax*_T(Indiso2)	0.071	0.060	0.029												
[5]*P.vivax*(Indiso1)	0.086	0.077	0.020	0.020											
[6]*P.fieldi*T	0.057	0.045	0.085	0.065	0.082										
[7]*P.fragile*T	0.031	0.035	0.070	0.051	0.068	0.041									
[8]*P.gonderii*T	0.093	0.091	0.123	0.107	0.119	0.086	0.087								
[9]*P.hylobati*T	0.042	0.035	0.074	0.057	0.072	0.042	0.030	0.083							
[10]*P.inui*T	0.038	0.034	0.070	0.053	0.068	0.042	0.031	0.086	0.026						
[11]*P.vivax*-Brazile	0.052	0.042	0.041	0.020	0.039	0.047	0.034	0.091	0.041	0.036					
[12]*P.knowlesi*T	0.043	0.044	0.079	0.062	0.078	0.051	0.030	0.092	0.041	0.041	0.048				
[13]*P.simiovale*T	0.046	0.038	0.080	0.060	0.078	0.035	0.034	0.085	0.038	0.039	0.041	0.043			
[14]*P.fal*T	0.115	0.112	0.122	0.119	0.115	0.111	0.101	0.115	0.101	0.106	0.106	0.111	0.111		
[15]*P.vivax*_Viet	0.052	0.042	0.041	0.020	0.039	0.047	0.034	0.091	0.041	0.036	0.000	0.048	0.041	0.106	

Table 7.2: Distance calculation of Synonymous Substitution in *tufA* gene using Nei - Gojobori model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
[1]*P.coatneyi*T															
[2]*P.cynomolgi*T	0.068														
[3]*P.vivax*_T(Indiso3)	0.099	0.097													
[4]*P.vivax*_T(Indiso2)	0.084	0.083	0.014												
[5]*P.vivax*(Indiso1)	0.089	0.088	0.010	0.005											
[6]*P.fieldi*T	0.075	0.073	0.096	0.082	0.086										
[7]*P.fragile*T	0.051	0.070	0.096	0.081	0.086	0.060									
[8]*P.gonderii*T	0.126	0.126	0.161	0.146	0.153	0.118	0.140								
[9]*P.hylobati*T	0.051	0.051	0.080	0.065	0.070	0.044	0.039	0.120							
[10]*P.inui*T	0.058	0.068	0.097	0.082	0.087	0.070	0.065	0.137	0.042						
[11]*P.vivax*-Brazile	0.070	0.068	0.038	0.024	0.029	0.070	0.067	0.135	0.051	0.068					
[12]*P.knowlesi*T	0.081	0.075	0.101	0.086	0.091	0.075	0.063	0.152	0.058	0.080	0.072				
[13]*P.simiovale*T	0.065	0.061	0.099	0.085	0.089	0.056	0.051	0.126	0.044	0.071	0.070	0.065			
[14]*P.fal*T	0.192	0.170	0.172	0.172	0.167	0.179	0.184	0.195	0.162	0.179	0.174	0.192	0.164		
[15]*P.vivax*_Viet	0.070	0.068	0.038	0.024	0.029	0.070	0.067	0.135	0.051	0.068	0.000	0.072	0.070	0.174	

Molecular clock was not introduced in the phylogeny study. From the obtained tree it was evident that *P. falciparum* is present away from the cluster of primate malaria parasites and from the cluster of all *P. vivax* isolates including Indian, Salvador and Vietnam strains. All the *P. vivax* sequences have formed a cluster and the Indian *P. vivax* isolates are present at the extended branch. The complete *P. vivax* cluster is present in the clade of primate malaria parasites. *P. gonderii* and *P. falciparum* are present at the base of the tree as separate clade representing sister taxa and earliest appearance of the two in parasite family. This is in accordance with the phylogeny established by various workers (Escalante *et. al.*, 1994; Escalante *et. al.*, 1995 a, b; Rathore *et. al.*, 2001; Escalante *et. al.*, 2005) that, *P. vivax* is phylogenetically more close to primate malaria parasites than to *P. falciparum*. The extended branch of Indian *P. vivax* isolates could be considered as divergence of parasite from the rest of the reported *P. vivax* sequences of different regions.

