

Mitochondrial genome sequence diversity of Indian *Plasmodium falciparum* isolates

Suchi Tyagi¹, Veena Pande², Aparup Das^{1/+}

¹Evolutionary Genomics and Bioinformatics Laboratory, Division of Genomics and Bioinformatics, National Institute of Malaria Research, New Delhi, India ²Department of Biotechnology, Kumaun University, Nainital, Uttarakhand, India

We have analysed the whole mitochondrial (mt) genome sequences (each ~6 kilo nucleotide base pairs in length) of four field isolates of the malaria parasite Plasmodium falciparum collected from different locations in India. Comparative genomic analyses of mt genome sequences revealed three novel India-specific single nucleotide polymorphisms. In general, high mt genome diversity was found in Indian P. falciparum, at a level comparable to African isolates. A population phylogenetic tree placed the presently sequenced Indian P. falciparum with the global isolates, while a previously sequenced Indian isolate was an outlier. Although this preliminary study is limited to a few numbers of isolates, the data have provided fundamental evidence of the mt genome diversity and evolutionary relationships of Indian P. falciparum with that of global isolates.

Key words: malaria - *Plasmodium falciparum* - mitochondrial genome - evolution - India

Analyses of whole mitochondrial (*mt*) genome sequences have proven to be highly relevant in inferring both inter and intra-specific evolutionary histories of many model and non-model organisms (Hedges 2000, Hurst & Jiggins 2005, Haag-Liautard et al. 2008) including the malaria parasites which have shown promising evolutionary outcomes (Joy et al. 2003, 2006, Mu et al. 2005). For example, analyses of whole *mt* genomes in the two most common human malaria parasites, *Plasmodium falciparum* and *Plasmodium vivax*, have provided evidence of the origin and evolutionary histories of these species (Joy et al. 2003, 2006, Mu et al. 2005). In addition, useful inferences on the probable host-switching mechanism of *P. falciparum*, although much debated (Prugnolle et al. 2011), have been revealed to a certain extent.

Malaria is highly endemic in India, with *P. falciparum* causing malaria havoc (Singh et al. 2009). However, the *mt* genome sequence of Indian *P. falciparum* has not yet been analysed and compared to global *mt* genome sequences of this species. Therefore, the evolutionary history of global *P. falciparum* isolates remains incomplete without Indian *mt* genome sequence data, although two *mt* genome sequences of Indian origin are available in the public domain (Sharma et al. 2001, Tyagi et al. 2014). In order to fill this gap, we have utilised whole *mt* genomes of four *P. falciparum* field isolates [3 new to this study and 1 previously reported (Tyagi et al. 2014)] from low, mild and high endemic localities of India and compared them with the global data. We have also included the already published *mt* genome sequence data of a single *P.*

falciparum isolate of unknown Indian origin (Sharma et al. 2001) in this study to obtain first-hand information on the extant diversity of Indian *P. falciparum* in comparison to global isolates.

We collected finger-prick blood samples from *P. falciparum*-infected malaria patients attending malaria clinics in three geographically distant localities of India with variable *P. falciparum* malaria epidemiology: Betul (state of Madhya Pradesh, high endemic), Goa (state of Goa, low endemic) and Mangalore (state of Karnataka, mild endemic). The samples were spotted (2, 3 spots) on Whatman filter paper, dried and transported to the laboratory in New Delhi. While the sample from Betul was collected by us, the samples from Goa and Mangalore were obtained from the parasite bank of the National Institute of Malaria Research (NIMR). The fourth sample came from Bilaspur [state of Chhattisgarh, high endemic (Tyagi et al. 2014)] (Fig. 1). This study was approved by NIMR, New Delhi, India, and written informed consents were obtained from all patients who participated in the study before the samples were collected. The locations of the sample collection sites in India are shown in Fig. 1.

For each collected sample, genomic DNA was isolated using the QIAamp mini DNA isolation kit (Qiagen, Germany) according to the manufacturer's instructions. Because both *P. falciparum* and *P. vivax* occur in India in almost equal proportion (Singh et al. 2009) and the *mt* genome is conserved among different species of human malaria parasites (Hikosaka et al. 2011), we first tested each sample for incidences of mixed malaria parasite infections using nested polymerase chain reaction (PCR) detection assays with genus and species-specific primers based on the 18S rRNA gene (Gupta et al. 2010). In order to PCR amplify and sequence the whole *mt* genome of 5,967 nucleotide base pairs, we used 19 recently designed primer pairs [for details of the primer sequences and PCR conditions, see Tyagi et al. (2014)]. For each individual DNA fragment, the PCR protocols suggested by Tyagi et al. (2014) were followed. To avoid

doi: 10.1590/0074-0276130531

Financial support: DBT, ICMR, Government of India

+ Corresponding author: aparup@mrcindia.org

Received 20 May 2013

Accepted 11 March 2014



Fig. 1: map of India showing the location of collection of samples of *Plasmodium falciparum* isolates.

