Genomic Diversity in North-East, India: Genetic Relationship Among Five Ethnic Populations From Manipur Based on VNTR Polymorphism

Dutta Ranjan[1], V.K. Kashyap[1], R. Trivedi[1] and T.S. Vasulu[2]

1. DNA Typing Unit, Central Forensic Science Laboratory, Kolkata 700 014, West Bengal, India
2. Indian Statistical Institute, Kolkata 700 108, West Bengal, India

KEY WORDS VNTR loci; average heterozygosity; genetic distance

ABSTRACT Polymorphisms at four mini-satellite loci (D1S7, D4S139, D5S110 and D17S79) were examined to assess the genetic relationship among five contiguous and socio-culturally distinct ethnic populations (Meitei, Kuki, Naga, Hmar and Manipuri Muslim) in northeastern India. All the studied loci were highly polymorphic, with almost no departure from HW-equilibrium. Heterozygosity values among the five groups for the minisatellite range from 68% to 94%. The gene differentiation for the VNTR loci was moderate (Gst = 0.030). Neighboring joining method of clustering suggests a close cluster for the tribal groups of Kuki, Naga and Hmar, while Manipur Muslim is distinct from other populations. The results show that geographic proximity has the least influence on genetic variation and support the results obtained from the traditional markers that northeast Mongoloid populations show genetic affinity irrespective of their linguistic background and geographic proximity.

INTRODUCTION

India is inhabited by diverse populations of tribes, castes, religious and migrant groups which could be referred as fascinating natural population genetics laboratory and provides interesting insights for studying microevolutionary process. The Northeast region is particularly interesting as represented by several tribal populations of Mongoloid origin, who have migrated from eastern and southeast Asian regions and settled in different parts during several waves of migration (Rapson 1955; Dani 1960). Though they share similar physical features but speak different languages and show differences in cultural, anthropological and genetic traits (characters) (Roychoudhury 1992; Bhasin and Walter 2001). The affinity and/or diversity and the microevolutionary nature of these groups have evoked considerable interest among the anthropologists and population geneticists. Several workers have studied the genetic diversity among some localized groups using classical genetic polymorphisms to understand the genetic structure and microevolution in northeast part of India (Das and Deka 1985; Walter 1986; Das et al. 1987; Chakraborty et al. 1986; Deka et al. 1988; Roychoudhury 1992; Bhasin et al. 1994). For example, investigating the genetic relationship based on available classical gene frequency data between four sets of populations of different castes and ethnicity Roychoudhury (1992) observed that "all the Mongoloid affiliated populations in eastern India show genetic similarity with respect to geographical proximity, no matter whether they originated from the same tribal group or same linguistic family in the past". In recent years the identification of hyper variable loci in human genome e.g., VNTR and STR, were found to be useful in studies concerning information on population structure and microevolution. Such studies were initiated at the global and racial level and studies at the regional or local levels, especially at the Indian context, is very few, sporadic and limited to some local populations (Deka et al. 1995; Papiha 1996; Papiha et al. 1996; Mukherjee et al. 1999; Mukherjee et al. 2000; Reddy et al. 2001; Dutta and Kashyap 2001a, b; Dutta et al. 2001).

Recently the Central Forensic Laboratory, Calcutta, has undertaken a National Human Genome Project to characterize and develop of Genomic Data Bank of a variety of castes, tribes and ethnic groups from each region of the country. A few populations have been studied in this regard. In this study we have analyzed genetic variation in VNTR loci among five groups in Manipur region. Manipur located at northeastern part of India, close to Myanmar, represent a unique population structure. It represents several tribal populations settled since centuries and some have remained as tribal populations and some have adopted Hindu caste social structure whereas some others have migrated from other parts of India. This situation of fusion-fission

process expected to result in wide genetic diversity between populations. Especially that some populations are expected to show greater affinity to other tribal populations of the region whereas others are closer to caste populations. The genetic structure and diversity of these populations have not been fully explored. The present study examines the use of VNTR loci in assessing the genetic affinity among the regional populations and based on genomic diversity of VNTR loci we test the hypothesis that Mongoloid populations show genetic similarity irrespective of their geographic contiguity, language affiliation or origin.

MATERIALS AND METHODS

The Sample: The five populations considered for the study is: Meitei, a caste group, three tribal populations (Naga, Kuki, Hmar) and a religious community (Manipuri Muslim). Meities are said to be the original inhabitants of Manipur and do not practice consanguineous marriages. Naga, Kuki and Hmar tribes are said to belong to the Naga-Kuki-Chin group believed to have migrated to Manipur 300-400 years ago (Saha and Tay 1990) and maintain high endogamy and practice consanguineous marriages. While Naga, Kuki and Hmar speak the local Manipur language, the Meities speak Hindi and Bengali more commonly.

The present study employs minisatellite loci: D1S7, D4S139, D5S110 and D17S79. These are well documented for analysis of genetic variation (Chakraborty 1990; Jin and Chakraborty 1991; Papiha et al. 1996).

Blood Samples: Blood samples were obtained by venipuncture in EDTA coated vaccutainers from unrelated donors. Collection of blood samples was carried out from several regions of the state of Manipur. The samples (randomly selected males and female individuals) from Meitei and Manipuri Muslims were obtained from Imphal city, while Kuki, Naga and Hmar tribal samples were collected from the hills bordering the states of Assam and Arunachal Pradesh. The Naga samples were mainly collected from Tankul and Kabui the two major subgroups of the population.

