**RESEARCH NOTE**

# Genetic differentiation of populations residing in areas of high malaria endemicity in India

SWAPNIL SINHA[1], VANDANA ARYA[2], SARITA AGARWAL[2], INDIAN GENOME VARIATION CONSORTIUM[3] and SAMAN HABIB[1]*

[1]*Division of Molecular and Structural Biology, Central Drug Research Institute, Lucknow 226 001, India*
[2]*Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute for Medical Sciences, Lucknow 226 014, India*
[3]*Institute for Genomics and Integrative Biology, New Delhi 110 007, India*

## Introduction

The frequency of five single nucleotide polymorphisms (SNPs) (*TNF*: rs1800629, rs361525; *ICAM-1*: rs5491; *TLR9*: rs187084 and *FCGR2A*: rs1801274), previously correlated with severity of falciparum malaria in some world populations, was used to analyse genetic differentiation among 55 ethnically-diverse populations of India. Significant differentiation of the north Indian Tharu tribe inhabiting the malaria-endemic Terai region and tribal populations of the falciparum-endemic eastern region of India was revealed. In addition to high frequency of the protective *FCGR2A* exon 4 *A* allele and low frequency of the *TLR9 C* susceptibility allele in these populations, near-fixation of the $\alpha$-thalassaemia $\alpha^{3.7}$ allele was observed in Indian Tharus, indicative of their genetic proximity to Tharus of western or central Nepal.

*Plasmodium falciparum* malaria has been a major selection force in human-evolutionary history and association of host genetic variants with susceptibility/resistance to falciparum malaria has been reported in genes involved in erythrocytic defects, as well as immune regulation and cytoadhesion (Kwaitkowski 2005). Analysis of population-specific differences with respect to malaria-related SNPs has been carried out mostly in populations of Africa and Southeast Asia. Some of these data have provided valuable information on the possible role of specific SNPs in influencing disease pathogenesis and resistance in certain populations (Gourley *et al*. 2002; Koch *et al*. 2005). *Plasmodium falciparum* malaria is endemic in many parts of India, with ~0.9 million cases reported annually in the past few years (http://www.nvbdcp.gov.in) and may have influenced the frequency profile of specific SNPs populations in these regions. There is limited information on the distribution pattern of SNPs in genes related to malaria-disease manifestation in Indian populations. We therefore analysed genetic differentiation among 55 populations representing the linguistic and ethnic diversity of India (Indian Genome Variation Consortium 2008) in terms of selected SNPs that have been associated with the outcome of malaria infection in the other world populations.

*For correspondence. E-mail: saman.habib@gmail.com; samamit@lycos.com.

## Materials and methods

Data from 1871 healthy individuals that were genotyped under the Indian Genome Variation Consortium (IGVC) (Indian Genome Variation Consortium 2008; genotype information is available on http://igvdb.res.in) was used in this analysis. These individuals were drawn from 55 endogamous populations of contrasting ethnicity belonging to diverse linguistic classes and geographical zones across India (Indian Genome Variation Consortium 2008). Population descriptors had linguistic affiliation (AA, Austro–Asiatic; DR, Dravidian; IE, Indo–European; TB, Tibeto–Burman) followed by the geographical location and ethnic category (LP, large population; IP, isolated population; SP, religious groups). OG is an outgroup population of known African descent. The criteria for selection of populations and numbers of individuals genotyped for each population are described in Indian Genome Variation Consortium (2008). Briefly, 46 adult individuals were genotyped from LPs and large IPs, while 21–23 adult individuals were genotyped from smaller IPs. The final 1871 individuals comprised of 1240 males and 631 females. Genotyping of SNPs was performed using the Sequenom mass array (Indian Genome Variation Consortium 2008). Linkage disequilibrium (LD) between *TNF* promoter SNPs was

**Keywords.** falciparum malaria; single nucleotide polymorphism (SNP); tumour necrosis factor; human genetics; Indian populations.

determined using Haploview (Broad Institute, Cambridge, USA). Population pair-wise $F_{ST}$ was calculated from allele frequencies of the five selected SNP loci using FSTAT (version 2.9.3.2) (Lausanne University, Lausanne, Switzerland). Principal component analysis (PCA) was carried out using XLSTAT (Addinsoft, Brooklyn, USA). The prevalence of the $\alpha$-thalassaemia-$\alpha^{3.7}$ deletion mutation in Indian Tharus was determined by gap-PCR (Baysal and Huisman 1994) of DNA from 63 individuals.