infections with multiple clones of *P. falciparum* in a single patient, we sequenced all 19 DNA fragments of the *mt* genome from both directions (2X coverage, see below) and excluded isolates showing double peaks in any single nucleotide position (Gupta et al. 2012). For each single-clone *P. falciparum* isolate, the 19 PCR-amplified products were separately purified with exonuclease and shrimp alkaline-phosphatase following standard protocols (Tyagi et al. 2014) and DNA sequencing was performed at an NIMR in-house facility following Sanger sequencing technology using an ABI 3730XL DNA Analyser (Applied Biosystems). For each *P. falciparum* isolate, all 19 DNA fragments downstream of the sequencer were individually edited using the SeqMan and EditSeq modules of the Lasergene v.7 computer program (DNASTAR, Madison, USA), respectively and the sequences were manually assembled to form one complete *mt* genome. The newly generated whole *mt* genome sequences of the three isolates were submitted to GenBank with accessions KJ418723, KJ418724 and KJ418725.

To estimate *mt* genome diversity in Indian *P. falciparum* and infer evolutionary patterns by comparing to global isolates, we have estimated several population genetic parameters. Specifically, the number of segregating sites, number of haplotypes, haplotype diversity (H_d) and a measure of nucleotide diversity (π) (Nei 1987), were calculated and compared with data from other parts of the globe using the DnaSP v.5 computer program (Librado & Rozas 2009). Furthermore, to perform comparative and evolutionary studies of global *P. falciparum* isolates using whole *mt* genome sequences, five Indian whole *mt* genome sequences were analysed [of which 3 were newly generated in the present study and 2, Blsp1 (Tyagi et al. 2014) (GenBank accession KJ144901) and PfPH10 (Sharma et al. 2001) (GenBank accession AJ298788) were previously reported]. In these analyses, one consensus sequence from each continent [Africa ($n = 29$), South America ($n = 28$) and Asia ($n = 29$)] and

country [Papua New Guinea (PNG) ($n = 10$)] (Joy et al. 2003) were generated by alignment of the whole *mt* genome sequences from the respective places using the MegAlign module of Lasergene v.7 computer program (DNASTAR, Madison, USA). The 100 *mt* genome sequences of *P. falciparum* utilised here for comparative and evolutionary studies bear the GenBank accessions AY282924–AY283019 and AJ276844–AJ276847. The whole *mt* genome sequence of the reference 3D7 isolate (GenBank accession AY282930) was also included in this study. Therefore, 10 sequences were utilised in this study [5 Indian, 1 African (consensus), 1 South American (consensus), 1 PNG (consensus), 1 Asian (consensus) and the single 3D7 strain]. All 10 sequences were aligned following the CLUSTALW algorithm and a phylogenetic tree was constructed using the neighbour-joining (NJ) method implemented in MEGA v.5 computer program (Tamura et al. 2011). In order to estimate the strength of each internal node of the NJ phylogenetic tree, the tree topologies were simulated 1,000 times.

Following the PCR diagnostic approach to detect malaria parasites and based on the peaks of the DNA sequence chromatogram, we were able to isolate three pure single-clone *P. falciparum* isolates (Bet12, Goa2 and Mang2) and successfully sequence their complete *mt* genomes. Because complete *mt* genome sequences from two other Indian isolates (Blsp1 and PfPH10) were available in the public domain (National Center for Biotechnology Information, available from: ncbi.nlm.nih.gov/), we also utilised the data from these two isolates (altogether 5 Indian *mt* genome sequences) to estimate *mt* genome sequence diversity and to infer first-hand evolutionary relationships of Indian *P. falciparum* with that of global isolates. The whole *mt* genome sequence alignment of five Indian isolates (Bet12, Goa2, Mang2, Blsp1 and PfPH10) with the reference sequence from the 3D7 strain revealed 26 variable nucleotide sites in Indian *P. falciparum* (Table I). Considering that the *mt* genome is fairly conserved across populations and species in *Plasmodium* (Hikosaka et al. 2011), the observed high incidences of nucleotide substitutions in Indian isolates was quite surprising. However, a closer look at the alignment revealed that as many as 21 unique nucleotide substitutions were present in the PfPH10 isolate (Table I). A very similar pattern was also observed when the PfPH10 isolate was compared with the Blsp1 and 3D7 isolates (Tyagi et al. 2014). Because there were such unusual patterns of single nucleotide mutations in the PfPH10 isolate, which might bias the overall outcome of the *mt* genome diversity, we did not consider the data of PfPH10 in further population genetic analyses. Therefore, we restricted the dataset to four *mt* genome sequences of Indian *P. falciparum*. Multiple alignments of *mt* genome sequences of four Indian *P. falciparum* isolates (Bet12, Goa2, Mang2 and Blsp1) with the reference 3D7 isolate revealed five nucleotide substitutions (Table II). Out of these five single nucleotide polymorphisms (SNPs), four were found in the four Indian isolates and one in the 3D7 isolate. Out of the four SNPs found in Indian isolates, two were present in the intergenic regions and two were in the *cox I* gene (Table II). The two SNPs present in

