Patterns of nucleotide sequence variation in *ICAM1* and *TNF* genes in twelve ethnic groups of India: roles of demographic history and natural selection

SANGHAMITRA SENGUPTA, SHABANA FARHEEN, NEELANJANA MUKHERJEE and PARTHA P. MAJUMDER*

Human Genetics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata 700108, India

Abstract

We have studied DNA sequence variation in and around the genes *ICAM1* and *TNF*, which play functional and correlated roles in inflammatory processes and immune cell responses, in 12 diverse ethnic groups of India, with a view to investigating the relative roles of demographic history and natural selection in shaping the observed patterns of variation. The total numbers of single nucleotide polymorphisms (SNPs) detected at the *ICAM1* and *TNF* loci were 29 and 12, respectively. Haplotype and allele frequencies differed significantly across populations. The site frequency spectra at these loci were significantly different from those expected under neutrality, and showed an excess of intermediate-frequency variants consistent with balancing selection. However, as expected under balancing selection, there was no significant reduction of F_{ST} values compared to neutral autosomal loci. Mismatch distributions were consistent with population expansion for both loci. On the other hand, the phylogenetic network among haplotypes for the *TNF* locus was similar to expectations under population expansion, while that for the *ICAM1* was as expected under balancing selection. Nucleotide diversity at the *ICAM1* locus was an order of magnitude lower in the promoter region, compared to the introns or exons, but no such difference was noted for the *TNF* gene. Thus, we conclude that the pattern of nucleotide variation in these genes has been modulated by both demographic history and selection. This is not surprising in view of the known allelic associations of several polymorphisms in these genes with various diseases, both infectious and noninfectious.

[Sengupta S., Farheen S., Mukherjee N. and Majumder P. P. 2007 Patterns of nucleotide sequence variation in *ICAM1* and *TNF* genes in twelve ethnic groups of India: roles of demographic history and natural selection. *J. Genet.* **86**, 225–239]

Introduction

The intercellular adhesion molecule one (*ICAM1*) and tumor necrosis factor α (*TNF*) genes are known to play important functional and correlated roles in inflammatory processes and immune cell responses in a wide range of diseases, both noninfectious and infectious (Bjornsdottir and Cypcar 1999; Dobbie *et al*. 1999; Fernandez-Arquero *et al*. 1999; Knight and Kwiatkowski 1999; McGuire *et al*. 1999; Negoro *et al*. 1999; Striz *et al*. 1999; Kawasaki *et al*. 2000; Zeggini *et al*. 2002; Thio *et al*. 2004). Both *ICAM1* and *TNF*, appear to play an important roles in malarial susceptibility (Hill 1992; Fernandez-Reyes *et al*. 1997; McGuire *et al*. 1999). The pathogenecity of *Plasmodium falciparum* has been ascribed to the ability of the infected red blood cells to adhere to capillary endothelium (Paloske and Howard 1994). ICAM1 has been shown to be an endothelial cell adhesion receptor for *Plasmodium falciparum* (Berendt *et al*. 1989). In a histopathological study, it was shown that the presence of parasitised erythrocytes in cerebral vessels colocalized with endothelial expression of *ICAM1*, indicating that *ICAM1* is an endothelial receptor for infected erythrocytes in cerebral malaria (Turner *et al*. 1994). Therefore, similar to the MHC locus (Grimsley *et al*. 1998), it is possible that heterozygotes for different variants at the *ICAM1* locus enjoy a selective advantage when exposed to various pathogens, since *ICAM1* acts as a receptor. Thus, balancing selection may play an important role in maintaining genetic variation at this locus.

Various alleles in the *TNF* promoter have been found to be associated with cerebral malaria and severe malarial ane-

*For correspondence. E-mail: ppm@isical.ac.in.

Keywords. human genetics; diversity; immunity; inflammation; selection; population expansion.

mia (McGuire *et al*. 1999). However, *TNF* seems to have both beneficial and detrimental functions. It can activate host defense and promote resistance to infectious diseases, and it can also be involved in toxicity (Kwiatkowski *et al*. 1993; Gimenez *et al*. 2003). Thus, natural selection may not operate in a homogeneous or unidirectional mode at this locus. It is also known that there is an interaction between the *ICAM1* and *TNF* gene products in the inflammatory processes and immune cell responses in a wide range of diseases. The cytokine TNF is known to upregulate the endothelial adhesion molecule ICAM1 (Meager 1999).

The facts stated above indicate that a complex set of interacting evolutionary forces may operate at the *ICAM1* and *TNF* loci in maintaining the DNA sequence variation. Moreover, this variation is also determined in part, by the evolutionary histories of the populations sampled to estimate it. We, therefore, sought to explore the relative roles of demographic history and natural selection on the nature and extent of the DNA sequence variability at these two interacting loci. For this, we have carried out a systematic survey, by DNA sequencing, of polymorphisms in and around these two genes in 208 individuals drawn from 12 population groups of India with diverse ethnic, ecological and epidemiological backgrounds. We have analysed these data, in conjunction with mitochondrial DNA (mtDNA) sequence data, to draw appropriate inferences.

Materials and methods

Populations

There are over 1000 endogamous ethnic groups present in India (Singh 1992). These groups are broadly classified into two major clusters—tribes and castes. The tribes are considered as the authochthones of India. The vast majority of tribal groups live in isolation, inhabit geographically remote areas and practice hunting and gathering or primitive forms of agriculture. The caste groups belong to the Hindu religious fold, and practice various occupations. It is generally acknowledged that there has been considerable admixture of the caste populations with local tribals and with immigrants from other regions of the world in prehistoric and historic times (Thapar 2003). Both tribal and caste populations are spread throughout India. Because of their different ancestral histories, in this study we have sought to obtain representation of both caste and tribal groups from diverse geographical regions of India, to further reduce the possibility of biases that may stem from regional differences in prevalence of infections and other diseases.

This study was initiated after obtaining appropriate ethical approvals. Blood samples were drawn with informed consent from normal, healthy individuals unrelated to the first cousin level. These individuals belonged to 12 distinct ethnic groups (six tribal and six caste) inhabiting five different geographical regions of mainland India and the Andaman and Nicobar Islands (figure 1).

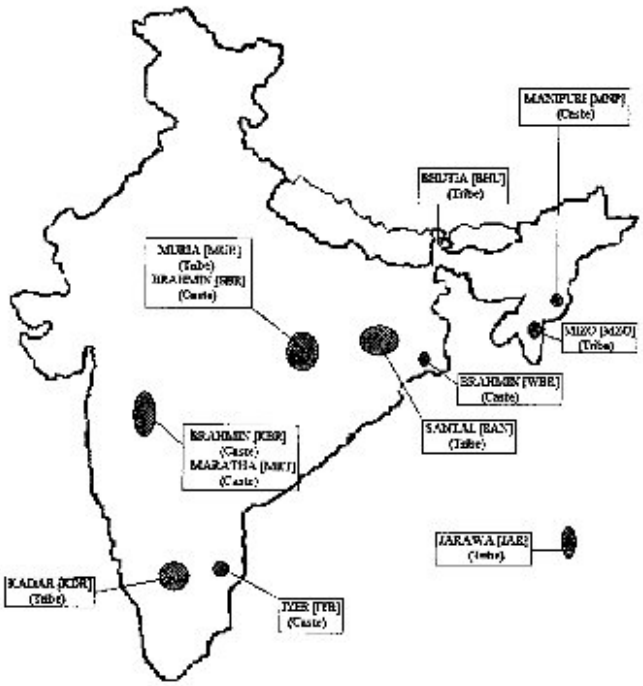

Figure 1. Geographical locations and background information regarding the study populations.

Anonymized blood samples from the Jarawas were collected and stored at the Regional Medical Research Centre, Indian Council of Medical Research, Port Blair and were used in this study, after obtaining the approval of the Ethics Committee of the Regional Medical Research Centre, Port Blair.

Experimental protocols

The *ICAM1* gene maps to 19p13.3–p13.2 and contains seven exons. The *TNF* gene maps to 6p21.3 and contains four exons. Genomic sequences of these two genes were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu). The genomic region encompassing *ICAM1* was repeat-masked using the program RepeatMasker2 (http://ftp.genome.washington.edu.cgi-bin/RepeatMasker). Appropriate primers to amplify the exons, introns, the 5′ and a portion of the 3′ untranslated regions (UTRs) of these genes (excluding the repeat-masked region of *ICAM1*) were designed. The total number of bases resequenced for each individuals were 6000 and 3046, respectively for the *ICAM1* and *TNF* genes.