ORF470/ sufB gene

A major ORF present in the *P. falciparum* apicoplast genome has recently been reported as a homologue of *sufB* gene. A blast search using this *sufB* gene isolated from Indian *P. vivax* shows match mainly with ORF's of only 3 – 4 *Plasmodium* species viz, *P. berghei, P. chabaudi, P. yoelii* and *P. falciparum* while a good number of matches were seen with other apicomplexan or bacterial species. Sequences of *Plasmodium* members were obtained and analysed against one complete and three partial sequences of Indian *P. vivax* isolates, in the similar manner as discussed for *tufA*.

An average distance of 0.343 was obtained for nucleotide substitutions using Kimura 2 parameter and an average of 0.261 was obtained for synonymous substitution. A phylogenetic tree (Figure 7.2) was built using NJ method. In the obtained tree, the three Indian *P. vivax* isolates formed a separate cluster but are close to *P. chabaudi* and *P. yoelii* and are away from *P. falciparum* which is not clear. The isolate with complete sequence of *sufB* gene showed closeness with *P. falciparum* with a bootstrap reliability of 100%. This *P. vivax* isolate is also showing slight extended branch as was seen in the *tufA* phylogeny data, though the branch is small enough to

conclude any thing. The separate positions of these four Indian *P. vivax* isolates may be clarified only with availability of more sequences from Indian isolates. Also, a better matrix design is required to understand this variation.

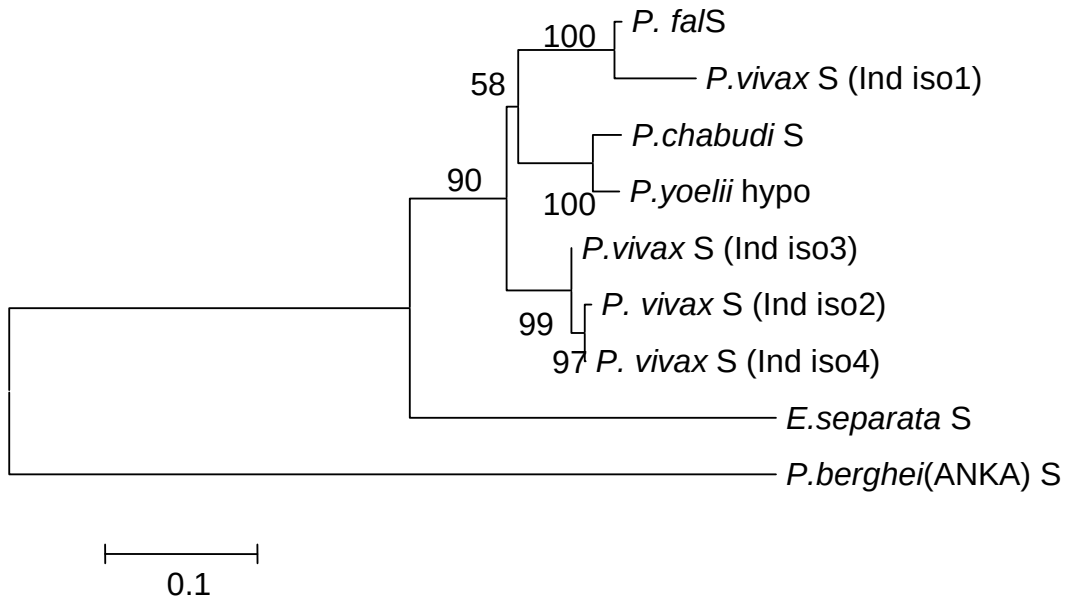

Figure 7.2: Phylogeny of *P. vivax sufB* gene using NJ method

Hypothetical protein of *P. yoelii* and *ORF470* gene of *P. chabaudi* are present in the same cluster. *P. berghei* has formed a separate branch which may indicate the early appearance of the parasite during evolutionary development. Due to the lack of more data on this gene, proper conclusion related to the grouping or origin of *P. vivax* cannot be drawn.

