Research

**Genetic structure of *Plasmodium falciparum* field isolates in eastern and north-eastern India**

Hema Joshi*1, Neena Valecha1, Anju Verma1, Asha Kaul1, Prashant K Mallick1, Sneh Shalini1, Surendra K Prajapati1, Surya K Sharma2, Vas Dev3, Sukla Biswas1, Nutan Nanda1, MS Malhotra1, Sarala K Subbarao1,4 and Aditya P Dash1

Address: 1National Institute of Malaria Research (ICMR), 22-Sham Nath Marg, Delhi-110 054, India, 2National Institute of Malaria Research, Field Unit IDVC, Sector 5, Rourkela-769 002, Orissa, India, 3National Institute of Malaria Research, Field Unit IDVC, PO Sonapur, District Kamrup, Assam-782 402, India and 4Indian Council of Medical Research, V. Ramalingaswami Bhawan, Ansari Nagar, New Delhi 110029, India

Email: Hema Joshi* - hema_joshi_mrc@yahoo.com; Neena Valecha - neenavalecha@gmail.com; Anju Verma - anju_mverma@yahoo.com; Asha Kaul - kaulasha@gmail.com; Prashant K Mallick - pkmmrc@gmail.com; Sneh Shalini - shalu_sneh@yahoo.co.in; Surendra K Prajapati - surendramrc@gmail.com; Surya K Sharma - suryaksarma@gmail.com; Vas Dev - nimr.sonapur@gmail.com; Sukla Biswas - suklabiswas@yahoo.com; Nutan Nanda - nutannmc@yahoo.co.in; MS Malhotra - msmalhotra@gmail.com; Sarala K Subbarao - saralaks@yahoo.com; Aditya P Dash - director@mrcindia.org

* Corresponding author

**Abstract**

Background: Molecular techniques have facilitated the studies on genetic diversity of *Plasmodium* species particularly from field isolates collected directly from patients. The *msp-1* and *msp-2* are highly polymorphic markers and the large allelic polymorphism has been reported in the block 2 of the *msp-1* gene and the central repetitive domain (block 3) of the *msp-2* gene. Families differing in nucleotide sequences and in number of repetitive sequences (length variation) were used for genotyping purposes. As limited reports are available on the genetic diversity existing among *P. falciparum* population of India, this report evaluates the extent of genetic diversity in the field isolates of *P. falciparum* in eastern and north-eastern regions of India.

Methods: A study was designed to assess the diversity of *msp-1* and *msp-2* among the field isolates from India using allele specific nested PCR assays and sequence analysis. Field isolates were collected from five sites distributed in three states namely, Assam, West Bengal and Orissa.

Results: *P. falciparum* isolates of the study sites are highly diverse in respect of length as well as sequence motifs with prevalence of all the reported allelic families of *msp-1* and *msp-2*. Prevalence of identical allelic composition as well as high level of sequence identity of alleles suggest a considerable amount of gene flow within the *P. falciparum* populations of different states. A comparatively higher proportion of multiclonal isolates as well as multiplicity of infection (MOI) was observed among isolates of highly malarious districts Karbi Anglong (Assam) and Sundergarh (Orissa). In all the five sites, R033 family of *msp-1* was observed to be monomorphic with an allele size of 150/160 bp. The observed 80–90% sequence identity of Indian isolates with data of other regions suggests that Indian *P. falciparum* population is a mixture of different strains.

Conclusion: The present study shows that the field isolates of eastern and north-eastern regions of India are highly diverse in respect of *msp-1* (block 2) and *msp-2* (central repeat region, block 3). As expected Indian isolates present a picture of diversity closer to southeast Asia, Papua New Guinea and Latin American countries, regions with low to meso-endemicity of malaria in comparison to African regions of hyper- to holo-endemicity.
Background

Information on the nature and extent of genetic diversity within *Plasmodium falciparum* is essential in understanding the mechanism underlying the pathology of malaria, the acquisition of immunity, the spread of drug resistance and the condition of transmission. Molecular techniques have facilitated the studies on genetic diversity of *Plasmodium* species particularly from field isolates collected directly from patients. Polymerase chain reaction assay has been specially a very useful tool in epidemiological studies. Polymorphic markers could be identified from genomic DNA isolated from small quantities of blood spotted on filter papers. Studies on genetic diversity, the differentiation of different strains within a *Plasmodium* species, presence of multiple parasite strains/types in individual host etc. are reported from different regions of the globe [1-13]. However, limited reports are available on the genetic diversity existing among *P. falciparum* population of India [14-19]. This paper reports polymorphism observed in merozoite surface protein genes (*msp-1* and *msp-2*) among field isolates of *P. falciparum* collected from north-eastern and eastern regions of India.