## Results and discussion

The extent of genetic differentiation among Indian populations with respect to SNPs previously correlated with malaria susceptibility was evaluated from allele frequency data for *TNF*-308 (G/A, rs1800629) and *TNF*-238 (G/A, rs361525), *ICAM*-$1^{Kilifi}$ (A/T, rs5491), *FCGR2A* exon 4 (G/A, rs1801274) and *TLR9*-1486 (T/C, rs187084) polymorphisms. These SNPs have been previously correlated with malaria susceptibility/resistance primarily in African populations (McGuire *et al*. 1994; Fernandez-Reyes *et al*. 1997; Ubalee *et al*. 2001; Mockenhaupt *et al*. 2006; Nasr *et al*. 2007). *TNF*-308 and *TNF*-238 SNPs are not in significant LD in Indian populations ($r^2$ = 0.022; determined from data on 55 Indian Genome Variation Consortium populations) and are thus unlikely to bias differentiation analysis (Sinha *et al*. 2008). Population pair-wise $F_{ST}$ identified seven isolated populations: IE-N-IP2, AA-E-IP3, AA-NE-IP1, DR-C-IP1, DR-C-IP2, DR-E-IP1 and DR-S-IP1 that exhibited the highest level of differentiation (mean $F_{ST}$ > 5%; figure 1,a).