TABLE I
Nucleotide sequence alignment of the whole mitochondrial genomes of *Plasmodium falciparum* isolates from different malaria endemic localities of India

Isolates	Nucleotide positions																								
	<i>cox III</i>					<i>cox I</i>					<i>cyt b</i>														
	208	222	230	276	510	615	725	1122	1339	2175	2763	3330	3444	3764	3766	3868	3985	4352	4353	4420	4640	4759	4952	5485	
3D7 ^a	G	A	G	G	A	C	C	G	A	T	C	T	A	T	T	A	A	C	A	T	T	A	T	T	T
Blsp1 ^b
Bet12 ^c	.	.	.	A	.	.	T	.	.	.	T	C	.
Goa2 ^c	C	C	.
Mang2 ^c	C	C	.
PfPH10 ^d	A	T	A	.	G	T	.	C	G	C	.	A	G	A	C	C	G	T	G	T	A	A	G	C	A

a: Joy et al. (2003); b: Tyagi et al. (2014); c: present paper; d: Sharma et al. (2001). The sequences of two earlier reported Indian sequences (Blsp1 and PfPH10) and the reference sequence 3D7 are also included in the alignment. Sites showing nucleotide variations in the alignment were shown.

TABLE II

Nucleotide sequence alignment of the whole mitochondrial genome sequence of the four Indian *Plasmodium falciparum* isolates (excluding the previously reported PfPH10 isolate) with the reference sequence of the 3D7 isolate

Isolates	Nucleotide positions				
	276	725	2175	2763	4952
3D7 ^a	G	C	T	C	T
Blsp1 ^b	C
Bet12 ^c	A	T	.	T	C
Goa2 ^c	.	.	C	.	C
Mang2 ^c	.	.	C	.	C

a: Joy et al. (2003); b: Tyagi et al. (2014); c: present paper; Only sites showing nucleotide variations in the alignment were shown. To be noted that due to occurrence of four single nucleotide polymorphisms in the four Indian isolates, three haplotypes are formed: GCTC (Blsp1), ATTT (Bet12) and GCCC (Goa2 and Mang2).

the *cox I* gene, however, were found to be synonymous. Further multiple sequence alignment of the four Indian *mt* genome sequences with the 100 published sequences (Conway et al. 2000, Joy et al. 2003) indicated that three out of the four SNPs found in Indian *P. falciparum* (at positions 276, 725 and 2763) were novel and India-specific (data not shown). The fourth SNP (at the 2175th position) seems to be Asian-specific in nature, as this SNP was never detected in any global samples and was only detected in some Asian isolates (Joy et al. 2003). In comparison to the reference 3D7, one nucleotide substitution was detected in the Blsp1 isolate and two were found in each of the Goa2 and Mang2 isolates (Table II). The Bet12 isolate, however, contained four nucleotide substitutions, indicating that this isolate is highly diverged from the rest of the three Indian isolates (Table II). The four SNPs segregating in the four Indian isolates produced three haplotypes: GCTC, ATTT and GCCC (Table II). While the GCCC haplotype (Goa2 and Mang2) was segregated in 50% frequency and found in *P. falciparum* isolates from the low and mild-malaria endemic localities (Goa and Mangalore), the other two haplotypes (ATTT and GCTC) were found to be unique, each with 25% frequency in *mt* genome sequences from the two highly endemic localities (Betul and Bilaspur) (Table II). We have further estimated the values of Hd and π in four Indian *P. falciparum* isolates to be 0.833 and 0.00036, respectively. These estimates were higher than similar estimates in the whole *mt* genome sequences of *P. falciparum* isolates sampled in Asia and South America, but comparable to samples from Africa (Hd = 0.865; π = 0.00025) (Joy et al. 2006). It is notable that *P. falciparum* malaria is not only highly endemic in Africa, but based on the high genetic diversity of isolates from this continent, Africa is considered to be the homeland of this species (Conway et al. 2000, Joy et al. 2003). Although the present Indian data on whole *mt* genome sequences are

limited to only four isolates, the high *mt* genome diversity observed was comparable to that of Africa. These results corroborate earlier observations on high genetic diversity in different genes in Indian *P. falciparum* (Singh et al. 2009) and therefore justify further in-depth analyses with a larger number of isolates from different eco-climatic and malaria endemic localities in India.