DNA amplification conditions by PCR were optimized using control samples. PCR products were cleaned using Exonuclease I and Shrimp Alkaline Phosphatase, and subjected to sequencing on an ABI-3100 automated sequencer using dye-terminator chemistry (primer sequences and PCR conditions are given in table 1 of appendix.) ABI trace files thus generated were analysed using the PHRED software (http://www.mbt.washington.edu/phrap.docs/phred.html), which assigns quality score to each base. The PHRED outputs for all the individuals for any given

PCR amplicon were aligned using PHRAP software (http://www.phrap.org/phredphrapconsed.html). The resulting assemblies were viewed using CONSED (http://www.phrap.org/phredphrapconsed.html) that allows identification of the putative sequence variants. All samples with putative variant alleles were resequenced in reverse direction for confirmation.

Statistical analysis

Allele frequencies at each variant site were computed by the gene-counting method. Maximum likelihood estimates of haplotype frequencies from the *ICAM1* and *TNF* polymorphic sites were obtained via the EM algorithm using the program HAPLOPOP (Majumdar and Majumder 1999). Estimation of standard diversity indices, mismatch distributions and statistics for testing neutrality, including coalescent simulations, were performed using the Arlequin (Schneider *et al*. 2000) and DnaSP (Rozas *et al*. 2003) packages.

A number of statistics for testing neutrality of mutations were computed, their tests of significance were performed by coalescent simulations (1000 simulation runs were performed for each case) using DnaSP. Observed and expected allele frequency spectra were computed using a computer program written by us. The expected number of sites at which the derived allele is present i times in a sample of size n was computed as, $\{s_n/a_n\}/i$, where s_n, denotes the observed number of sites and $a_n = \sum_{i=1}^{n-1}(1/i)$ (Watterson 1975; Fu 1995). Phylogenetic relationships among haplotypes were obtained by the Network software (http://www.fluxus-engineering.com/sharenet.htm).

Results and discussion

At the *ICAM1* locus, 29 variant sites were identified by resequencing the *ICAM1* gene in 208 individuals drawn from the 12 different ethnic groups. These have been reported in Sengupta *et al*. (2004) and are summarized in table 1. Tribal groups possess 22 of these 29 sites, while caste groups possess 21. Transition substitutions are more prevalent (64%) than transversions (35%); one insertion/deletion (indel) polymorphism was observed. All variant sites are biallelic, except for one site where a third T-allele appeared as GT heterozygotes in two Konkan Brahmins of Maharashtra (we removed these two individuals from the allele frequency estimation for that site, and also from haplotype reconstruction.) Interestingly, we observed two fairly common nonsynonymous SNPs (Glycine to Threonine) in our samples at nucleotide positions 13487 and 13542, that have not been reported earlier. The 29 SNPs detected by resequencing represent an overall occurrence of 1 SNP per 213 bp; 1 per 207 bp in introns and 1 per 177 bp in exons. The minor allele frequencies of six of the seven nonsynonymous SNPs are above 5% in one or more ethnic groups in our sample. Only five of 29 sites are shared among all the 11 ethnic groups inhabiting mainland India. A wide differences in allele frequencies across groups are observed (table 1). The Jarawas are monomorphic for 25 of 29 sites.

At the *TNF* locus, 12 SNPs (nine transitions and three transversions) and two indels were identified. Four new SNPs were discovered, of which three are present only in the Jarawa. One of these private sites among the Jarawa (C500T) is highly polymorphic, the frequency of the rarer allele at this site is 0.343. There is a wide variation in allele frequencies across populations (table 2).

Nucleotide diversity values ($\times 10^4$) across populations are very similar (2.5–5.0) for *ICAM1* (table 1), while there is slightly greater variability (1.5–5.4) for *TNF* (table 2). Unfortunately, no comparable data on neutral autosomal loci are available in Indian population groups. However, the nucleotide diversities in Indian groups estimated from mtDNA *HVS1* sequence data are in the range of 0.015–0.022 (Basu *et al*. 2003). Although it appears to be a reduction of nucleotide diversity at the *ICAM1* and *TNF* loci by two orders of magnitude compared to the mtDNA, it must be remembered that the rate of nucleotide substitution in the *HVS1* region of mtDNA is known to be substantially higher than in nuclear genomic regions. The average nucleotide sequence diversity in autosomal regions has been estimated to be about 7.5×10^{-4} (Sachidanandam *et al*. 2001), although it can vary by an order of magnitude across genomic regions (Reich *et al*. 2002). Thus, there is no significant evidence of reduction or enhancement of nucleotide diversity in the *ICAM1* and *TNF* genes.

However, when the nucleotide diversities were calculated separately for various regions of the genes (table 3), we found that there was a ten-fold reduction in nucleotide diversity in the promoter region of the *ICAM1* gene compared to the introns or exons of this gene, which exhibited similar levels of nucleotide diversity. Such a difference was, however not found in the case of the *TNF* gene. This finding is indicative of positive selection pressure in the promoter region of *ICAM1*.

Frequencies of haplotypes at the *ICAM1* locus were estimated (table 4; table 2 of appendix) using genotype frequency data of only those 17 polymorphic sites at which the frequency of the rarer allele exceeded 0.05 in at least one of the 12 populations. A total of 61 haplotypes are present, about 34% (19 of 61) of which are shared by at least two groups. Three haplotypes—H1 (21% in the pooled sample), H5 (14%) and H9 (12%)—are the most frequent ones. The southern-Indian Brahmin group, Iyer harbour the largest number of haplotypes (16), while the Jarawas harbour the lowest number (8). At the *TNF* locus, 36 haplotypes are observed (table 5; table 3 of appendix), of which 11 are shared among groups. Haplotype H1 frequency is 62.5% in the pooled sample. The vast majority of the haplotypes observed at both the loci have arisen by recombination.

To investigate the distribution of genetic variation at these two loci, we computed the site frequency spectra for tribal and caste groups, separately for the *ICAM1* (figure 2a,b) and

Table 1. Minor allele[a] frequencies at observed single nucleotide polymorphisms in and around the *ICAM1 gene in 12 ethnic groups of India and estimated nucleotide diversities.*

Position & nucleotide change[b]	Characteristics: Region	Amino acid change	Population code: BHU Tribe (13)	MZO Tribe (21)	MNP Caste (11)	SAN Tribe (16)	WBR Caste (16)	KAD Tribe (16)	IYR Caste (17)	MUR Tribe (16)	SBR Caste (16)	MRT Caste (15)	KBR Caste (16)	JAR Tribe (35)
A-785del	Promoter		0.038	0.024			0.031	0.063	0.088	0.031	0.031	0.033	0.094	
C-667T	Promoter				0.045									
A493C	Intron-1								0.029					
C503T	Intron-1		0.038	0.024			0.031	0.063	0.088	0.031	0.031	0.033	0.094	
T840C	Intron-1									0.031				
C958G	Intron-1									0.063				
C1066G	Intron-1			0.024	0.045	0.031	0.031	0.031		0.063				
G1076A	Intron-1								0.029					
G1110C	Intron-1			0.024	0.045	0.031		0.045		0.033				
G1195C	Intron-1								0.029					
C3642T	Intron-1												0.033	
G3757A	Exon-2	Q54Q						0.063						
A3762T	Exon-2	K56M		0.024	0.045	0.031	0.031	0.031		0.063				
G3784A	Exon-2	P63P	0.038	0.024										
C3965G[c]	Intron-2		0.385	0.286	0.227	0.438	0.375	0.281	0.412	0.375	0.313	0.214	0.462	0.206
T7175C	Intron-2				0.045	0.031								
G8880C	Intron-2		0.462	0.619	0.545	0.250	0.438	0.500	0.441	0.406	0.531	0.367	0.500	0.514
G12625A	Exon-3	R193Q								0.031				
C12739T	Intron-3			0.024										
G13014A	Exon-4	G241R					0.063						0.094	
C13430T	Exon-5	P352L									0.063			
C13470T	Exon-5	N365N			0.045	0.031		0.031		0.063				
G13487T	Exon-5	C371F			0.136			0.031	0.206			0.067	0.063	
G13542T	Exon-5	E389D	0.308	0.368*	0.364	0.063	0.031	0.156	0.147	0.375*	0.094	0.467*	0.219	
C13668T	Intron-5							0.031						
C13900T	Exon-6	T467T				0.063								
A13905G	Exon-6	E469K	0.192	0.286	0.409	0.531	0.594	0.406	0.471	0.531	0.469	0.533	0.313	0.486
G14195A	Exon-7 (3′UTR)						0.031	0.094	0.059	0.031	0.033		0.125	
C14588T	Exon-7 (3′UTR)		0.346	0.200	0.136	0.031	0.219	0.094	0.147	0.094	0.094	0.033	0.156	0.071
Nucleotide diversity (π) $\times 10^4$			4.103	4.060	4.833	3.335	3.866	4.688	4.978	5.026	3.528	3.625	5.026	2.465

Figures in parentheses indicate the numbers of individuals sampled.