Two highly polymorphic and widely used markers are *msp-1* and *msp-2* and the large allelic polymorphism has been reported in the block 2 of the *msp-1* gene and the central repetitive domain (block3) of the *msp-2* gene. Families differing in nucleotide sequences and in number of repetitive sequences (length variation) were used for genotyping purposes. In *msp-1* (block 2) three distinct allelic families have been described: K1, MAD20 and RO33 while *msp-2* has two distinct families, 3D7 (Indo-China) and FC27 [12,20,21].

Methods

*P. falciparum* isolates were collected from microscopically diagnosed *P. falciparum* positive subjects in three states with varied malaria epidemiology (Figure 1). Blood was spotted on filter paper strips (Whatman 3 mm) by pricking a finger after obtaining the consent of patient or guardian in case of children. This study has approval of the ethical committee of the National Institute of Malaria Research.

Study areas

**Assam**

The Assam state contributes 64% of the malaria positive cases and 75% of the *P. falciparum* cases in the north eastern region. The area is highly endemic for malaria and is known for its persistent transmission [22], with the prevalence of multi drug resistant strains. The area is dominated by mongoloid tribes namely Khasis, Bodos and Mikir (Karbi). Samples studied were from two endemic districts Karbi Anglong and Kamrup.

**West Bengal**

An eastern state accounting for about 10% of the total malaria cases in the country and areas along international borders pose a serious problem. The study district Darjeeling, is dominated by Gorkhas (Nepali) population, however, labourers from adjoining state Bihar also frequent the tea gardens. The transmission is seasonal from May to July and then again September to November. *P. falciparum* proportion is in the range of 50–70.

**Orissa**

State located in eastern plateau, contributes 22% of total malaria cases, 43% of falciparum cases and 50% of all reported deaths due to malaria although it constitutes only 4% of the total population of India [23]. *P. falciparum* is the major cause of malaria and accounts for 80–90% of malaria cases. Samples were collected from two malaria endemic districts namely Sundergarh and Keonjhar. Area is dominated by tribals and predominant tribes being Oram, Khadia, Pradhan and Munda. Malaria transmission in District Sundergarh has been reported to be meso to hyper endemic [24].

Genomic DNA isolation and PCR amplification

Genomic DNA of *P. falciparum* from parasitized blood spots was isolated using QiaAmp DNA minikit as per manufacturer instructions (Qiagen, Hilden, Germany). Primers and PCR protocols were followed as previously described by Snounou et al [11] for family specific allele analysis of *msp-1* (block 2) and *msp-2* (block 3). PCR amplification was performed on thermal cycler (Perkin Elmer 9700/2400, UK) in a final volume of 20μl. The
PCR products were visualized by UV transillumination at 302 nm on gel documentation system (Syngenta, USA) after electrophoresis on 2% agarose gel (Promega/Boehringer) using 0.5 x TBE buffer at 80–100 volts. Allele sizes were calculated using Genetool programme.

A limited number of isolates representing different families of msp-1 and msp-2 were sequenced using the Big Dye Terminator cycle-sequencing kit (Applied Biosystems, Foster City, CA) and the ABI Prism 310 automated DNA sequencer (Applied Biosystems). The sequences were then analysed using the DNASTAR software package (DNASTAR, Madison, WI). Sequence data have been submitted to GenBank vide accession numbers DQ485417 to DQ485451.

To understand the identity of Indian isolates with respect to isolates of other regions, sequence data available in public domains were downloaded for allelic families of msp-1&2 and details are given below; Thailand(K1-M77730, MAD20-M77722, R033-AAA29684, 3D7-U91676), Vietnam(K1-AF509651, MAD20-AF509653&94, FC27-AF104696, 3D7-AF104693), Tanzania(K1-AF061134, FC27-AF532386), Brazil(K1-AF509714, MAD20-AY115485, FC27-DQ115973, 3D7-AF177389), China(MAD20-AF251345), Sudan(MAD20-AF034635), Iran(MAD20-AY138509, R033-AF138507, FC27-DQ38451), Indonesia(K1-AF191061, R033-FAF18431), Western Africa(R033-PFAMSA1), Kenya(R033-AAM21583), Ghana(FC27-AF329577), PNG(FC27-AF329579), Gambia(FC27-U91668, 3D7-U91665) and Nigeria(3D7-AF148224).

