Genetic Affinity Between Two Ethnically Diverse Caste Groups of North India: A Study Based Upon 15 Microsatellite Loci

M. Tandon[1], T.S. Vasulu[2], R. Trivedi[1] and V.K. Kashyap[1]*

1. DNA Typing Unit, Central Forensic Science Laboratory, Kolkata, India
2. Anthropology and Human Genetics Unit, Indian statistical Institute, Kolkata, India

ABSTRACT This study attempts to understand the genetic structure and affinity among two major ethnic caste groups viz., Jat and Kurmi of north India in order to examine the effect of geographical as well as occupational proximity among these groups and their genetic relationship with two other predominant populations of the same region. The genetic structure and diversity of the studied populations was examined by analyzing polymorphism at 15 microsatellite loci. The Jat and Kurmi populations displayed high degree of heterozygosity with average heterozygosity in the range of 0.788 to 0.796 at all the studied loci. The most frequent allele common for the populations was observed at four loci (D21S11, vWA, D8S1179 and D16S539) while the remaining eleven loci showed differences in their allele frequency. The level of intra-population genetic diversity is found significantly high. The coefficient of gene differentiation wars found to be low (average G_{ST} is 1.1 per cent and ranges between 0.2% at D16S539 to 2.5% at D13S317) across the loci, indicating close affinity between the two occupational caste groups. Clustering pattern based on neighbouring joining tree reveals two clear clusters, one for Kurmi and Jat and other for Khatri and Thakur populations. The results strongly demonstrate that the geographical as well as occupational proximity attributes significantly to genetic affinity between the studied populations.

INTRODUCTION

The vast spectrum of genetic diversity in Indian populations is a subject of great interest to population geneticists and evolutionary biologists. Studies based on morphological, classical genetic marker data show greater diversity among Indian populations (Bhasin and Walter 2001), comparable to global populations. The enormous genetic diversity among various subcastes and castes is primarily guided by ethno-historical factors, geographical proximity and socio-cultural rules of marriage practices (Malhotra 1979; Malhotra and Vasulu 1993; Cavalli Sfroza et al. 1994; Bhasin et al. 1994). These studies indicate that in some regional populations ethno-historical factors play a significant role than other factors while in certain cases geographical proximity constitute a strong force than socio-cultural factors.

Although in India most of the Hindu communities follow strict endogamy but gene flow among the people with similar occupation,

Address for correspondence: Dr. V. K. Kashyap, Central Forensic Science Laboratory 30, Gorachand Road, Kolkata, West Bengal, India
Telephone: +91-33-2284-1638;
Fax: +91-33-2284-1753;
E-mail: vkk2k@hotmail.com

tradition and inhabiting same geographical territories cannot be ruled out. In general, among tribal populations isolation and endogamy play a greater role in maintaining the genetic diversity whereas in some regional populations, especially among castes and subcastes in northern parts of the country, geographical proximity is observed as major influencing factor than socio-cultural variables.

The Gangetic belt in northern parts of the country shows an interesting geographical feature with least geographical barriers limited to rivers to hillocks and the rich delta region which were suitable for human inhabitation since prehistoric times (Ahmed and Saxena 2001). Conducive conditions for human inhabitation found in this region led to several waves of migration and settlement and might have been the directive force in the formation of various castes, subcastes and scheduled castes. Historically this region has played a major role in the socio-political scenario and has seen the rise and fall of several dynasties from pre-Mauryan period (325-200 BC) to Moguls and British occupation, which had changed the population profile of the entire country. Because of several historical wars, epidemiological events and natural calamity, this region had experienced demographic upheaval