With the whole *mt* genome sequence information of Indian *P. falciparum*, we were interested to know the evolutionary relationships of global isolates using phylogenetic approaches. To do this, we used single consensus sequences of whole *mt* genomes from Africa, Asia, South America, PNG, the reference 3D7 and five *mt* genomes of Indian *P. falciparum* (Bet12, Goa2, Mang2, Blsp1 and PfPH10). The NJ phylogenetic tree (Fig. 2) showed congregation of global *P. falciparum* isolates (Africa, Asia, South America, PNG, 3D7) with two Indian sequences (Bet12 and Blsp1) to form a large single clade with high bootstrapped values of the internal node (Fig. 2). However, two other Indian isolates (Goa2 and Mang2) were found in a small separate clade with very weak bootstrapped values (Fig. 2). This might be because both Goa2 and Mang2 isolates bear a similar haplotype (GCCC) (Table II) that is different from the other two isolates (Bet12 and Blsp1). Moreover, Goa and Mangalore are from the south-western part of India and are in close geographic proximity to each other (Fig. 1). The observed genetic similarities between the two geographically close *P. falciparum* populations can be explained by the fact that *P. falciparum* populations often follow the isolation-by-distance (IBD) model of population structure (Tanabe et al. 2010) and are also maintained as genetically sub-structured populations (Anderson et al. 2000). Whether Indian *P. falciparum* populations follow the IBD and/or genetically sub-structured models of population structure remains to be seen with a larger sample size of *P. falciparum* isolates collected from a wide geographic distribution within India.

The distant placement of the PfPH10 isolate from the other *P. falciparum* isolates in the phylogenetic tree (Fig. 2) was not surprising considering the fact that this sequence bears a large number of mutations (21 nucleotide substitutions) (Table I). This corroborates earlier observations on the high sequence divergence of PfPH10 from the Blsp1 and the 3D7 isolates (Tyagi et al. 2014). Due to the unknown origin within India of the PfPH10 clinical isolate (Sharma et al. 2001) we are not in a position to discuss the observed high *mt* genomic differentiation among the Indian isolates, the reference 3D7 isolate and global isolates of *P. falciparum*. The placement of all global isolates in a single cluster supports the hypothesis of a comparatively lower polymorphic nature of *mt* genomes in relation to other recombined genomes (e.g., nuclear genomes) and reasserts the idealistic nature of *mt* genomes for evolutionary inferences (Conway et al. 2000). In general, the four Indian *P. falciparum* isolates possess an appreciable level of nucleotide diversity (as measured by π) that corresponds to the African isolates. The observed results, although with a very limited sample size, provide sufficient background information for future studies, including the sequencing more Indian

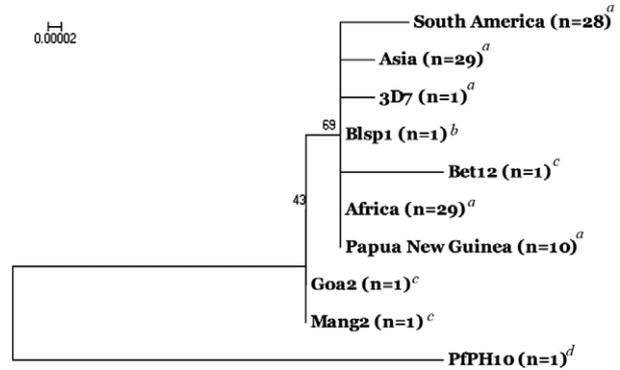


Fig. 2: neighbor-joining phylogenetic tree based on whole mitochondrial genome sequence alignment showing genetic interrelationships among *Plasmodium falciparum* isolates from different endemic localities of the globe. a: Joy et al. (2003); b: Tyagi et al. (2014); c: present paper; d: Sharma et al. (2001).