**Results**

One hundred and thirty one *P. falciparum* isolates analysed during the study have demonstrated highly diverse nature of field isolates in respect of msp-1 (block 2) and msp-2 (central repeat region, block3).

All the three reported families of msp-1(K1, MAD20 and RO33) and two of msp-2 (FC27 and 3D7), were observed among the isolates of all the five study sites (Table 1). Proportion of isolates with K1 family ranged from 33.3% to 72.7% with 6 alleles in the range of 140 to 280 bp. Proportion of isolates with MAD20 family ranged between 13.6% to 72.7% and a total of six alleles were observed within 120 to 240 bp. RO33 proportions ranged from 15.0% to 41.7% and the family was observed to be monomorphic with an allele size of 150/160 bp. Observed proportions, numbers and size range of alleles among the isolates of different study sites are given in Table 2.

In msp-2, the reported families FC27 and 3D7 were observed among the isolates of all the 5 study sites (Table 1). Proportion of FC27 family varied from 18.2% to 85.0% and that of 3D7 ranged from 57.9% to 82.1%. Proportional prevalence of FC27 and 3D7 families was significantly different between the study sites ($\chi^2 = 16.5$, $P = 0.002$), particularly between Orissa and Assam ($\chi^2 = 10.20$, $P = 0.001$), Orissa and West Bengal ($\chi^2 = 9.65$, $P = 0.002$). It is seen from Table 2 that in Assam and West Bengal isolates, prevalence of 3D7 family was higher than FC27 family and vice versa in Orissa isolates. Six alleles of FC27 in the range of 250–500 bp and 9 alleles of 3D7 in the range of 420–600 bp were observed (Table 2).

Proportion of multiclonal isolates (multiple infection) in different study sites is given in Table 1 and proportion ranged from 13.6% in Darjeeling, West Bengal to 83.3% in Karbi Anglong, Assam. Mproportion of multiclonal isolates, all possible combinations of msp-1 families namely (K1+MAD20, MAD20+R033, K1+R033 and

| Table 1: Distribution of size variants of msp-1 and msp-2 in Indian P. falciparum isolates |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | Assam           | Orissa          | West Bengal     |                 |                 |
|                 | Karbi Anglong   | Kamrup          | Keonjhar        | Sundergarh      | Darjeeling      |
| msp-1           |                 |                 |                 |                 |                 |
| K1             | n = 18          | n = 27          | n = 24          | n = 40          | n = 22          |
| MAD20          | 1               | 8               | 7               | 17              | 12              |
| RO33           | 2               | 5               | 7               | 13              | 1               |
| K1+MAD20       | 0               | 5               | 8               | 4               | 6               |
| K1+RO33        | 10              | 3               | 0               | 4               | 2               |
| MAD20+RO33     | 1               | 3               | 1               | 1               | 1               |
| K1+MAD20+RO33  | 0               | 2               | 1               | 1               | 0               |
| Multiclonal isolates % | 83.3 | 33.3 | 8.3 | 15.0 | 13.6 |
| msp-2           |                 |                 |                 |                 |                 |
| FC27           | n = 22          | n = 28          | n = 19          | n = 40          | n = 22          |
| 3D7            | 4               | 5               | 8               | 9               | 4               |
| FC27+3D7       | 13              | 21              | 6               | 6               | 18              |
| Multiclonal isolates % | 22.7 | 7.1 | 26.3 | 62.5 | 0.0 |
K1+MAD20+RO33) were observed and proportion of K1+MAD20 was highest 54.3%(19/35). Complexity of infection (multiplicity of infection, MOI) was highest (2.05) in Karbi Anglong (Assam) and lowest (1.0) in Darjeeling, West Bengal (Table 2). MOI was estimated by dividing the total number of fragments detected in the individual system by the number of samples positive in the particular system (either *msp-1* or *msp-2*).