clpC gene

Phylogeny using partial region (641bp) of *clpC* gene has been reported (Rathore *et. al.*, 2001), which follows the established pattern stating the close relation of *P. falciparum* and *P. reichenowi* and common clade of *P. vivax* with *P. cynomolgi* and other primate parasites. Using the available sequences of *clpC* gene from the database, phylogeny was tried against Indian *P. vivax* isolate sequence. The average distance of nucleotide substitution was 0.130 while for synonymous substitution it was 0.089.

The phylogeny tree was built (Figure 7.3) and analysed. In the obtained tree, the Indian *P. vivax clpC* isolate has shown closeness with that of *P. falciparum*, while the Salvador *P. vivax* sequence is still showing presence in the clade of *P. cynomolgi* and *P. knowlesi* as per the established facts by various workers.

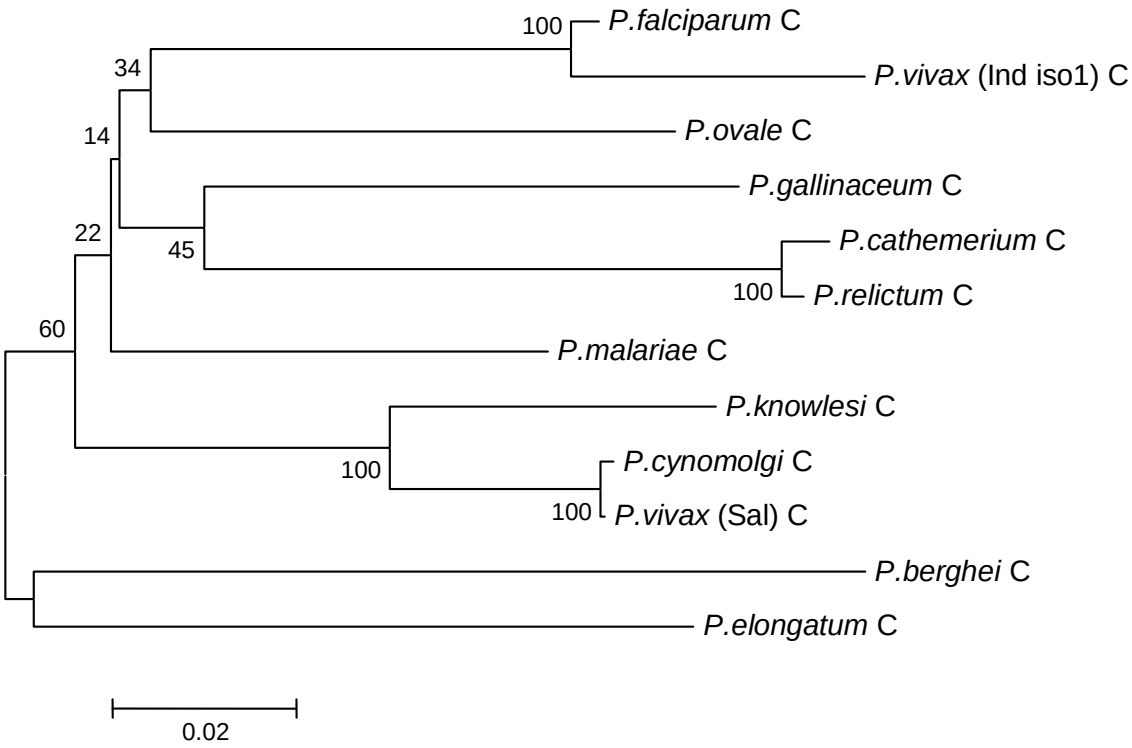

Figure 7.3: Phylogeny tree of *clpC* gene using NJ method

A similar relation between the *ORF470* genes of the same Indian *P. vivax* isolate and *P. falciparum* sequence was seen (Figure 7.2). This could mean that the *sufB* gene and the *clpC* genes have remain conserved throughout the evolutionary changes. While the divergence shown by the *tufA* gene (Figure 7.1) is due to changes in the gene which could have happened due to codon bias in *P. vivax*. Again in the tree, *P. berghei* and *P. elongatum* are appearing in the separate branch showing their early appearance.