[a] The allele with a lower frequency in the pooled sample is designated as the minor allele. Blank cells frequencies indicate zero frequencies;

[b] Nucleotide positions have been counted from the transcriptional start site. Nucleotides in italics are the derived ones, determined by comparing the human sequence with that of the chimpanzee. SNPs indicated in boldface have been considered for haplotype determination;

[c] A third allele *T* was detected as *GT* heterozygotes in two KBR individuals. These two individuals have been excluded from allele frequency estimation;

* Significantly ($P < 0.05$) deviated from Hardy–Weinberg equilibrium.

Table 2. Minor allele [a] frequencies at observed single nucleotide polymorphisms in and around the *TNF gene in 12 ethnic groups of India and estimated nucleotide diversities.*

Position & nucleotide change[b]	Region	Population code											
		BHU	MZO	MNP	SAN	WBR	KAD	IYR	MUR	SBR	MRT	KBR	JAR
		Tribe	Tribe	Caste	Tribe	Caste	Tribe	Caste	Tribe	Caste	Caste	Caste	Tribe
		(13)	(21)	(11)	(16)	(16)	(16)	(17)	(16)	(16)	(15)	(16)	(35)
A-572C	Promoter							0.059		0.031	0.067*	0.067*	
G-308A	Promoter	0.038				0.031	0.033	0.029	0.200*	0.061			
G-303A	Promoter												0.100
G-238A	Promoter	0.154	0.024			0.125	0.031	0.029	0.031	0.033	0.033	0.094	
T-77A	Promoter					0.031	0.125	0.029	0.031	0.067			
C-4T	Promoter												0.157
C56T	Exon1 (5′UTR)		0.048	0.182									
G420A	Intron1	0.154	0.024			0.125	0.031	0.029	0.031	0.033	0.033	0.094	
G489A	Intron1	0.077	0.167		0.219	0.094	0.094	0.118	0.187	0.133	0.2	0.062	
C500T	Intron1												0.343
AATG Indel at 625	Intron1		0.095			0.031	0.031		0.062	0.067			
AG Indel at 731	Intron1		0.048			0.031	0.094		0.031	0.033			0.129
A1304G	Intron3	0.154	0.024			0.156	0.187	0.059*	0.062	0.100	0.036	0.133	0.147
A2053C	Exon4 (3′UTR)				0.062*	0.031	0.087		0.031	0.067			0.143
Nucleotide diversity (π) $\times 10^4$		3.401	2.570	1.020	1.516	3.952	4.780	2.206	3.906	3.860	2.160	2.710	5.377

Figures in parentheses indicate the numbers of individuals sampled.

[a] The allele with a lower frequency in the pooled sample is designated as the minor allele. Blank cells frequencies indicate zero frequencies;

[b] Nucleotide positions have been counted from the transcriptional start site. Nucleotides in italics are the derived ones, determined by comparing the human sequence with that of the chimpanzee. SNPs indicated in boldface have been considered for haplotype determination;

[c] A third allele *T* was detected as *GT* heterozygotes in two KBR individuals; These two individuals have been excluded from allele frequency estimation;

* Significantly ($P < 0.05$) deviated from Hardy–Weinberg equilibrium.

Table 3. Nucleotide diversities ($\times 10^4$) in different regions of the *ICAM1* and *TNF* genes among tribal and caste populations of India.

Gene	Region	Tribe	Caste
ICAM1	Promoter	0.281	0.568
	Introns	4.178	4.193
	Exons	5.847	6.612
	Exons + UTRs	5.194	5.910
*TNF**			
	Introns	6.235	4.368
	UTR	3.597	2.716

*There are no polymorphic sites in the exons.

the *TNF* (figure 3a,b) loci. The differences between the observed and expected site frequency spectra are statistically significant for both *ICAM1* and *TNF*. The *P*-values corresponding to the Kolmogorov–Smirnov test statistic were < 0.001 for each of the loci (the observed site frequency spectrum is significantly different from that expected under neutrality for most populations for each of the two loci, details are not presented for brevity). At both loci, there is evidence of a significantly higher frequency of intermediate-frequency variants, which can result from balancing selection (Bamshad and Wooding 2003).

While various demographic processes can also affect the distribution of genetic variation, the effects of these processes are more-or-less uniform over the entire genome. On the other hand, natural selection affects functional and nonfunctional regions of the genome differentially (Bamshad and Wooding 2003). We have therefore, also computed the observed and expected site frequency spectra for the *ICAM1* locus separately for the promoter region, exons and introns for tribal (figure 2c, d, e) and caste (figure 2f, g, h) groups. Results for the *TNF* locus for these genomic regions are presented in figure 3 c,d,e for the tribal groups, and in figure 3 f,g,h for the caste groups. The site frequency spectra for these regions—promoter, exons and introns—show the same excess of intermediate-frequency variants compared to expectations under neutrality mentioned earlier, except for the intron region of *ICAM1* (figure 2e, h) where the pattern is similar to that expected for a neutral locus (Bamshad and Wooding 2003). These excesses are more pronounced for *TNF* than for *ICAM1*. No formal statistical tests for differences between the observed and expected site frequency

Table 4. Estimated frequencies of major[a] haplotypes at the *ICAM1* locus in 12 ethnic groups in India.

		Frequency[b]											
ID #	Haplotype[c]	BHU (26[d])	MZO (42)	MNP (22)	SAN (32)	WBR (32)	IYR (34)	KAD (32)	MUR (32)	SBR (32)	MRT (30)	KBR (32)	JAR (70)
H1	ACCCGACCGCCGGCAGC	0.417	0.371	0.227	0.104	0.046	0.284	0.244	0.187	0.266	0.040	0.423	0.069
H2	. . . . . GG . . . . . T	0.189	0.125		0.031	0.147				0.100			
H3	. . . . . GG . . . . T . . . T	0.118	0.085	0.090		0.031			0.046		0.035		
H4	. . . . . G . . . . T . G . .	0.107	0.132					0.025			0.064		
H5	. . . . . G . . . . . G . .	0.046		0.227	0.312	0.067	0.058	0.093	0.249	0.133	0.324		0.183
H6	. . . . . T	0.044	0.076	0.136							0.205		
H9	. . . . . G . .		0.083	0.136	0.051	0.352	0.068	0.062		0.193			0.272
H10	. . . . . GG		0.026	0.022	0.270	0.021	0.176	0.187		0.073	0.131	0.038	0.032
H14	. . . . . T . G . .		0.022		0.031			0.036	0.125	0.040	0.075		
H18	. . . . . GG . . . . T			0.022				0.031	0.140	0.060	0.011		
H25	. . . . . G					0.065		0.036			0.040		0.268
	Other 50 haplotypes[e]	0.079	0.080	0.140	0.201	0.271	0.414	0.286	0.253	0.135	0.075	0.539	0.176
	No. of haplotypes	8	11	11	10	14	16	15	13	10	11	14	8
	Haplotype diversity (±se)	0.781 ±.064	0.829 ±.046	0.902 ±.034	0.824 ±.044	0.860 ±.049	0.899 ±.034	0.885 ±.04	0.881 ±.032	0.873 ±.034	0.843 ±.045	0.824 ±.074	0.812 ±.022
	Tajima's D	0.943	0.318	0.728	−0.460	−0.444	0.849	−0.566	−0.165	−0.107	−0.437	−0.294	1.651
	Fu & Li's D^*	−0.062	−1.177	−0.811	1.268	0.192	−0.267	0.182	0.182	−0.146	0.204	0.182	1.219
	Fu & Li's F^*	0.273	−0.781	−0.504	1.530	0.355	−0.112	0.341	0.395	0.327	0.192	0.232	1.724*
	Fu's F_S	−0.962	−2.661*	−3.161*	−3.210*	−6.786*	−7.755*	−7.206*	−4.305*	−2.743*	−3.047*	−6.166*	−1.054

[a] A haplotype with an estimated frequency > 5 in the pooled sample is designated as a major haplotype; [b] Blank cells represent zero frequencies; [c] Based on 17 polymorphic sites identified in table 1; [d] Figures in parentheses indicate the numbers of sampled chromosomes; [e] The complete list of haplotypes is given in table 2 of appendix; * $P < 0.05$.