Analysis of *msp-1* and *msp-2* sequence data revealed above 80% identity of study isolates among themselves in general and above 70% with isolates of other countries with a few exceptions. In K1 family, 88 to 100% similarity was observed with K1 allelic sequences reported for isolates of Indonesia, Thailand, Vietnam, Tanzania and Brazil. Similar identity was observed in MAD20 family with isolates of China, Brazil, Vietnam, Thailand, Sudan and Iran. RO33 allelic sequences of Indian isolates were all very identical and had shown above 94% similarity with sequences reported for isolates of Thailand, Indonesia, Iran and Western Africa except Kenya (Figure 2a–c). Allelic families of *msp-2*, FC27 and 3D7 showed above 70% identity with isolates of Vietnam, Ghana, PNG, Thailand and Gambia. On the other hand, with Iran and Tanzania isolates, identity ranged between 60–70% for Indian FC27 allele and less than 30% with Nigeria and Brazil isolate for 3D7 sequences of the present study (Figures 3a&b).

**Discussion**

Observed highly diverse nature of *P. falciparum* isolates of eastern and northeastern regions of India in respect of length as well as sequence motifs with prevalence of all the allelic families of *msp-1* and *msp-2* is in agreement with the earlier reports on Indian isolates [16-19,25].

Prevalence of identical allelic composition as well as high level of sequence identity of alleles in five study areas of three states suggest for a considerable amount of gene flow between the *P. falciparum* populations of different states. Labourers from Orissa, Bihar and other parts of West Bengal come to work in tea gardens of Darjeeling (WB) and Assam and their to and fro movement from native place to work place may be facilitating the gene flow between the *P. falciparum* populations of different states.

### Table 2: Observed proportions of various families of *msp-1* & *2*, allele numbers and size range among study isolates.

| Markers | Assam | Orissa | West Bengal |
|---------|-------|--------|-------------|
|         | Karbi Anglong | Kamrup | Keonjhar | Sundergarh | Darjeeling |
| **msp-1** | | | | | |
| K1 Obs. Nos. (%) | 16 (72.7) | 15 (53.6) | 8 (33.3) | 22 (55.0) | 15 (68.2) |
| Allele | 5 | 2 | 3 | 5 | 1 |
| Size range (bp) | 140–240 | 170–240 | 170–240 | 170–280 | 140 |
| MAD20 Obs. Nos. (%) | 16 (72.7) | 11 (39.3) | 8 (33.3) | 18 (45.0) | 3 (13.6) |
| Allele | 3 | 3 | 1 | 3 | 1 |
| Size range (bp) | 120–220 | 200–240 | 220 | 120–180 | 120 |
| RO33 Obs. Nos. (%) | 5 (22.7) | 11 (39.3) | 10 (41.7) | 6 (15.0) | 7 (31.8) |
| Allele | 1 | 1 | 1 | 1 | 1 |
| Size range (bp) | 150/160 | 150/160 | 150/160 | 150/160 | 150/160 |
| **MOI** | 2.1 | 1.4 | 1.1 | 1.2 | 1.1 |
| **msp-2** | | | | | |
| FC27 Obs. Nos. (%) | 9 (40.9) | 7 (25.0) | 13 (68.4) | 34 (85.0) | 4 (18.2) |
| Allele | 6 | 1 | 1 | 3 | 1 |
| Size range (bp) | 250–500 | 300 | 300 | 300–400 | 300 |
| 3D7 Obs. Nos. (%) | 18 (81.8) | 23 (82.1) | 11 (57.9) | 31 (77.5) | 18 (81.2) |
| Allele | 6 | 2 | 3 | 8 | 2 |
| Size range (bp) | 480–600 | 450,550 | 450–500 | 420–580 | 450,500 |
| **MOI** | 1.2 | 1.1 | 1.3 | 1.6 | 1.0 |

**MOI** – Multiplicity of Infection
### Alignment for the amino acid sequences corresponding to a) K1, b) MAD20 and c) RO33 families of \textit{msp-1}.

Sequences shown are either of isolates collected during the study or from the GenBank database. Deletions are represented by minus sign (-) and alphabets represent a change in amino acid at the position.