Mitochondrial Genes

To compare and verify our plastid-based tree results with *Plasmodium* trees from other sources, we constructed two parallel trees using sequences that were of mitochondrial (*cdc2* and *Cytochrome b*) origin. Primers for these genes were obtained from Ananias Escalante, Arizona State University, Arizona, US. The genes were amplified and sequenced from the same *P. vivax* isolates as plastid DNA genes.

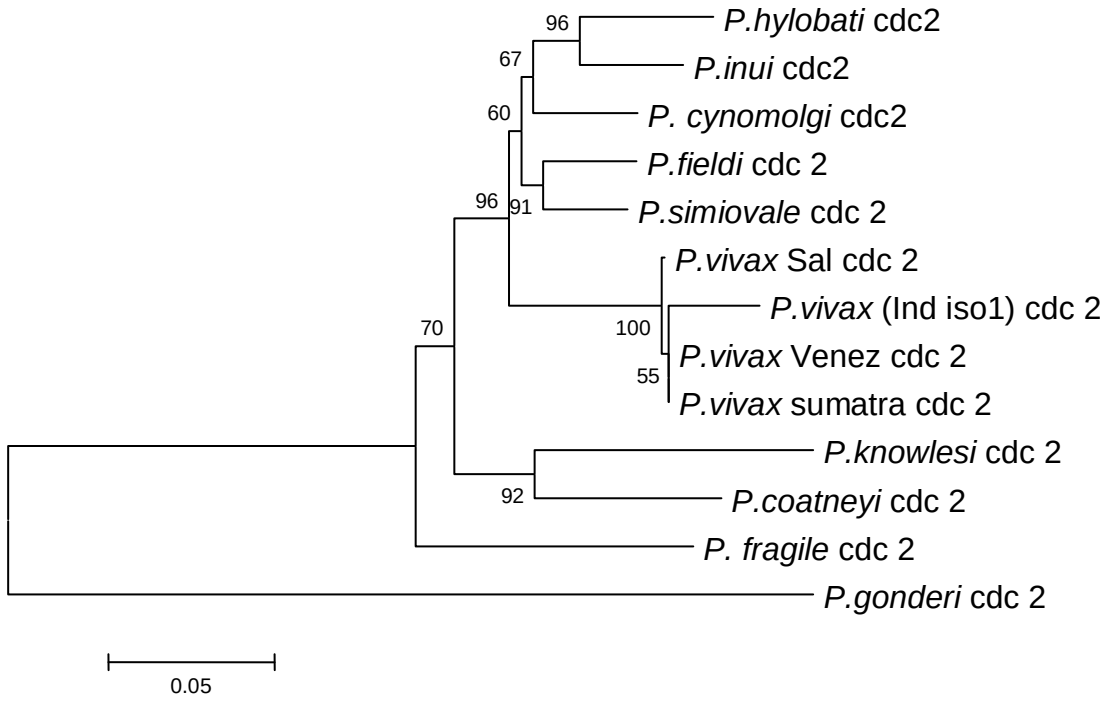

Figure 7.4: Phylogeny of *cdc2* mitochondrial gene using NJ method

Same species of *Plasmodium* were used for both the trees with exception of unavailability of *P. falciparum*, *P. malariae* and *P. ovale* sequences for the *cdc2* gene (Figure 7.4). The trees were obtained for the two genes using NJ method. *P. falciparum* has formed a common clade at the base of *Cytob* tree (Figure 7.5) with *P. reichenowi,* indicating towards its very first divergence. Both the obtained trees show a common cluster of all *P. vivax* genes from different regions, in the clade of *P. cynomolgi, P. hylobati, P. inui, P. simium, P. fieldi* and *P. simiovale*. This may indicate towards the previously stated hypothesis by other workers about the