of population expansion, bottlenecks and wide migration and settlement of various ethnic, racial, cultural, religious populations from different parts of the world. Further socio-religious movements of Hinduism, Buddhism, Islamic, Jainism and Christian elements had influenced the population structure since the beginning of human occupation. Therefore, the present-day population of the region is the successful residual of the past demographic upheavals and has the genetic-imprint of the historical events. The genetic diversity or affinity among the major dominating castes and subcastes of this region is due to the relative roles of historical events, geographical proximity and socio-cultural variables of the population structure. The relative roles of these influencing factors especially whether socio-cultural or geographic or historical events had played a crucial role in shaping the genetic structure of the populations has been one of the interesting research investigations among population biologists. Though only limited studies have been conducted to characterize the genetic structure and to understand the micro-evolutionary process of these regions (Bhasin et al. 1994; Papiha 1996; Mastana 1999; Bhasin and Walter 2001) yet several interesting populations are to be explored. For example, morphological diversity among diverse castes and subcastes showed that the caste hierarchy is not strictly reflected in their biological affinity (Mahalanobis et al. 1941). The sub castes and other major castes are genetically more similar irrespective of socio-cultural or religious barriers, whereas scheduled castes show greater diversity from Brahmins and upper caste groups (Mahalanobis et al. 1941). Further genetic variation studies show ABO*B allele more frequent in higher castes than in lower caste groups, whereas other traditional polymorphic system show low levels of genetic differentiation among different caste groups (Bhasin et al. 1992; Bhasin et al. 1994; Papiha et al. 1972). And an extensive study based on 24 genetic markers (9 blood groups, 11 red cell enzymes and 4 serum proteins) among 6 groups showed low genetic variation. And the FIS estimate was found to be higher than FST values, possibly due to the influence of high rates of endogamy of the caste structure (Lanchbury et al. 1996).

With the development of molecular tools of genomic analysis, which has an immense power to address the origin and microevolutionary process of the human populations, the above finding can now be addressed with much more clarity and depth that was not possible with the erstwhile classical genetic markers (Bowcock et. al. 1994; Deka et al. 1995a; Jorde et al. 1997; Perez-Lezaun et al. 1997, Calafell et al.1998; Kimmel et al. 1998). In the present study microsatellite markers were chosen, as they are highly informative and provide greater insights and understanding of population structure (Deka et al. 1999; Mukherjee et al. 1999; Dutta and Kashyap 2001; Destro-Bistol et al. 2000; Reddy et al. 2001). However studies on genetic polymorphism based on microsatellite markers in this region are very rare. Here we have considered two predominant socio-economic caste populations of the Uttar Pradesh State viz., the Jat and Kurmi to investigate their genomic structure and the affinities. Jat and Kurmi along with Thakur and Khatri are some of the major populations widely spread in western parts of Uttar Pradesh adjoining Punjab and Haryana. Historically Jat population appears to have come from Central Asia and have widespread and settled. The origin and history of Kurmi is not clear however the ethno historical accounts suggests that they migrated from central India and possibly originally belong to austroloid ethnic tribes several centuries ago. Jat and Kurmi are two compatriot caste, share same occupation but with diverse ethnic and migration background. Thakurs are believed to have migrated from western Eurasia during early medieval period. They belong to Indo-Caucasoid ethnic group and are traditional warriors and rulers. Khatri are migrants of western India, they claim to be Rajput and believe that their community name is corrupt form of Kshatriya. There are hardly any study on the genetic structure of these populations to investigate their affinity and diversity. The present study is perhaps the first report on their genetic structure and is part of nationwide analysis of genomic diversity of Indian populations. Jat and Kurmi are widely distributed in northern region and their large population sizes would naturally result to the promotion of breeding populations possibly with little external gene flow.

The basic objective of present study is to investigate whether geographical and occupational proximity could be the cause of genetic affinity between the populations with different ethnicity and also to explore that the genetic

affinity is least with other predominant populations with different traditional occupation despite the geographic proximity.

MATERIALS AND METHODS

Populations Studied: Blood samples were collected from 99 healthy, normal and unrelated donors belonging to two endogamous populations viz. the Jat and Kurmi from places near Dehradun town in Uttar Pradesh State. The Jat belongs to Indo-Caucasoid while the Kurmi belong to Austroloid ethnic groups. Linguistically also the Jat are Indo-European while Kurmi are Austro-Asiatic. Their main occupation of both the castes is agriculture, they celebrate same festivals and their regional deities are also similar.