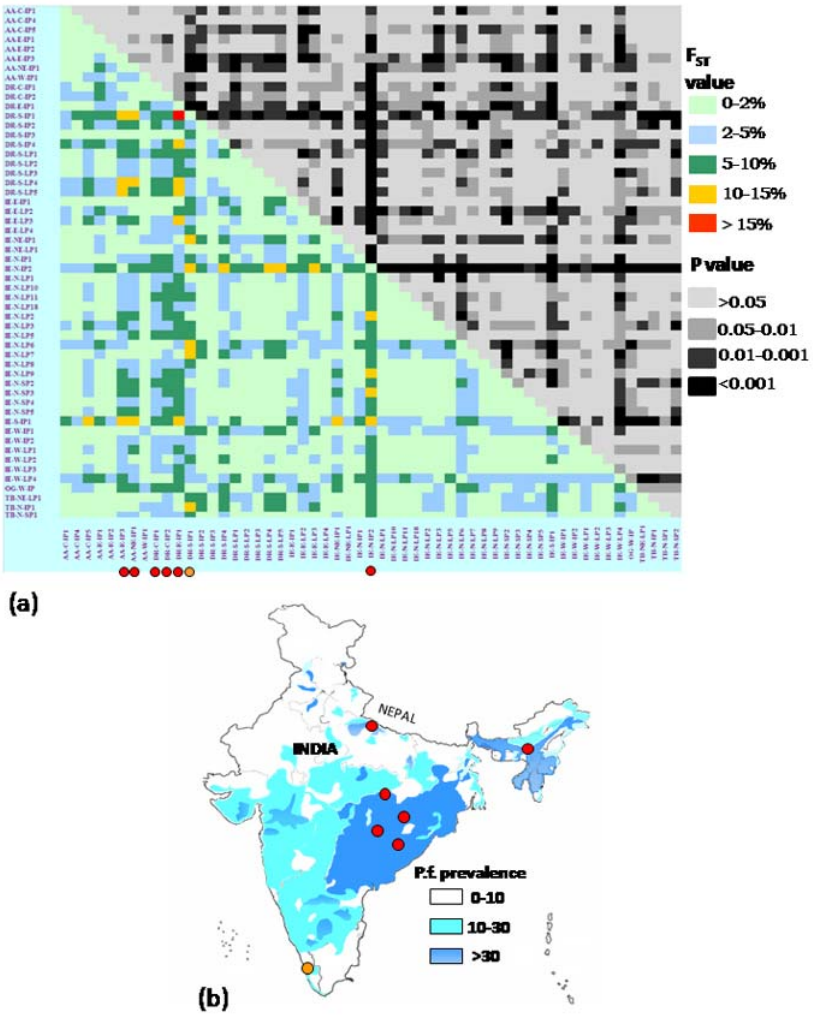


**Figure 1.** Genetic differentiation and falciparum malaria incidence in India. (a) Pairwise $F_{ST}$ calculated using data from five SNPs in 55 populations. Colour scale indicates $F_{ST}$ values in percentage and the grayscale indicates the *P* values. (b) Map depicting *P. falciparum* malaria incidence across India as percentage of all malaria cases. The locations of the highly differentiating populations are indicated.

Interestingly, all these populations belong to tribal groups, and all except DR-S-IP1 inhabit regions endemic to falciparum malaria in India (figure 1,b). DR-C-IP1, DR-C-IP2, AA-E-IP3 and DR-E-IP1 inhabit the *P. falciparum*-endemic regions of central and eastern India, AA-NE-IP1 resides in northeastern India, IE-N-IP2 is a tribal population of north India, while DR-S-IP1 is a tribe of southern India. DR-S-IP1 exhibited high $F_{ST}$ with most other populations with the highest $F_{ST}$ observed between this population and DR-E-IP1 ($F_{ST}$ = 15.87%) and IE-N-IP2 ($F_{ST}$ = 13.25%) (figure 1,a).

PCA on the basis of the five SNPs also revealed maximum distance between DR-S-IP1 and the malaria-endemic-region populations IE-N-IP2 and DR-E-IP1 (figure 2). Of the populations identified in the $F_{ST}$ analysis, only IE-N-IP2, DR-E-IP1 and AA-E-IP3 were maximally separated in the first component. These three populations have a high frequency of the *FCGR2A* exon 4 *A* allele that has recently been reported to be associated with protection from falciparum malaria in India (Sinha *et al*. 2008) and low frequency of the *TLR9 C* allele that has been associated with susceptibility to pregnancy-associated malaria in Africa (Mockenhaupt *et al*. 2006). High frequency of a protective allele and low frequency of a susceptibility allele in these populations are suggestive of selection by falciparum malaria. The separation of DR-S-IP1 with all the other populations, predominantly in the second component, is explained by high frequency of the *ICAM*-$1^{Kilifi}$ SNP and monomorphism at the *TNF*-308 and *FCGR2A* exon 4 SNP loci in this population.

IE-N-IP2 showed markedly high $F_{ST}$ throughout (mean $F_{ST}$ = 8.2%) and exhibited a dramatically high frequency of *TNF*-308 *A* allele (29%) as compared to the mean minor allele frequency (MAF) across all populations (6.06%). Additionally, the frequency of the malaria protective *FCGR2A AA* genotype was very high (0.72) in this population as well as in DR-E-IP1 (0.74), as compared to the average of 0.39. In the analysis of genetic affinities among the 55 populations based on data from 405 SNPs in 75 genes and a Chr22 region (Indian Genome Variation Consortium 2008), IE-N-IP2 consistently clustered with TB-speaking and IE-speaking populations of the Himalayan belt. However, when analysed in the context of malaria-related SNPs, IE-N-IP2 did not cluster with populations of the Himalayan belt in PCA (figure 2) and also exhibited high $F_{ST}$ with these populations (figure 1), indicating that the observed differentiation of IE-N-IP2 was possibly a result of natural selection by malaria and not population history alone. IE-N-IP2 (Tharu) inhabits the low-land Terai region of the Uttar Pradesh province of India, bordering southern Nepal. The Terai region of India and Nepal is endemic for malaria and the Tharus have been reported to show decreased morbidity to *P. falciparum* malaria