divergence of *P. vivax* parasite from others during a similar period. The other features noted in these trees from mitochondrial genes are divergence of *P. malariae* and *P. ovale* well before the *P. vivax* clade.

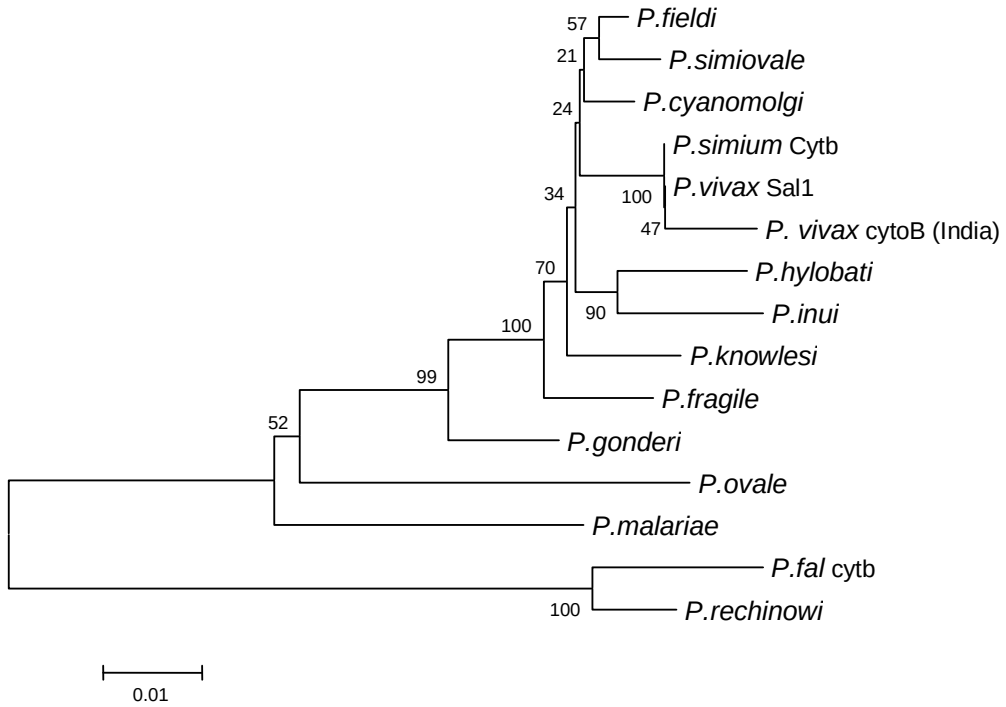

Figure 7.5: Phylogeny of mitochondrial gene *Cytochrome b* using NJ method

All the obtained results indicate towards a divergence of Indian *P. vivax* as evident from the extended branch in all the trees. The *P. vivax* plastid *sufB* and *clpC* genes seem to be conserved with those of *P. falciparum* genes. The *P. vivax* plastid *tufA* gene has shown high divergence.

This is a preliminary study and requires further refinements using more sequences for analysis and better tools for tree building. The study gives a little insight into changes that may be taking place in the genes of *P. vivax* Apicoplast genome. These changes may be due to the codon bias problem which varies from species to species. There may be changes occurring within the isolates of the same species as seen with the three isolates in *sufB* gene analysis. Without the availability of more data it will be too early to make out any strong conclusion from this data.