Markers Studied: The fifteen microsatellite loci (Powerplex 16™ Kit) studied are Penta E(15q), Penta D(21q), D18S51(18q21.3), D21S11(21q11-21q21), THO1(11p15.5), D3S1358(3p), FGA(4q28), TPOX(2p23-2pter), D8S1179(8q), vWA(12p12-pter), CSF1PO(5q33.3-34), D16S539(16q24-qter), D7S820(7q11.21-22), D13S317(13q22-q31) and D5S818(5q23.3-32). These loci were chosen for two reasons. Firstly, they consist of repetitions of tetranucleotide or pentanucleotide repeat units and are therefore less prone to slippage of polymerase during enzymatic amplification (He et al. 1996) thus enabling unambiguous allele typing. Secondly they are located on different chromosomes except D5S818 and CSF1PO (both on chromosome 5) and D21S11 AND Penta D (both on chromosome 21) so there is no possibility of mitotic recombination as they are present far apart from each other. All 15 studied loci are substantially unlinked, which makes them ideal tools to study genomic variation.

DNA Extraction, Multiplex PCR Amplification and Genotyping: Blood was collected in EDTA coated tubes. DNA was isolated from lymphocytes by phenol/chloroform extraction followed by ethanol precipitation. The DNA precipitate was dissolved in 10mM TE buffer and stored at $-20\,^0$C (Sambrook et al. 1989). The fifteen autosomal STR loci were co amplified with 2ng of template DNA in a reaction mixture containing Gold ST*R 10X Buffer, 10X Primer pair mix and Taq DNA polymerase in buffer B (Promega Corporation, Madison,U.S.A.), while ABI Prism™ 377 DNA sequencer was used for genotyping. Resultant data analysis and allele designation were carried out using the Genescan™ and Genotyper™ softwares.

Statistical Analysis: Gene counting method has been used to estimate the allele frequency with DNATYPE software (Chakrobarty et al.). Hardy-Weinberg equilibrium (HWE) was assayed by the exact test (Guo and Thompson 1992). DISPAN software (Ota 1993) was used to compute G_{ST} and D_A distances between populations. D_A was then employed for the construction of the neighbour –joining (NJ) phylogenetic tree

RESUTLS

The data on allelic variation at 15 microsatellite loci among the studied populations exhibit extensive polymorphism at all the loci and these were published elsewhere (Tandon et al. 2002) The heterozygosity per locus and the overall average heterozygosity values in the Jat and Kurmi populations is shown in Table 1. The values range from 0.520 at TPOX to 0.966 at FGA in Kurmi and 0.691 at TPOX and 0.913 at Penta D and Penta E in Jat. The patterns of allelic distribution at 15 loci among the two populations are shown in Fig.1. The four loci namely Penta E, FGA, D18S51 and D21S11 show a vast level of allelic distribution. The range of allele at the studied loci varies from 5 at Penta E to 35 at D21S11 in Jat while it is 36.2 at D21S11 in Kurmi. For both the communities, the most frequent alleles are 14 29, 17, at D21S11, vWA, D8S1179 and D16S539 loci respectively (Table 2). The estimates of G_{ST} value, a measure of the extent of genetic differentiation at all the loci showed considerable heterogeneity. It is as high as 2.5%

Table 1: Locuswise average heterozygosity and G_{ST} values at 15 microsatellite loci in Jat and Kurmi Population of North India.