Table 5. Estimated frequencies of major[a] haplotypes in *TNF* gene in 12 ethnic groups in India.

		Frequency[b]											
ID #	Haplotype	BHU (26[d])	MZO (42)	MNP (22)	SAN (32)	WBR (32)	IYR (34)	KAD (32)	MUR (32)	SBR (32)	MRT (30)	KBR (32)	JAR (70)
H1[c]	AGGGTCCGGC6IAA	0.726	0.706	0.818	0.750	0.647	0.781	0.533	0.567	0.665	0.714	0.750	0.279
H2	. . . A . . . A . . . . G .	0.082				0.027			0.033	0.033	0.035	0.071	
H3	. . . . . A	0.043	0.103		0.188	0.065	0.031	0.100	0.133	0.100	0.178	0.071	
H6	. A	0.038				0.031			0.167	0.035			
H16	. . . . . DGC					0.031		0.100	0.033				0.118
H33	. . . . . T												0.338
H34	. . . . . T												0.147
	Other 29 haplotypes[e]	0.111	0.191	0.182	0.062	0.199	0.188	0.267	0.067	0.167	0.073	0.108	0.118
	No. of haplotypes	7	9	2	4	9	8	8	7	9	5	5	7
	Haplotype diversity (±se)	0.470 ±.119	0.486 ±.093	0.311 ±.106	0.413 ±.094	0.570 ±.102	0.701 ±.084	0.395 ±.110	0.650 ±.084	0.556 ±.106	0.470 ±.102	0.436 ±.112	0.775 ±.025
	Tajima's D	−0.591	−1.410	0.237	−0.092	−1.402	−0.887	−1.806	−1.519	−1.652	−1.262	−1.159	0.698
	Fu & Li's D^*	0.308	−1.002	0.992	0.026	0.562	−0.389*	0.431	0.250	−1.146	1.058	1.051	1.067
	Fu & Li's F^*	0.055	−1.181	0.939	0.066	0.022	−0.809	0.283	0.078	−0.558	0.618	0.730	1.251
	Fu's F_S	−2.867*	−5.763*	0.648	−1.281	−4.211*	−5.794*	−2.500	−2.214	−4.400*	−1.643	−1.467	0.006

[a] A haplotype with an estimated frequency > 5 in the pooled sample is designated as a major haplotype; [b] Blank cells represent zero frequencies; [c] Based on 14 polymorphic sites identified in table 2. Six indicates (AATG) copy number at position 625; I and D represent AG insertion and deletion, respectively, at position 731; [d] Figures in parentheses indicate the numbers of sampled chromosomes; [e] The complete list of haplotypes is given in table 3 of appendix; * $P < 0.05$.

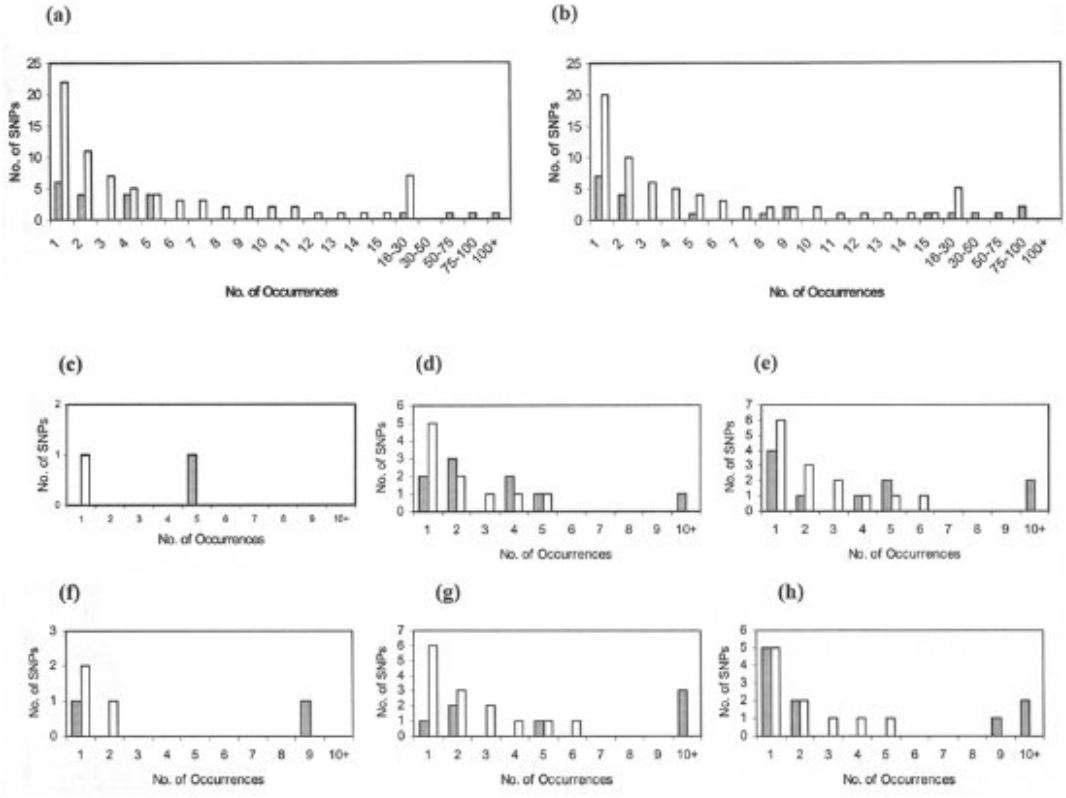

Figure 2. Observed (unfilled bars) and expected (filled bars) site frequency spectra among tribal and caste populations of India for *ICAM1*: (a) total gene–tribe, (b) total gene–casete, (c) promotor–tribe, (d) exon–tribe, (e) intron–tribe, (f) promoter–caste, (g) exon–caste and (h) intron-caste.

spectra were done for these separate genomic regions since the observed numbers of sites in these regions were small. However, although for the *ICAM1* promoter region a reduction in nucleotide diversity, consistent with positive selection was observed, there is no evidence of excess of low-frequency alleles that is expected under positive selection.

Various statistics, notably Tajima (1989) D, Fu and Li's (1993) $D*$ and $F*$, and Fu's (1997) F_S, have been proposed to examine various characteristics of the site frequency spectra for testing selective neutrality of mutations. Since population amalgamation may significantly affect the values of the test statistics, we have computed these statistics separately for each population (tables 4, 5). Our results show that for *ICAM1*, the F_S values are statistically significant for 10 of the 12 populations (table 4), while the other statistics are not (except for one F^* corresponding to the Jarawa). For *TNF*, the F_S values are statistically significant for five of the 12 populations (table 5). The other statistics are not significant for any of the populations, except for one D value. Through computer simulations, Fu (1997) has shown that F_S is particularly sensitive to demographic history, in the sense that if only F_S is significant while the other statistics are not, then it is more likely to be due to population expansion than natural selection. One way to resolve this confounding effect of positive or background selection and population growth is to investigate the mismatch distribution, which is expected to be smooth and unimodal in an expanding population (Rogers and Harpending 1992), but not necessarily so under selection. We have plotted the mismatch distributions among the tribes and castes using data of the *ICAM1* and *TNF* loci and, to obtain an independent calibration, also the data of the mtDNA *HVS1* region taken from Basu *et al.* (2003), pertaining to nine of the 12 populations considered here (figure 4). The mtDNA mismatch distributions are unimodal for both the tribes and castes with raggedness values (Rogers and Harpending 1992) of 0.02 and 0.03, respectively. The mismatch distributions for both *ICAM1* and *TNF* are also unimodal. The raggedness values for *ICAM1* are 0.06 for tribes and 0.05 for castes; the corresponding values for *TNF* are 0.07 and 0.04, respectively. The notable feature of the mismatch distributions for *TNF* is that these have modes at 1 and 0, respectively, for tribes and castes. This feature was not observed either for the *ICAM1* or for the mtDNA data. Thus, while the mismatch distributions for both the genes (*ICAM1* and *TNF*) are in good agreement with a population expansion model, the distribution for *TNF* is similar to that expected under a recent population expansion. However, since demographic history is a characteristic of the population, the implications of these distinct patterns are unclear. It is probably not due to a recent selective sweep operating at the *TNF*