#### a) K alignments

|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |

**Majority**

| 10    | 10     | 10     | 10     | 10     | 10     |
|-------|--------|--------|--------|--------|--------|
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam | 7 Kollam | 8 Kollam | 9 Kollam | 10 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam | 7 Kollam | 8 Kollam | 9 Kollam | 10 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam | 7 Kollam | 8 Kollam | 9 Kollam | 10 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam | 7 Kollam | 8 Kollam | 9 Kollam | 10 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam | 7 Kollam | 8 Kollam | 9 Kollam | 10 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam | 7 Kollam | 8 Kollam | 9 Kollam | 10 Kollam |

#### b) M alignments

|       | 1       | 2       | 3       | 4       | 5       | 6       |
|-------|---------|---------|---------|---------|---------|---------|
|       | 1       | 2       | 3       | 4       | 5       | 6       |
|       | 1       | 2       | 3       | 4       | 5       | 6       |
|       | 1       | 2       | 3       | 4       | 5       | 6       |
|       | 1       | 2       | 3       | 4       | 5       | 6       |
|       | 1       | 2       | 3       | 4       | 5       | 6       |
|       | 1       | 2       | 3       | 4       | 5       | 6       |

**Majority**

| 5       | 5       | 5       | 5       | 5       | 5       |
|---------|---------|---------|---------|---------|---------|
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |

#### c) R alignments

|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |

**Majority**

| 10    | 10     | 10     | 10     | 10     | 10     |
|-------|--------|--------|--------|--------|--------|
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |
|       | 1 West Bengal | 2 Sonapur | 3 Madhupur | 4 Mussoorie | 5 Sundergarh | 6 Kollam |

### Figure 2
Alignment for the amino acid sequences corresponding to a) K1, b) MAD20 and c) RO33 families of \textit{msp-1}. Sequences shown are either of isolates collected during the study or from the GenBank database. Deletions are represented by minus sign (-) and alphabets represent a change in amino acid at the position.
Alignment of the amino acid sequences corresponding to a) FC27 and b) 3D7 families of msp-2. Sequences shown are either of isolates collected during the study or from the GenBank database. Deletions are represented by minus sign (-) and alphabets represent a change in amino acid at the position.
flow. Observed high proportion of multiclonal isolates was in accordance with the reports of other workers [17,18] on Indian isolates. A comparatively higher proportion of multiclonal isolates as well as multiplicity of infection (MOI) was observed among isolates of district Karbi Anglong (Assam) and district Sundergarh (Orissa). Both the districts are highly endemic for malaria with *P. falciparum* proportion being above 80%. Similar observations that the extent of diversity and multiplicity of infection in an area is related to level of malaria endemicity has been made by Ranjit & Sharma [18], Babiker et al [26] and Zwetyenga et al [27].

Present study for the first time revealed that population structure of *P. falciparum* is identical in two highly malarious regions (Assam and Orissa states) of India as revealed by presence of common allelic composition in both the states as well as high level of identity among allelic sequences of isolates from two areas. However, a higher MOI based on *msp-1* in Assam and based on *msp-2* in Orissa, suggests that local factors such as vector population, human host as well as drug susceptibility pattern of the parasites in an area may be playing a role in defining the population structure of the field isolates. A recent study from Iran reports high level of diversity of *msp-1* and *msp-2* markers along with high proportion of multiclonal isolates (87%) and MOI (3.06), due to emergence of drug-resistant *P. falciparum* [13].

Distribution of families of *msp-1* and *msp-2* and their allelic variations were similar to that reported from other countries with low or meso-endemicity of malaria i.e. Southeast Asia, Latin America and Papua New Guinea [2,4,5,9,24,28,29]. Monomorphic nature of RO33 family of *msp-1* has also been reported earlier in isolates of other regions including India and observed 150/160 bp was the most commonly reported allele in other continents also [1,7,8,10]. Observed good sequence identity of Indian isolates with data of other regions suggests that Indian *P. falciparum* population is a mixture of different strains. Further studies on the genetic diversity of *P. falciparum* isolates from other regions with varied malaria epidemiology as well as longitudinal studies to understand the clonal fluctuations associated with transmission intensity are important, more so from Indian subcontinent which is a large country with varied malaria paradigm.

**Conclusion**

The present study shows that field isolates of eastern and north-eastern regions of India are highly diverse in respect of *msp-1* (block 2) and *msp-2* (central repeat region, block 3) with identical population structure and exhibit a level of diversity similar to that in Papua New Guinea, Southeast Asia and South and Central America, regions with low to meso endemicity of malaria.