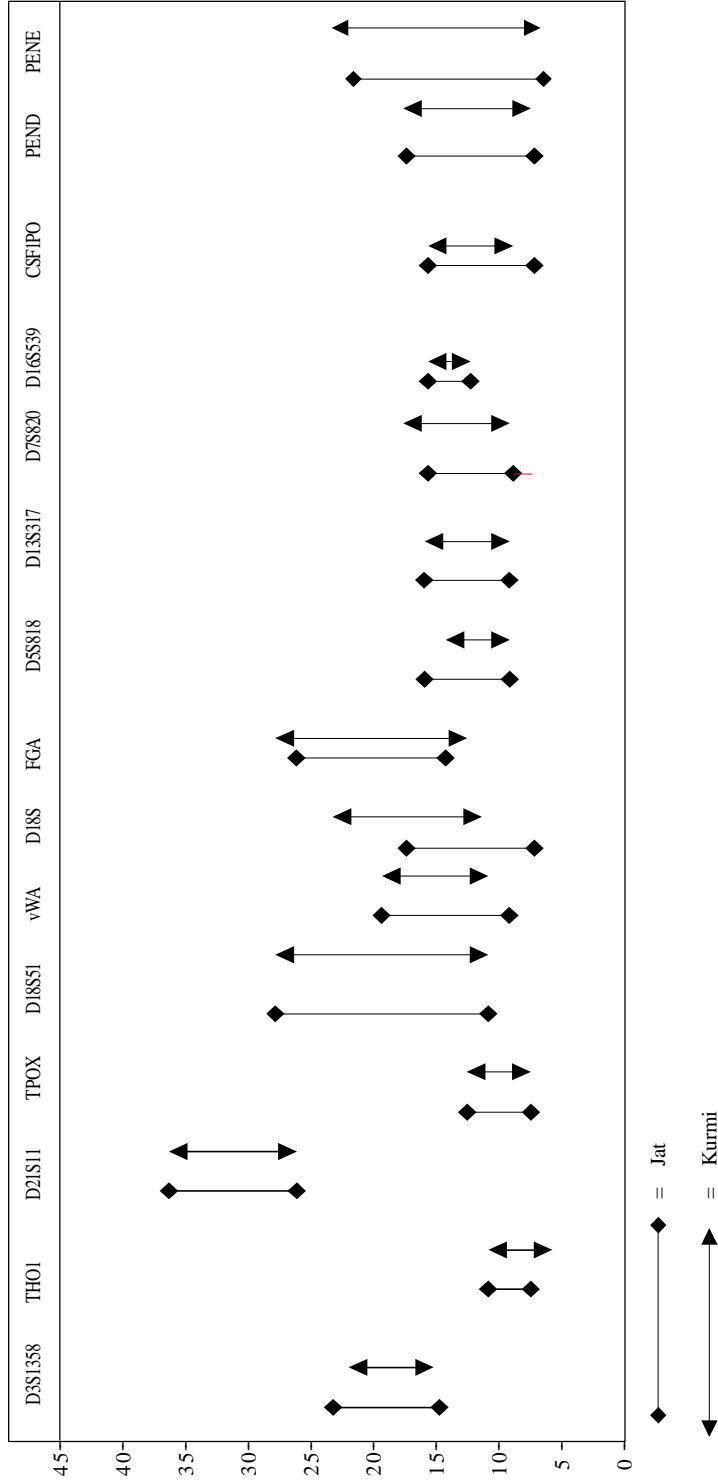| Locus	Jat	Kurmi	G_{ST}
D3S1358	0.85	0.799	0.013
THO1	0.76	0.681	0.011
D21S11	0.76	0.903	0.007
TPOX	0.691	0.52	0.011
D18S51	0.74	0.764	0.005
PENTA E	0.913	0.941	0.014
VWA	0.83	0.882	0.017
D8S1179	0.733	0.78	0.016
FGA	0.833	0.966	0.01
D5S818	0.73	0.884	0.013
D13S317	0.75	0.711	0.025
D7S820	0.714	0.723	0.005
D16S539	0.81	0.823	0.002
CSF1PO	0.79	0.8	0.007
PENTA D	0.913	0.76	0.016
Average	0.788	0.796	0.011

Fig. 1. Comparison of Allele Diversity at 15 microsatellite loci in Jat and Kurmi populations of UP.