P. falciparum isolates and performing associated comparative and evolutionary genomic studies. Such studies will help to understand the population structure and demography of Indian *P. falciparum* isolates and revisit the evolutionary history of global *P. falciparum*. Most importantly, such studies will help discover whether the antimalarial drug atovaquone could be beneficial malaria chemotherapy in India. This is because the *cyt b* gene present in the *P. falciparum* mitochondrion is considered to be the target of atovaquone (Vaidya et al. 1993, Biagini et al. 2006) and low or no diversity in the *cyt b* gene in populations would be the ideal condition for using atovaquone to treat *P. falciparum* malaria.

ACKNOWLEDGEMENTS

To Prof AP Dash, former director of the NIMR, for motivation and encouragements, to Dr N Valecha, present director, for facilities, to Drs Anup Anvikar and Vineeta Singh, for help in *P. falciparum* isolates from the malaria parasite bank of NIMR, and to the two anonymous reviewers, for critical and constructive comments on an earlier version of the paper.

REFERENCES

- Anderson TJC, Haubold B, Williams JT 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* 17: 1467-1482.
- Biagini GA, Viriyavejakul P, O'Neill PM, Bray PG, Ward SA 2006. Functional characterization and target validation of alternative complex I of *Plasmodium falciparum* mitochondria. *Antimicrob Agents Chemother* 50: 1841-1851.
- Conway DJ, Fanello C, Lloyd JM, Al-Joubori BM, Baloch AH, Somanath SD, Roper C, Oduola AM, Mulder B, Povoa MM, Singh B, Thomas AW 2000. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol Biochem Parasitol* 111: 163-171.
- Gupta B, Gupta P, Sharma A, Singh V, Dash AP, Das A 2010. High proportion of mixed species *Plasmodium* infections in India revealed by PCR diagnostic assay. *Trop Med Int Health* 15: 819-824.
- Gupta B, Srivastava N, Das A 2012. Inferring the evolutionary history of Indian *Plasmodium vivax* from population genetic analyses of multilocus nuclear DNA fragments. *Mol Ecol* 21: 1597-1616.

- Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol* 6: e204.
- Hedges SB 2000. Human evolution: a start for population genomics. *Nature* 408: 652-653.
- Hikosaka K, Watanabe Y-I, Kobayashi F, Waki S, Kita K, Tanabe K 2011. Highly conserved gene arrangement of the mitochondrial genomes of 23 *Plasmodium* species. *Parasitol Int* 60: 175-180.
- Hurst GD, Jiggins FM 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc Lond B Biol Sci* 272: 1525-1534.
- Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su X-z 2003. Early origin and recent expansion of *Plasmodium falciparum*. *Science* 300: 318-321.
- Joy DA, Mu J, Jiang H, Su X 2006. Genetic diversity and population history of *Plasmodium falciparum* and *Plasmodium vivax*. *Parassitologia* 48: 561-566.
- Librado P, Rozas J 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
- Mu J, Joy DA, Duan J, Huang Y, Carlton J, Walker J, Barnwell J, Beerli P, Charleston MA, Pybus OG, Su X-z 2005. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol Biol Evol* 22: 1686-1693.
- Nei M 1987. *Molecular evolutionary genetics*, Columbia University Press, New York, 512 pp.
- Prugnolle F, Durand P, Ollomo B, Duval L, Arriey F, Arnathau C, Gonzalez J-P, Leroy E, Renaud F 2011. A fresh look at the origin of *Plasmodium falciparum*, the most malignant malaria agent. *PLoS Pathog* 7: e1001283.
- Sharma I, Rawat DS, Pasha ST, Biswas S, Sharma YD 2001. Complete nucleotide sequence of the 6 kb element and conserved cytochrome b gene sequences among Indian isolates of *Plasmodium falciparum*. *Int J Parasitol* 31: 1107-1113.
- Singh V, Mishra N, Awasthi G, Dash AP, Das A 2009. Why is it important to study malaria epidemiology in India? *Trends Parasitol* 25: 452-457.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
- Tanabe K, Mita T, Jombart T, Eriksson A, Horibe S, Palacpac N, Ranford-Cartwright L, Sawai H, Sakihama N, Ohmae H, Nakamura M, Ferreira MU, Escalante AA, Prugnolle F, Björkman A, Färnert A, Kaneko A, Horii T, Manica A, Kishino H, Balloux F 2010. *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr Biol* 20: 1283-1289.
- Tyagi S, Pande V, Das A 2014. Whole mitochondrial genome sequence of an Indian *Plasmodium falciparum* field isolate. *Korean J Parasitol* 52: 99-103.
- Vaidya AB, Lashgari MS, Pologe LG, Morrisey J 1993. Structural features of *Plasmodium* cytochrome b that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones. *Mol Biochem Parasitol* 58: 33-42.