**Competing interests**

The author(s) declare that they have no competing interests.

**Authors’ contributions**

HI – Conception & designing of the study, co-ordination of molecular analysis work, analysis & interpretation of data, manuscript preparation. NV – Planning, designing and co-ordination of field studies. AV – Sequencing assays, data analysis & interpretation. AK, PKM, SS, SKP – PCR assays, compilation and analysis of data, drafting of manuscript. SKS, VD, SB, NN, MSM – Epidemiological studies. SKS and APD – Guidance.

**Acknowledgements**

Authors are grateful to Prof. N.K. Ganguly, Director-General, ICMR for granting the permission to conduct the study and funding the study partially under genomics to senior author. Authors acknowledge the support and co-operation of staff members of NIIR, HQ as well as field units. Excellent technical assistance for molecular genotyping by Mrs. Alka Kapoor and Mrs. Lalita Gupta is sincerely acknowledged. Thanks are due to Dr. Patricia Campbell, MSF, Assam for providing Karbi Anglong samples. Authors are thankful to Ms. Pinky Saxena for typing the manuscript.

**References**

1. Aubouy A, Migot-Nabias F, Deloron P: Polymorphism in two merozoite surface proteins of *Plasmodium falciparum* isolates from Gabon. *Malar J* 2003, 2:12.

2. Ariey F, Chalvet W, Hommel D, Peneau C, Hulin A, Mercereau-Puijalon O, Duchemin JB, Sarthou JL, Arthou JL, Reyes JM, Fandeur T: *Plasmodium falciparum* parasites in French Guiana: limited genetic diversity and high selfing rate. *Am J Trop Med Hyg* 1999, 61:978-985.

3. Babiker HA, Ranford-Cartwright LC, Currie D, Charlwood JD, Billingsley P, Teuscher T, Walliker D: Random mating in a natural population of malaria parasite *Plasmodium falciparum*. *Parasitology* 1994, 109:413-421.

4. Gomez D, Chaparro J, Rubiano C, Rojas MO, Wasserman M: Genetic diversity of *Plasmodium falciparum* field samples from an isolated Colombian village. *Am J Trop Med Hyg* 2002, 67:611-616.

5. Haddad D, Sunounou G, Mattei D, Enamorado IG, Figueroa J, Stahl S, Berzins K: Limited genetic diversity of *Plasmodium falciparum* field isolates from Honduras. *Am J Trop Med Hyg* 1999, 60:30-34.

6. Irion A, Felger I, Abdulla S, Smith T, Mull R, Tanner M, Beck HP, Beck HP: Distinction of recrudescence from new infections by PCR-RFLP analysis in a comparative trial of cgp 56 67 and chloroquine in Tanzanian children. *Trop Med Int Health* 1998, 3:490-494.

7. Jordan S, Jelinek T, Aida AO, Peyerl-Hoffmann G, Heuschkel C, el Valy AO, Christophel EM: Population structure of *Plasmodium falciparum* isolates during an epidemic in southern Mauritania. *Trop Med Int Health* 2001, 6:761-766.

8. Konate L, Zwetyenga J, Rogier C, Bischoff E, Fontenelle D, Tall A, Spiegel A, Trape JF, Mercereau-Puijalon O: Variation of *Plasmodium falciparum msp-1* block 2 and *msp-2* allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. *Trans R Soc Trop Med Hyg* 1999, 93:21-28.

9. Montoya L, Maestre A, Carmona J, Lopes D, Do Rosario V, Blair S: *Plasmodium falciparum*: diversity studies of isolates from two Colombian regions with different endemicity. *Exp Parasit* 2003, 104:14-19.

10. Peyerl-Hoffmann G, Jelinek T, Kilian A, Kabagambe G, Metzger WG, von Sonnenburg F: Genetic diversity of *Plasmodium falciparum* and its relationship to parasite density in an area with differ-
ent malaria endemicities in West Uganda. Trop Med Int Health 2001, 6:607-613.

11. Snounou G, Zhu X, Siripoon N, Jarra W, Thaithong S, Neil Brown K, Vrijeyakosol S: Biased distribution of msp-1 and msp-2 allelic variants in Plasmodium falciparum population in Thailand. Trans R Soc Trop Med Hyg 1999, 93:369-374.