Table 2: Most frequent allele for 15 STR loci in Jat and Kurmi populations on North India

Population	Loci/Allele	D3S1358	THO1	D21S11	TPOX	D18S51	vWA	PENTA E	D8S1179	FGA	D5S818	D13S317	D7S820	D16S539	CSF1PO	PENTA D
Jat	Most Frequent	16	9	29	8	15	17	11	14	23	11	12	11	11	12	11
Kurmi	Most Frequent	15	8	29	11	14	17	16	14	21	12	11	8	11	11	10,13

in case of D13S317 and lowest (0.2%) at D16S539 locus. The average G_{ST} was found to be very low (1.1%). To assess the genetic distance between the two caste groups D_A distance was computed (Nei et al. 1983) and compared with other two predominant castes -Thakur and Khatri, cohabiting the region. The genetic distance of the other comparable populations of the same geographical area is larger than between the Jat and the Kurmi. The genetic affinity between the four populations was investigated using neighbour-joining (NJ) method (Saitou and Nei 1987) and the pattern obtained (Fig. 2.) which clearly revealed two distinct clusters, one consisting of the Kurmi and Jat with a low genetic distance of 0.071 and the other of the Khatri and Thakur populations.

DISCUSSION

The anthropological and archaeological records reveal that the fertile Gangetic region of Uttar Pradesh had nurtured many civilizations after settlement of modern humans in this area. Genetic studies are however scanty to substantiate succession of civilization in this region. Even the few studies undertaken to reveal genetic diversity are based on classical genetic markers and anthropological data (Tiwari and Bhasin 1968). This study perhaps, is the first attempt of genomic variation among two predominant occupational caste groups of the region. The Jat and Kurmi are two widely distributed populations in several regions of northern and western India. Both the communities have different historical backgrounds and have migrated from other territories and settled

in western UP at different periods in the past. As per documented history Jat population is the offshoot of Scythian population of Central Asia and known to have migrated to India between 1 BC to 5 AD (Russell and Hiralal 1916). The origin and history of Kurmi is not clear. In physical appearance Kurmi bears resemblance to Austroloid ethnic elements: short height, with dark brown skin colour, stubbed nose, thick lips and prominent eyebrow-ridges. The ethnic notes suggest the possibility of two groups of Kurmi: one from Chota Nagpur and Orissa belt and others in the northern region. It is possible that Kurmi of Uttar Pradesh were basically Austroloid ethnic population migrated from central India and during generations of their stay with other neighbouring populations and possible admixture with the castes around might have acquired some of the Caucasoid morphological features or traits. Both are endogamous occupational caste groups with different historical backgrounds but live and share the same geographical area (since their settlement) along with other similar widely distributed caste groups e.g., Thakur, Khatri etc. Given the situation, apart from other influencing factors, the genetic affinity between these two are expected to be guided by the relative roles of geographic proximity, which might allow admixture and gene flow, and in contrast the caste endogamy which will maintain the genetic homogeneity. The extent of population differen-tiation (G_{ST}= 1.1%) is rather low, clearly indicates that within a geographical unit the population group show greater genetic similarity. Previous studies based on classical genetic polymorphism (Papiha et al. 1996; Bhasin and Khanna 1992) to
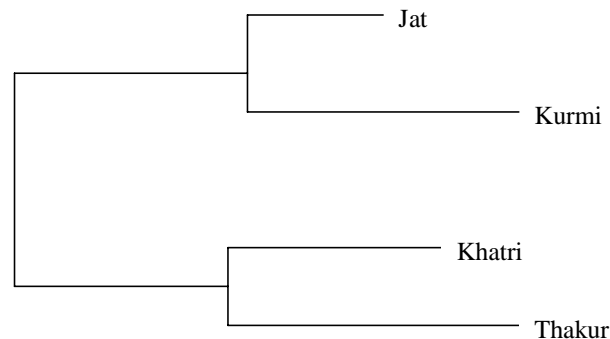