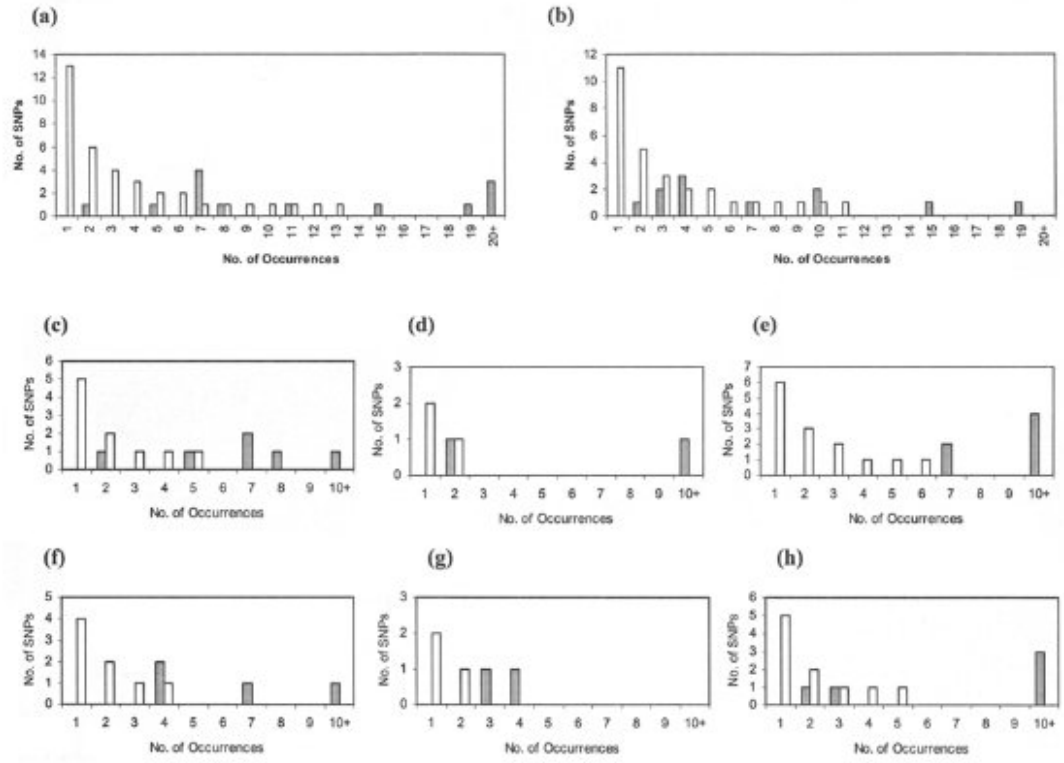

Figure 3. Observed and expected site frequency spectra among tribal and caste populations of India for *TNF*: (a) total gene–tribe, (b) total gene–caste, (c) promoter–tribe, (d) exon-tribe, (e) intron–tribe, (f) promoter–caste, (g) exon–caste and (h) intron–caste.

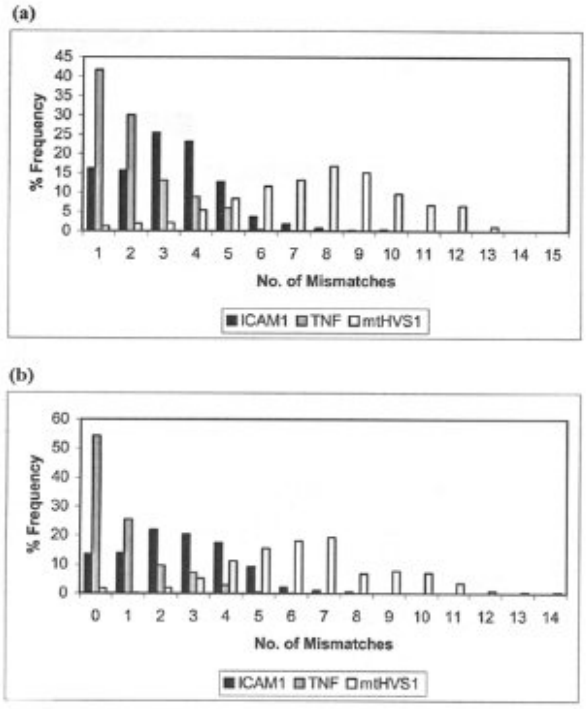

Figure 4. Mismatch distributions pertaining to, (a) tribal and (b) caste populations for *ICAM1* (■), *TNF* (▣) and mitochondrial *HVS1* (□)

locus, because if there had been a recent selective sweep, one would expect an excess of rare frequency alleles, while in fact an excess of intermediate frequency alleles is observed at this locus.

We have also examined the F_{ST} values at these loci, and have computed them separately for promoter, intronic and exonic polymorphisms for *ICAM1* (figure 5a) and *TNF* (figure 5b). The F_{ST} value over all polymorphic sites for *TNF* (0.08; $P < 0.001$) is marginally higher than that for *ICAM1* (0.06; $P < 0.05$). These values are only slightly higher than observed (0.04) for neutral autosomal loci (Basu *et al.* 2003). Since balancing selection is expected to reduce the F_{ST} value compared to neutral loci, our finding does not indicate any strong effect of balancing selection at the loci under study. There is, however, considerable variation in F_{ST} values between tribes and castes: for *ICAM1* these values are, respectively 0.05 and 0.02, while for *TNF*, the values are 0.09 and 0.02, respectively. All the F_{ST} values are statistically significant ($P < 0.05$). The tribal groups are more differentiated than that of the caste groups, which may be a result of their isolation for a longer period of time than that of the caste groups. Further, the locus-specific F_{ST} values are highly structured by the position within the gene. For *ICAM1* (figure 5a), the F_{ST} values for polymorphic loci that are in exons are substantially higher than those located in the promoter region or in the introns. For *TNF* (figure 5b), the

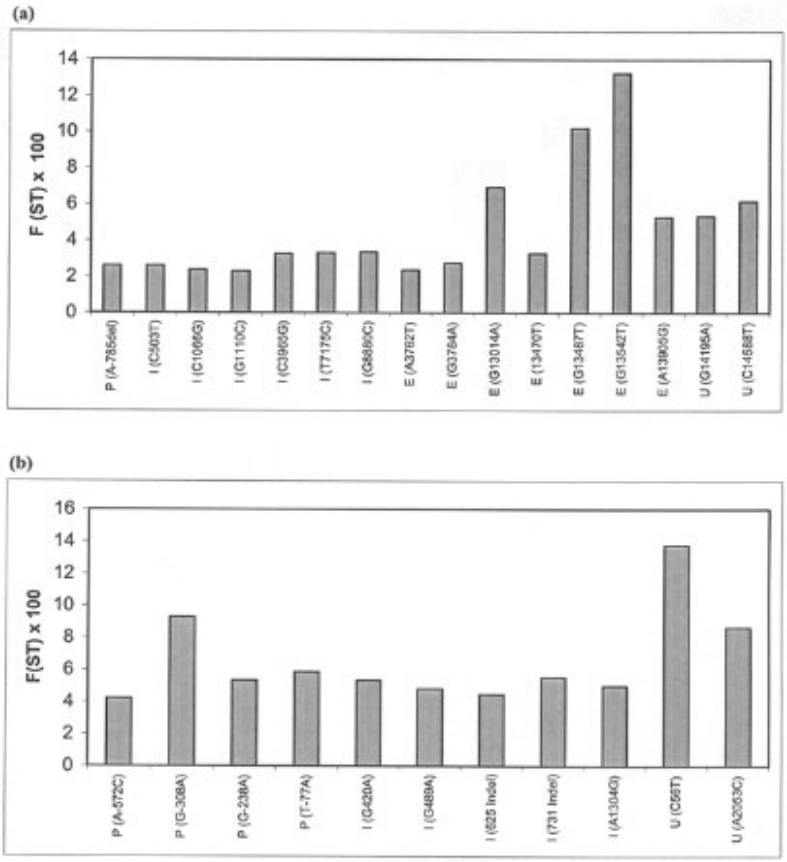

Figure 5. F_{ST} values for various polymorphisms observed at the (a) *ICAM1* and (b) *TNF* loci. P, promoter, I, intronic; E, exonic; U, 3′ untranslated region of polymorphisms.