12. Tanabe K, Mackay M, Goman M, Scala JG: Allelic dimorphism in a surface antigen of the malaria parasite Plasmodium falciparum. J Mol Biol 1987, 195:73-287.

13. Zakeri S, Bereczky S, Naimi P, Pedro-Gill J, Djadid ND, Farnert A, Snounou G, Bjorkman A: Multiple genotypes of the merozoite surface proteins 1 and 2 in Plasmodium falciparum infections in a hypoendemic area in Iran. Trop Med Int Health 2005, 10:1060-1064.

14. Bhattacharya PR: Plasmodium falciparum: genetic polymorphism of the merozoite surface antigen 2 gene of strains from India. Trans R Soc Trop Med Hyg 1998, 92:225-226.

15. Bhattacharya PR, Kumar M, Das RH: Surprising little polymorphism in the merozoite-surface-protein-2 (MSP-2) gene of Indian Plasmodium falciparum. Ann Trop Med Parasit 1999, 93:561-564.

16. Joshi H: Markers for population genetic analysis of human plasmodia species, P. falciparum and P. vivax. J Vector Borne Dis 2003, 40:78-83.

17. Raj DK, Das BR, Dash AP, Supakar PC: Genetic diversity in the merozoite surface protein 1 gene of Plasmodium falciparum in different malaria-endemic localities. Am J Trop Med Hyg 2004, 72:285-289.

18. Ranjit MR, Sharma YD: Genetic polymorphism of falciparum malaria vaccine candidate antigen genes among field isolates in India. Am J Trop Med Hyg 1999, 61:103-108.

19. Ranjit MR, Das A, Das BP, Das BN, Dash BP, Chhotray GP: Distribution of Plasmodium falciparum genotypes in clinically mild and severe malaria cases in Orissa, India. Trans R Soc Trop Med Hyg 2005, 99:389-395.

20. Smythe JA, Peterson MG, Coppel RL, Saul AJ, Kemp DJ, Anders RF: Structural diversity in the 45-kilodalton merozoite surface antigen of Plasmodium falciparum. Mol Biochem Parasitol 1999, 99:227-234.

21. Smythe JA, Coppel RL, Day KP, Martin RK, Odoula AM, Kemp DJ, Anders RF: Structural diversity in the Plasmodium falciparum merozoite surface antigen 2. Proc Natl Acad Sci USA 1991, 88:1751-1755.

22. Dev V, Hira CR, Rajkhowa MK: Malaria attributable morbidity in Assam, north-eastern India. Ann Trop Med Parasitol 2001, 95:789-792.

23. Sharma RS, Sharma GK, Dhillon GPS: Epidemiology and control of malaria in India. NMEP, Directorate General of Health Services, Delhi; 1996:1-752.

24. Sharma SK, Tyagi PK, Padhan K, Upadhyak AK, Haque MA, Nanda M, Joshi H, Biswas S, Adak T, Das BS, Chauhan VS, Chitnis CE, Subbarao SK: Epidemiology of malaria in the forest and plain ecotopes of Sundergarh District, Orissa, India. Trans R Soc Trop Med Hyg 2006, 100:917-925.

25. Okeyo JN, Pillai CR, Chitnis CE: Plasmodium falciparum field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of Glycophorin A. Infection Immun 1999, 67:5784-5791.

26. Babiker HA, Lines J, Hill WG, Walliker D: Population structure of Plasmodium falciparum in villages with different malaria endemicity in east Africa. Am J Trop Med Hyg 1997, 56:141-147.

27. Zewzenga J, Rognier C, Tall A, Fontenille D, Snounou G, Trape JF, Mercereau-Puijalon O: No influence of age on infection complexity and allelic distribution in Plasmodium falciparum infections in Ndiop, a Senegalese village with seasonal, mesoendemic malaria. Am J Trop Med Hyg 1998, 59:726-735.

28. Paul REL, Packer MJ, Walmsey M, Lagow M, Ranford-Cartwright LC, Paru R, Day KP: Mating patterns in malaria parasite populations of Papua New Guinea. Science 1995, 269:1709-1711.

29. Paul REL, Hackford I, Brockman A, Muller-Gräf C, Price R, Luxembourg C, White NJ, Nosten F, Day KP: Transmission intensity and Plasmodium falciparum diversity on the Northwestern border of Thailand. Am J Trop Med Hyg 1998, 58:195-203.