Fig. 2. Neighbour-joining tree depicting genetic variation among Jat and Kurmi with Thakur and Khatri populations of North India

investigate genetic affinity among these groups revealed low level of genetic differentiation among them but greater significant differences with the other neighbouring populations. Though the genomic studies of these region is not explored till date, the present study based on 15 microsatellite loci suggest that the genetic affinity between Jat and Kurmi is greater influenced by geographic proximity, despite the two groups are ethnically different and follow caste endogamy but their same occupation might have been the major factor responsible for gene flow among the two populations. Polymorphism at fifteen microsatellite loci reveals that the Jat, in general, shows greater diversity than Kurmi at all the studied loci, except at FGA locus. This possibly suggests differential rate of endogamy and variable population size of the two caste groups. Kurmi caste group is numerically smaller and highly endogamous over a narrow geographical area, which is expected to lead lower range of allelic variation. It is interesting to find that both Jat and Kurmi show the least number of alleles at D16S539 locus and highest number of alleles at D18S51 and Penta E loci. This could be due to greater mutation rates at these loci than other studied genes or could be due to presence of some specific influencing factors of selection or past demographic events affecting some susceptible loci. The result shows close affinity between Jat and Kurmi. Their genetic structure is primarily guided by endogamy however certain extent of gene flow cannot be ruled out. This professional proximity could have been the strong influencing factor for genetic affinity between them. Theoretically, gene flow can have stronger and greater influence than other competing co-evolutionary factors and can have a greater and stronger influence on genetic structure of the population. In contrast the endogamy and inbreeding are slow process and its influence on the genetic differentiation can be nullified by other factors e.g., bottleneck events, populations' expansion and gene flow. Though the sample sizes is limited and apart form other influencing factors what can be inferred from the results is that greater genetic affinity between Jat and Kurmi is attributed to gene flow. This can possibly be verified by future model based studies on gene flow among the regional populations. The greater genetic diversity with respect to other neighbouring populations Thakur and Khatri of the same geographical area

is possibly due to least gene flow and the genetic differentiation is more influenced by their caste endogamy. One of the interesting aspect of these widely distributed caste groups is that they form localized groups forming separate gene pools. To investigate the genetic differentiation of these widely distributed castes groups in different regional breeding populations and the microevolutionary process, study especially based on genomic data will be of much interest. Hence it can be concluded that geographical proximity and occupational affiliations seem to be strong determinants of genetic affinities among populations. In near future we expect that such well genomic designed studies among regional populations will give further insights, of the influence of geographical and occupational proximity on population structure and the genomic differentiation.

ACKNOWLEDGEMENTS

REFERENCES

Ahmed I, Saxena NC 2001. Caste, land and political power in UP. In: KL Sharma (Ed.): *Caste and Class in India*. Jaipur and New Delhi: Rawat Publishers, pp.171-201.
Bhasin MK, Walter H, Danker-Hopfe H 1992. *The Distribution of Genetical, Morphological and Behavioural Traits Among the Peoples of Indian Region (Bangladesh Bhutan, India, Maldives, Nepal, Pakistan, SriLanka).* Delhi: Kamla-Raj Enterprises.
Bhasin MK, Walter Danker-Hopfe H 1994. *An Investigation of Biological Variability in Ecological, Ethno-Economic and Linguistic groups.* Delhi: Kamla-Raj Enterprises.
Bhasin MK, Walter H 2001. *Genetics of Castes and Tribes of India.* Delhi: Kamla-Raj Enterprises.
Bowcock AM, Linares RA, Tomferhrde J, Minch E et al. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature,* **368:** 455-457.
Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK 1998. Short tandem repeat polymorphism evolution in humans. *Eur J Hum Genet,* **6:** 38-49.
Cavalli-Sfroza LL, Menozzi P, Piazza A 1994. *The History and Geography of Human Genes.* New Jercy: Princeton University Press. Pp. 238-242.
Chakraborty R, Stivers D, Zhong Y YEAR. *DNATYPE Software/(Windows 95/NT Ver).* Human Genetic

Centre, Houston Texas: University of Texas Health Science Centre.

Deka R, Shriver MD, Yu LM, Herdriech EM, Jin L, Zhong Y, McGarvy ST, Agarwal SS, Bunker CH, Miki T, Hundrieser J, Yin S, Raskin S, Barrantes R, Ferrell RE, Chakrobarty R 1999. Genetic variation at twenty three microsatellite loci in sixteen human populations. *J Genet,* **78:** 99-121.