exonic polymorphisms are located only in the untranslated regions of the exons—these loci have higher values than those located in the introns or in the promoter. The only exception is the *G-308A* promoter polymorphism at which a high F_{ST} value was observed; this polymorphism is known to be associated with the susceptibility to severe malaria, leishmaniasis, scarring trachoma and lepromatous leprosy (Knight and Kwiatkowski 1999). One reason for observing higher F_{ST} values for exonic polymorphisms like that of the *TNF* promoter polymorphism *G-308A*, is that these polymorphisms may also be associated with certain diseases that possibly have variable prevalence across populations. Thus, these loci may be under selective influence in some, but not all populations, resulting in wide differences in allele frequencies across populations and consequently higher F_{ST} values. Alternatively, high F_{ST} values may simply be because of genetic drift which, however, is unlikely because the observed pattern of F_{ST} values by genomic region (promoter, exon, intron) would then not be expected.

To further examine whether the observed patterns of genetic variation, especially at the *ICAM1* locus, are consistent with population expansion, we have constructed median-joining networks of the major haplotypes observed at these loci (figure 6). Under a population expansion model, a star like phylogeny of haplotypes is expected (Takahata and Nei 1990; Rogers and Harpending 1992). Balancing selection, on the other hand, is expected to retain multiple lineages for a long time, resulting in a network in which there are some high-frequency clusters and some low-frequency clusters with long branches (Takahata and Nei 1990). Such a pattern is observed for *ICAM1*, and to some extent for *TNF*. However, for *TNF*, the network is essentially star like, consistent with population expansion, as earlier inferred from the mismatch distributions.

To summarize, nucleotide diversity levels in the genes or their component regions (promoter, exon, intron) do not show any statistically significant evidence of reduction or enhancement compared to other autosomal genes. The only exception is the promoter region of *ICAM1*, where we have noted a significant reduction of nucleotide diversity consistent with positive selection. If a genomic region is under positive selection, then it is expected that there will be a significant excess of low-frequency alleles compared to neutral expectations. This, however, was not observed. In fact, consistent with balancing selection, at both the loci and in their component regions there were significant excesses of

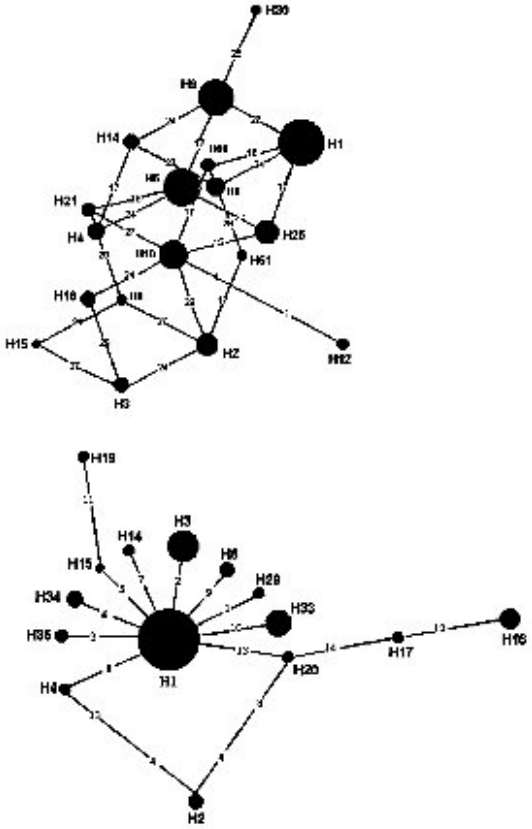

Figure 6. Median-joining networks depicting phylogenetic relationships among haplotypes observed in at least three individuals at the, (a) *ICAM1* and (b) *TNF* loci. [The identification numbers of haplotypes (nodes) and of polymorphisms given on the edges joining the nodes are provided in table 2 and 3 of appendixes and for *ICAM1* and *TNF*, respectively. Many apparent reticulations are due to recombination.]

intermediate-frequency alleles; these excesses were greater for *TNF* than for *ICAM1*. Thus, we did not observe a pattern of nucleotide variation that is consistent with a simple and uniform mode of natural selection in either genes. Since it is known (Bamshad and Wooding 2003) that demographic histories of populations can result in patterns of nucleotide variation that are similar to those expected under various models of natural selection, we have calculated statistics relevant for inferring demographic histories. Fu's (1997) F_S statistic and unimodality of mismatch distributions indicated that both the tribal and caste populations underwent significant population expansion. The median-joining network of haplotypes at the *TNF* locus was star like, consistent with population expansion, but that of the haplotypes at the *ICAM1* locus was not. The coefficient of population differentiation F_{ST} also did not show any significant excess or reduction, although higher values were observed for the promoter and exon regions of both *ICAM1* and *TNF*, consistent with natural selection. Thus, we see that the patterns of nucleotide variation in these genes, that perform related functions, is complex and is not consistent with a simple model of selection. Our results indicate that both natural selection and differential demographic histories have jointly contributed to the observed patterns of nucleotide diversity and haplotype structure. The effect of natural selection seems more pronounced in the promoter regions of these genes, although it is unclear whether the selective pressure is balancing or positive.

The complexity of our results is comparable to those found at the Duffy blood group (*DARC*) locus (Hamblin and Di Rienzo 2000). The *TNF* gene is located between genes that comprise the *HLA* gene cluster on chromosome six, and there are functionally important genes (e.g., intercellular adhesion molecule genes, erythropoietin receptor and low-density lipoprotein receptor) located around the *ICAM1* gene on chromosome 19. The pattern of selection operating on the *HLA* gene cluster is known to be complex (Takahata *et al*. 1992; Klein *et al*. 1993; Satta *et al*. 1994). Since the *TNF* gene is located within this cluster, it is possible that hitchhiking effects may have contributed to the pattern of nucleotide sequence variation in the *TNF* gene. The same phenomenon may have also operated on the *ICAM1* gene, if indeed selective effects have been strong on the nearby genes. Moreover, multiple distinct episodes of selection may have operated on the *TNF* and *ICAM1* genes, in view of their central importance in interacting with pathogens and in other noninfectious diseases. The $ICAM1^{Kilifi}$ variant has been found to predispose individuals in Kenya to cerebral malaria (Fernandez-Reyes *et al*. 1997). Although this variant was found in several populations in our study, its frequency is much lower than in Kenya. Similarly, many variants at the *TNF* locus that have been found to be associated with various diseases, both infectious and noninfectious (Gimenez *et al*. 2003), are found in widely differing frequencies (e.g., *G-238A*, *G-308A*), or not found at all in Indian population groups. Thus, it is possible that temporal and spatial variations in prevalence of pathogens and diseases, together with variable ancestral histories of population groups, have resulted in the complex pattern of nucleotide sequence variation at these two loci.

Appendix

Table 1. Sequences of oligonucleotide primers and protocols used for amplification and sequencing of *ICAM1* and *TNF* genes.