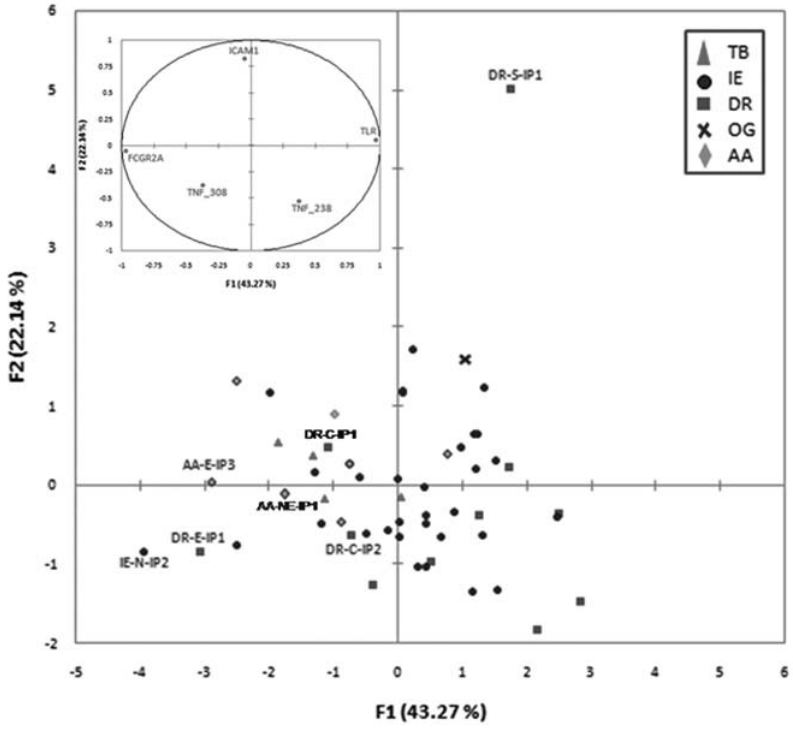


**Figure 2.** Two-dimensional principal component analysis computed using allele frequencies from five malaria-related SNPs. Symbol key indicating linguistic affiliations of populations is shown. Inset shows the correlation between variables and components.

(Terrenato *et al*. 1988). One possible explanation for this is the reported near fixation of $\alpha$-thalassaemia in central and western Tharus of Nepal ($-\alpha^{3.7}$, frequency = 0.8) (Modiano *et al*. 1991). We have genotyped the $-\alpha^{3.7}$ deletion mutation to detect the prevalence of $\alpha$-thalassaemia in Indian Tharus of Uttar Pradesh. The frequency of $-\alpha^{3.7}$ allele was markedly high (0.94) indicating near fixation of the $\alpha$-thalassemia genotype in the Indian Tharu population as well. The high frequency of $-\alpha^{3.7}$ allele in Tharus of Uttar Pradesh is indicative of their genetic proximity to central or western Tharus of Nepal rather than the eastern Tharus that have $-\alpha^{3.7}$ frequency of 0.04 (Passarino *et al*. 1992).

We observed striking genetic differentiation among populations inhabiting regions of high malaria endemicity on the basis of five SNPs. Apart from the fixation of $\alpha$-thalassaemia in Indian Tharus, the high frequency of the *FCGR2A* exon 4 *AA* genotype that is significantly associated with protection from manifestation of falciparum malaria in India (Sinha *et al*. 2008) may contribute to the resistance of Tharus to malaria. The association of the *FCGR2A* exon 4 *AA* genotype with protection from severe malaria has also been reported in New Halfa town population of eastern Sudan (Nasr *et al*. 2007), as well as in our study on tribal populations of the Sundargarh district in Orissa (Sinha *et al*. 2008). Of the remaining four SNPs that have been correlated with malaria severity/resistance in populations from Africa and Southeast Asia, none have been analysed for their association with falciparum malaria in India. Detailed case–control studies using these markers may provide insights about their observed distribution, disease association and selection in the context of falciparum malaria in Indian populations.