Destro-Bisol G, Boshi I, Caglia A, Tofanelli S, Pascali V, Paoli G, Spedini G 2000. Microsatellite variation in central Africa: An analysis of Intrapopulation and Interpopulation genetic diversity. *Am J Phys Anthropol,* **112:** 319-337.

Dutta R, Kashyap VK 2001. Genetic Variation observed at 3 tetrameric STR loci HumTHO1, TPOX and CSF1PO in 5 ethnic population groups of northeastern India. *Hum Biol,* **74:** 33-49

Guo SW, Thompson EA 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics,* **48:** 361-372.

He L, Morris S, Lennon A, St Clair DM, Porteous DJ, Wright AF, Muir WJ, Blackwood DH 1996. A genome-wide search for linkage in a large bipolar family: comparison of genotyping accuracy using di-and tetranucleotide repeat microsatellite markers. *Psychiatr Genet,* **6:** 123-129.

Jorde LB, Rogers AR, Bamshed M et al. 1997. Microsatellite diversity and demographic history of modern humans. *Proc Natl Acad Sci, USA,* **94:** 109-124.

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB 1998. Signatures of population expansion in microsatellite repeat data. *Genetics,* **148:** 1921-1930.

Lanchbury JS, Agarwal SS, Papiha SS 1996. Genetic differentiation and population structure of some ocupational caste groups in Uttar Pradesh, India. *Hum Biol,* **68:** 679-705.

Mahalanobis PC, Majumdar DN, Rao CR.1948-49. Anthropometric survey of the United Provinces 1941: A Statistical study. *Sankhya,* **9:** 80-324

Malhotra KC 1979. Excommunication as a process leading to the formation of new groups. *East Anthropologist*, **32:** 49-53.

Malhotra KC, Vasulu TS 1993. Structure of Human population in India. In: *Human Population Genetics.* PP Majumdar (Ed.). New York: Plenum, pp. 207-233.

Mastana SS 1999. Genetic analysis of the D1S80 locus in 5 North Indian population. *Ann Hum Biol,* **26:** 405-411.

Mukherjee N, Majumdar PP, Roy B, Roy M, Chakraborty M, Banerjee S 1999. Variation in 4 short tandem repeat loci in 8 population groups of India. *Hum Biol*, **71:** 439-446.

Ota T 1993. DISPAN: *Genetic Distance and Phylogenetic Analysis*. University Park, PA: Tatsuya Ota and Pennsylvania State University.

Papiha SS 1996. Genetic variation in India. *Hum Biol*, **68(5):** 607-628.

Papiha SS, Roberts DF, Wig NN, Singh S 1972. Red cell enzyme polymorphism in Punjabis in North India. *Am J Phys Anthropol*, **37:** 293-299.

Perez Lezaun A, Calafell F, Mateu E, Comas D, Ruiz Pacheco R, Bertranpetit J 1997. Microsatellite variation and the differentiation of modern humans. *Hum Genet*, **99:** 1-7.

Reddy MB, Sun G, Luis JR, Crawford MH, Heman NS, Deka R 2001. Gnomic diversity at 13 STR loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Hum Biol*, **73:** 175-190.

Rupsell RV, Hiralal 1916 (1975). *The Tribes and Castes of the Central Provinces of India*. London: Macmillan and Company, 1916 rpt.1975, Delhi: Cosmo Publications. Vol II

Saitou N, Nei M 1987. The neighbour joining method: A new method for reconstructing phylogenetics trees. *Mol Biol Evol*, **4:** 406-425.

Sambrook J, Fritsch EF, Maniatis T 1989. *Molecular cloning. A laboratory manual*. 2nd Ed. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Tandon M et al. 2002 Genomic diversity at 15 fluorescent-labeled short tandem repeat loci in few important populations of State of Uttar Pradesh, India. *Forens Sci Int,* **128(3):** 190-195.

Tiwari SC, Bhasin MK 1968. The blood groups of the Brahmin and Rajputs of Garhwal. *Hum Biol*, **40:** 386-395.