Primer ID	Oligonucleotide Sequence (5′ → 3′)		Amplicon size
	Forward	Reverse	
*ICAM*1.1/1.2	GGAGTCTCAGTTTACCGCTTTG	CTACCTAAGCATGCATGACCTG	590
*ICAM*2.1/2.2	GTCTTGTTAAGGCTGTGCCTCCAG	TGACCCTACGAGCAAGTGGCAAAG	475
*ICAM*3.1/3.2	CAGTTCGTCTGTTAGGCAGGCAG	CTCAGCAGCCTAGGTCACATACG	518
*ICAM*4.1/4.2	GTGTTCTAGGCGTATGTGACCT	CCTCTGGCTTCGTCAGAATCAC	543
*ICAM*5.1/5.2	GTGACCATCTACAGTAAGAAGG	GACTTGGAGGCAAGACCTTATG	432
*ICAM*6.1/6.2	CTTCGTGTCCTGTGTGAGTGG	GGTAGGTGTAGCTGCATGGCA	745
*ICAM*7.1/7.2	ATTTAGTGCATGAGCCTGGGTTCGAG	CAAAGCGGTAAACTGAGACTCCAGG	638
*ICAM*8.1/8.2	GGTATGCATGCTTAGGTAGCTGT	CCAAGTAGAAGCAGCCCTGGACTT	517
*ICAM*9.1/9.2	TTCCAGAGCTGACTTATCCGTG	TTGGATCAGGTCCCTCAGATTC	714
*ICAM*10.1/10.2	AAGTAGCTTTGGGATTAGCCTTG	TTTCCAGAAGCTGCTGGGAATGT	288
*ICAM*11.1/11.2	GAGGTTGGCAGAGCCTTGAA	CTCTTACCACCTCCTATAGACT	387
*ICAM*12.1/12.2	ATATGCCATGCAGCTACACCTA	CACTCTCCTGCAGTGTACAACCT	573
*TNF*1.1/1.2	CTTAACGAAGACAGGGCCAT	ATTTGTGTGTAGGACCCTGGA	592
*TNF*2.1/2.2	GAAGGAAACAGACCACAGACCT	CTTTCAGTGCTCATGGTGTCCT	566
*TNF*3.1/3.2	GAAAGGACACCATGAGCACTG	CACCTTCCAGGCATTCAACAG	674
*TNF*4.1/4.2	CTCAGGGAAAGAGTCGTTGAATGC	CCAAGTTCCAAGACACATCCTCAG	477
*TNF*5.1/5.2	GTGACAAGCCTGTAGCCCATGTTG	TGATGGTGTGGGTGAGGAGCAAT	517
*TNF*6.1/6.2	CGTGGAGCTGAGAGATAACCAG	TTGCCAGCACTTCACTGTGCAG	513

The reaction mixture for amplification comprised 100–200 ng of genomic DNA, 50 ng of each primer, 100 μM of dNTPs, 2 U of *Taq* DNA polymerase and 2.0 mM magnesium chloride in a total volume of 15 μl. After the initial denatuation step of 95°C for 10 min, a touchdown regime was used. The PCR cycling conditions of initial 14 cycles were (94°C for 20 s, 63°C for 30 s (−0.5 °C per cycle) and 72°C for 1min), followed by another 25 cycles of (94°C for 20 s, 56°C 45 s and 72°C for 1min), and a final extension of 5 min at 72°C.

Table 2. Sixty-one *ICAM1* haplotypes present in 12 ethnic groups of India.

Nt Position	A-75Ade1	C503T	C958G	C1066G	G3757A	A3762T	C3965G	G8888C	G13014A	C13430T	C13470T	G13487T	G13542T	C13900T	A13905G	G14195A	C14588T
ID	1	4	6	7	1	1	1	1	2	2	2	2	2	2	2	2	2
	2	3	5	7	0	1	2	3	4	6	7	8	9				
H1	A	C	C	C	G	A	C	C	G	G	C	A	G	C			
H2	.	.	.	.	.	.	G	G	.	.	.	.	.	T			
H3	.	.	.	.	.	.	G	G	.	.	.	.	T	.	.	T	
H4	.	.	.	.	.	.	.	G	.	.	.	.	T	.	G	.	
H5	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.		
H6	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	
H7	D	T	.	.	.	.	G	G	.	.	.	.	T	.	.	.	
H8	.	.	.	.	.	.	G	G	.	.	.	.	.	G	.	T	
H9	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		
H10	.	.	.	.	.	.	G	G	.	.	.	.	.	.	.		
H11	.	.	.	G	.	T	G	G	.	.	.	T	.	G	.		
H12	D	T	.	.	.	.	G	G	.	.	.	.	.	.	.		
H13	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	
H14	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	
H15	.	.	.	.	.	.	G	G	.	.	.	.	T	.	G	T	
H16	.	.	.	.	.	.	G	G	.	.	.	T	T	.	.	T	
H17	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	
H18	.	.	.	.	.	.	G	G	.	.	.	.	T	.	.	.	
H19	.	.	.	G	.	T	G	G	.	.	T	T	.	G	.		
H20	.	.	.	G	.	T	G	G	.	.	T	T	T	.	G	.	
H21	.	.	.	.	.	.	G	G	.	.	.	.	.	G	.		
H22	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	
H23	.	.	.	.	.	.	G	G	.	.	.	.	T	.	G	.	
H24	.	.	.	G	.	T	G	G	.	.	T	.	.	G	.		
H25	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.		
H26	.	.	.	.	.	.	.	A	.	.	.	.	.	G	.		
H27	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.		
H28	.	.	.	G	.	T	G	G	.	.	.	.	.	.	.		
H29	.	.	.	.	.	.	.	G	A	.	.	.	.	G	.		
H30	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.		
H31	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.	T	
H32	.	.	.	.	.	.	.	G	.	.	.	T	T	.	.	.	
H33	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.		
H34	.	.	.	.	A	.	.	.	.	.	.	T	.	G	.	T	
H35	D	T	.	G	.	T	G	G	.	.	T	.	.	G	.	T	
H36	D	T	.	.	.	.	.	G	.	.	.	.	.	G	.		
H37	.	.	.	.	.	.	G	G	.	.	.	T	T	.	.		
H38	.	.	.	.	.	.	.	G	.	.	.	T	.	G	.	T	
H39	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.		
H40	.	.	.	.	.	.	.	.	.	.	.	T	T	.	GA	.	
H41	D	T	.	.	.	.	.	G	.	.	.	.	.	G	.	T	
H42	D	T	.	.	.	.	.	G	.	.	.	.	.	G	.	T	
H43	.	.	.	.	.	.	.	G	G	.	.	.	T	.	.	A	.
H44	.	.	.	.	.	.	.	G	G	.	.	.	T	T	.	G	.
H45	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	
H46	.	.	.	.	.	.	.	G	G	.	.	.	.	.	.	A	T
H47	.	.	G	.	.	.	.	.	.	.	.	.	.	G	.		
H48	.	.	G	.	.	.	G	.	.	.	.	.	.	G	.		
H49	D	T	.	.	.	.	G	G	.	.	.	.	.	G	.		
H50	.	.	.	G	.	T	G	G	.	.	T	.	.	G	.	T	
H51	.	.	.	.	.	.	G	G	.	T	.	.	.	G	.		
H52	.	.	.	.	.	.	.	G	.	.	.	.	.	G	A		
H53	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	G	.
H54	D	T	.	.	.	.	G	G	.	.	.	.	T	.	G	.	
H55	.	.	.	.	.	.	G	G	A	.	.	.	.	.	T		
H56	.	.	.	.	.	.	.	G	.	.	.	.	T	.	G	A	
H57	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	
H58	.	.	.	.	.	.	G	G	A	.	.	.	T	.	G	.	
H59	.	.	.	.	.	.	G	G	A	.	.	.	.	G	.	T	
H60	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.		
H61	.	.	.	.	.	.	.	G	.	.	.	.	.	.	T		

The numbers in italics correspond to the 17 polymorphic sites in the 5′ → 3′ direction in the *ICAM1* gene (table 1) used to reconstruct the haplotypes.

Table 3. Thirty-six *TNF* haplotypes present in 12 ethnic groups of India.