**Acknowledgements**

Funding by the Council for Scientific and Industrial Research (CMM0016) and the Department of Biotechnology (BT/PR6065/MED/14/738/2005) of the Government of India is acknowledged. This is CDRI communication no. 7518.

## References

Baysal E. and Huisman T. H. J. 1994 Detection of common deletional $\alpha$-thalassemia-2 determinants by PCR. *Am. J. Hematol*. **46**, 208–213.

Fernandez-Reyes D., Craig A. G., Kyes S. A., Peshu N., Snow R. W. and Berendt A. R. 1997 A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum. Mol. Genet*. **6**, 1357–1360.

Gourley I. S., Kurtis J. D., Kamoun M., Amon J. J. and Duffy P. E. 2002 Profound bias in interferon-$\gamma$ and interleukin 6 allele frequencies in western Kenya, where severe malarial anemia is common in children. *J. Infect. Dis*. **186**, 1007–1012.

Indian Genome Variation Consortium 2008 Genetic landscape of the people of India: a canvas for disease gene exploration. *J. Genet*. **87**, 3–20.

Koch O., Rockett K., Jallow M., Pinder M., Sisay-Joof F. and Kwaitkowski D. 2005 Investigation of malaria susceptibility determinants in the IFN gene/IL26/IL22 genomic region. *Genes Immun*. **6**, 312–318.

Kwaitkowski D. 2005 How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet*. **77**, 171–192.

Mockenhaupt F. P., Hamann L., von-Gaertner C., Bedu-Addo G., von-Kleinsorgen C., Schumann R. R. and Bienzle U. 2006 Common polymorphisms of toll-like receptors 4 and 9 are associated with clinical manifestation of malaria during pregnancy. *J. Infect. Dis*. **194**, 184–188.

Modiano G., Morpurgo G., Terrenato L., Novelletto A., Di Rienzo A., Colombo B. *et al*. 1991 Protection against malaria morbidity: near fixation of the $\alpha$-thalassemia gene in a Nepalese population. *Am. J. Hum. Genet*. **48**, 390–397.

McGuire W., Hill A. V. S., Allsopp C. E. M., Greenwood B. M. and Kwiatkowski D. 1994 Variation in the TNF-$\alpha$ promoter region associated with susceptibility to cerebral malaria. *Nature* **371**, 508–510.

Nasr A., Iriemenam N. C., Troye-Blomberg M., Giha H. A., Balogun H. A., Osman O. F. *et al*. 2007 Fc gamma receptor IIa (CD32) polymorphism and antibody responses to asexual blood-stage antigens of *Plasmodium falciparum* malaria in Sudanese patients. *Scand. J. Immunol*. **66**, 87–96.

Passarino G., Semino O., Pepe G., Shrestha S. L., Modiano G. and Santachiara Benerecetti A. S. 1992 MtDNA polymorphisms among Tharus of eastern Terai (Nepal). *Gene Geogr*. **6**, 139–147.

Sinha S., Mishra S. K., Sharma S., Patibandla P. K., Mallick P. K., Sharma S. K. *et al*. 2008 Polymorphisms of *TNF*-enhancer and gene for Fc$\gamma$RIIa correlate with the severity of falciparum malaria in the ethnically diverse Indian population. *Malaria J*. **7**, 13.

Terrenato L., Shrestha S., Dixit K. A., Luzzatto L., Modiano G., Morpurgo G. and Arese P. 1988 Decreased malaria morbidity in the Tharu people compared to sympatric populations in Nepal. *Ann. Trop. Med. Parasitol*. **82**, 1–11.

Ubalee R., Suzuki F., Kikuchi M., Tasanor O., Wattanagoon Y., Ruangweerayut R. *et al*. 2001 Strong association of a tumor necrosis factor-$\alpha$ promoter allele with cerebral malaria in Myanmar. *Tissue Antigens* **58**, 407–410.

Received 18 June 2008, in revised form 13 August 2008; accepted 21 August 2008
Published on the Web: 13 March 2009