Nt Position	A-572C	G-308A	G-303A	G-238A	T-77A	C-4T	C56T	G420A	G489A	C500T	Indel at 625	Indel at 731	A1304G	A2053C
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
											0	1	2	3 4
H1	A	G	G	G	T	C	C	G	G	C	6	I	A	A
H2	.	.	.	A	.	.	.	A	.	.	.	.	G	.
H3	.	.	.	.	.	.	.	A	.	.	.	.	.	.
H4	.	.	.	.	.	.	.	A	.	.	.	.	.	.
H5	.	.	.	A	.	.	.	.	.	.	.	.	G	.
H6	.	A	.	.	.	.	.	.	.	.	.	.	.	.
H7	.	.	.	A	.	.	.	A	A	.	.	.	G	.
H8	.	.	.	.	.	.	.	.	.	.	5	.	.	.
H9	.	.	.	.	.	.	.	.	.	D	.	.	.	.
H10	.	.	.	A	.	.	.	A	.	.	D	G	.	.
H11	.	.	.	.	.	.	T	.	.	.	5	.	.	.
H12	.	.	.	.	.	.	T	.	A	.	.	.	.	.
H13	.	.	.	.	.	.	.	.	.	A	.	5	.	.
H14	.	.	.	.	.	.	T	.	.	.	.	.	.	.
H15	.	.	.	.	A	.	.	.	.	.	.	.	.	.
H16	.	.	.	.	.	.	.	.	.	.	D	G	C	
H17	.	.	.	.	.	.	.	.	.	.	G	C		
H18	.	A	.	.	.	.	.	.	.	.	C			
H19	.	.	.	A	.	.	.	.	.	5	.	.	.	
H20	.	.	.	.	.	.	.	.	.	G	.			
H21	.	.	.	.	.	.	.	A	.	.	G	.		
H22	C	.	.	A	.	.	.	.	.	.	.			
H23	.	.	.	A	.	.	.	A	.	.	.			
H24	.	A	.	.	.	.	.	A	.	.	.			
H25	.	.	.	.	.	.	.	.	.	.	C			
H26	.	.	.	.	.	.	.	A	.	.	C			
H27	.	.	.	A	.	.	.	A	.	.	.			
H28	.	A	.	.	.	.	.	.	5	.	.			
H29	C	.	.	.	.	.	.	.	.	.	.			
H30	.	.	.	.	.	.	.	A	.	.	D	G	C	
H31	.	A	.	.	A	.	.	.	.	5	.	.	.	
H32	C	.	.	.	.	.	.	A	.	.	.			
H33	.	.	.	.	.	.	.	.	T	.	.	.		
H34	.	.	.	.	.	T	.	.	.	.	.			
H35	.	.	A	.	.	.	.	.	.	.	.			
H36	.	.	A	.	.	.	.	.	.	D	G	.		

Acknowledgements

This study was supported in part by a grant from the Department of Biotechnology, Government of India, to PPM.

References

Bamshad M. and Wooding S. P. 2003 Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **4**, 99–111.

Basu A., Mukherjee N., Roy S., Sengupta S., Banerjee S., Chakraborty M. *et al.* 2003 Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* **13**, 2277–2290.

Berendt A. R., Simmons D. L., Tansey J. C., Newbold M. and Marsh M. 1989 Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for *Plasmodium falciparum*. *Nature* **341**, 57–59.

Bjornsdottir U. S. and Cypcar D. M. 1999 Asthma: An inflammatory mediator soup. *Allergy* **54**, 55–61.

Dobbie M. S., Hurst R. D., Klein N. J. and Surtees R. A. 1999 Upregulation of intercellular adhesion molecule-1 expression on human endothelial cells by tumour necrosis factor-alpha in an in vitro model of the blood-brain barrier. *Brain Res.* **830**, 330–336.

Fernandez-Arquero M., Arroyo R., Rubio A., Martin C., Vigil P., Conejero L. *et al.* 1999 Primary association of a *TNF*-gene polymorphism with susceptibility to multiple sclerosis. *Neurology* **53**, 1361–1363.

Fernandez-Reyes D., Craig A. G., Kyes S. A., Peshu N., Snow R. W., Berendt A. R. *et al.* 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.

Fu Y.-X. 1995 Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**, 172–197.

Fu Y.-X. 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

Fu Y.-X. and Li W. -H. 1993 Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Gimenez F., de Lagerie S. B., Fernandez F., Pino P. and Mazier D. 2003 Tumor necrosis factor α in the pathogenesis of cerebral malaria. *Cell. Mol. Life Sci.* **60**, 1623–1635.

Grimsley C., Mather K. A. and Ober C. 1998 *HLA-H*: a pseudogene with increased variation due to balancing selection at neighboring loci. *Mol. Biol. Evol.* **15**, 1581–1588.

Hamblin M. T. and Di Rienzo A. 2000 Detection of the signature of natural selection in humans: evidence from the Duffy blood group. *Am. J. Hum. Genet.* **66**, 1669–1679.

Hill A. V. 1992 Malaria resistance genes: a natural selection. *Trans. R. Soc. Trop. Med. Hyg.* **86**, 225–226.

Kawasaki A., Tsuchiya N., Hagiwara K., Takazoe M. and Tokunaga K. 2000 Independent contribution of *HLS-DRB1* and *TNF alpha* promoter polymorphisms to the susceptibility to Crohn's disease. *Genes Immun.* **1**, 351–357.

Klein J., Satta Y., O'hUigin C. and Takahata N. 1993 The molecular descent of the major histocompatibility complex. *Annu. Rev. Immunol.* **11**, 269–295.

Knight J. C. and Kwiatkowski D. 1999 Inherited variability of tumor necrosis factor production and susceptibility to infectious diseases. *Proc. Assoc. Am. Physicians* **111**, 290–298.

Kwiatkowski D., Molyneux M. E., Stephens S., Curtis N., Klein N., Pointaire P. *et al.* 1993 Anti-TNF therapy inhibits fever in cerebral malaria. *Q. J. Med.* **86**, 91–98

Majumdar P. and Majumder P. P. 1999 HAPLOPOP: A computer program package to estimate haplotype frequencies from genotype frequencies via the EM algorithm. *AHGU Tech Rep* 1/99, Indian Statistical Institute, Kolkata.

McGuire W., Knight J. C., Hill A. V., Allsopp C. E., Greenwood B. M. and Kwiatkowski D. 1999 Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles. *J. Infect. Dis.* **179**, 287–290.

Meager A. 1999 Cytokine regulation of cellular adhesion molecule expression in inflammation. *Cytokine Growth Factor Rev.* **10**, 27–39.

Negoro K., Kinouchi Y., Hiwatashi N., Takahashi S., Takagi S., Satoh J. et al. 1999 Crohn's disease is associated with novel polymorphisms in the 5′-flanking region of the tumor necrosis factor gene. *Gastroenterology* **117**, 1062–1068.

Paloske B. L. and Howard R. J. D. 1994 Malaria, the red cell, and the endothelium. *Annu. Rev. Med.* **45**, 283–295.

Reich D. E., Schaffner S. F., Daly M. J., McVean G., Mullikan J. C., Higgins J. M. *et al*. 2002 Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genet.* **32**, 135–142.

Rogers A. R. and Harpending H. 1992 Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Bio. Evol.* **9**, 552–569.

Rozas J., Sanchez-DelBarrio J. C., Messeguer X. and Rozas R. 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Sachidanandam R., Weissman D., Schmidt S. C., Kakol J. M., Stein L. D., Marth G. *et al*. 2001 A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933.

Satta Y., O'hUigin C., Takahata N., and Klein J. 1994 Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci.* USA **91**, 7184–7188.

Schneider S., Roessli D. and Excoffier L. 2000 Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Switzerland, Geneva, Switzerland.

Sengupta S., Farheen S., Mukherjee N. Dey B., Mukhopadhyay B., Sil S. K. *et al*. 2004 DNA Sequence variation and haplotype structure of the *ICAM-1* and *TNF* genes in twelve ethnic groups of India reveal patterns of importance in designing association studies. *Ann. Hum. Genet.* **68**, 574–588.

Singh K. S. 1992 *People of India: an introduction*. South Asia Books, New Delhi.

Striz I., Mio T., Adachi Y., Romberger D. J. and Rennard S. I. 1999 IL-4 induces ICAM-1 expression in human bronchial epithelial cells and potentiates *TNF-alpha*. *Am. J. Physiol.* **277**, L58–L64.

Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Takahata N. and Nei M. 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**, 967–978.

Takahata N., Satta Y. and Klein J. 1992 Polymorphism and balancing selection at the major histocompatibility loci. *Genetics* **130**, 925–938.

Thapar R. 2003 *Early India: from the origins to AD 1300*. University of California Press, Berkeley, CA.

Thio C. L., Goedert J. J., Mosbruger T., Vlahov D., Strathdee S. A., O'Brien S. J. *et al*. 2004 An analysis of tumor necrosis factor alpha gene polymorphisms and haplotypes with natural clearance of hepatitis C virus infection. *Genes Immun.* **5**, 294–300.

Turner G. D., Morrison H., Jones M., Davis T. M., Looareesuwan S., Buley I. D. *et al*. 1994. An immunohistochemical study of the pathology of fatal malaria: evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am. J. Pathol.* **145**, 1057–1069.

Watterson G. A. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276.

Zeggini E., Thomson W., Kwiatkowski D., Richardson A., Ollier W. and Donn R. 2002 Linkage and association studies of single-nucleotide polymorphism-tagged tumor necrosis factor haplotypes in juvenile oligoarthritis. *Arthritis Rheum.* **46**, 3304–3311.

Received 26 February 2007