Samples from the respective communities were stored in ice containers during transportation from the field of collection to CFSL, Calcutta. The collected samples Kuki (75), Naga (78), Meitei (128), Hmar (60) and Manipuri Muslim (65) was screened for testing. DNA was extracted by the organic methods of extraction (Maniatis et al. 1982). Quality and quantity assessment of the isolated DNA was performed using 0.8% agarose yield gel and slot blot procedures (Waye et al. 1989).

Digestion, Electrophoresis and Hybridization: High molecular weight DNA (‘200ng) was digested with Hae III at 37° C overnight. The digested samples were run on 1-% agarose gels in IX TAE buffer. Following electrophoresis the fragments were transferred onto nylon membranes and hybridized with chemiluminescent probes MS1, pH30, LH1 and VI specific for loci D1S7, D4S139, D5S110 and D17S79 respectively. The membranes were finally treated with lumiphos reagent in dark before exposing them to films for development of profiles. Developed profiles were compared with respect to the 22kb ladder and their sizes were ascertained with computer assisted image analysis software (Biorad). Approximate molecular weight (base pair) sizing estimates for each samples was obtained. Following development of profiles, membranes were stripped and rehybridized with subsequent probes.

Statistical Analysis: The size estimates was done with respect to 0.5 – 22 kb ladder (Gibco BRL) using computer assisted software supplied with phosphor Imager GS 505 (Biorad). Measured fragments were grouped with fixed-bin classification of the fragment sizes, following the fixed-bin boundaries listed by Budwole et al. (1991). The unbiased estimate of the VNTR profiles and number of possible genotypes were used for determination of the expected heterozygosity (Edward et al. 1992). Possible divergence from Hardy-Weinberg expectation (HWE) was determined by calculating the unbiased estimates of the expected homozygotes and heterozygote frequencies (Chakraborty et al. 1988; Nei and Roychoudhury 1974; Nei 1978); the likelihood ratio test (Chakraborty et al. 1994; Weir 1992) and exact test (Guo and Thomson 1992). The null allele frequency ‘r’ was computed using the methods described by Chakraborty et al. (1994). The level of significance of the test, where applicable, were determined by shuffling the

observed fragment sizes across the individual and replicating the allele shuffling 2000 times for each test (Chakraborty et al. 1994). An interclass correlation criterion (Karlin et al. 1981) for two-locus association was used for detection of the disequilibrium between the set of VNTR loci. In addition, to determined the linkage disequilibrium between the set of VNTR loci, variance (S_k^2) of the number of heterozygotes loci in the population sample was also performed (Brown et al. 1980; Chakraborty et al. 1994). Gene diversity analysis (Nei 1973) was performed and genetic distances (D_A) was computed according to methods described by Nei et al. (1983). Neighbor-joining trees were constructed for depicting the genetic affinities between populations (Saitou and Nei 1987), with the reliability of the phylogenetic relationships examined by the method of bootstrapping with 500 replications.

RESULTS

Distribution of Fixed Bin Frequencies at the different loci among the five populations is shown in figure 1. A visual inspection of this figure show that the frequent fragments were found either in the same or adjacent bins. Occurrences of bins with the highest frequency as above have also been noticed for other Indian and world populations (Sovinsky et al. 1996; Dutta and Kashyap 2001a; Laboratory Division FBI 1993). HWE-Homozogosity Test, Likelihood Ratio Test and Exact Test were computed to observe possible departure from HWE (Table 1). The table showed an apparent deviation from the HWE expectation at the D5S110 locus for the Kuki and Hmar population. These loci showing the devia-tions were reanalyzed for the presence of non-detectable allele 'r' (Chakraborty et al. 1994). The reanalyzed data was in conformity with HWE expectations. The value of 'r' was found between 0-4% among the 20 population-VNTR loci combinations. These values were consistent with frequency of non-detectable alleles observed among other populations using VNTR markers (Sovinsky et al. 1996; Dutta and Kashyap 2001a; Laboratory Division FBI 1993; Chakraborty et al. 1994; Karutha Pandian et al. 1995a,b). Lack of significant HWE results, (after taking undetected alleles into account) is an interesting result for the three tribal populations where there is a practice of consanguineous marriage and high endogamy. Linkage disequilibrium in multilocus genotype data from population surveys may be detected by the observed variance of the number of heterozygous loci (S_k^2) statistics (Brown et al. 1980).

Table 1: Hardy-Weinberg equilibrium test for binned genotype frequencies at the four VNTR loci in northeastern population groups, India

Locus	*Metei*	*Kuki*	*Naga*	*Hmar*	*M. Muslim*
N	*128*	*75*	*78*	*60*	*68*
D1S7					
HWE-Test	0.053	0.158	0.387	0.463	0.944
Exact Test	0.087	0.184	0.246	0.085	0.111
Likelihood Ratio Test	0.066	0.123	0.301	0.150	0.051
'r'	0.017	0.021	0.0	0.022	0.0
D4S139					
HWE-Test	0.186	0.101	0.339	0.350	0.785
Exact Test	0.077	0.185	0.467	0.552	0.540
Likelihood Ratio Test	0.149	0.361	0.310	0.652	0.373
'r'	0.026	0.041	0.0	0.008	0.0
D5S110					
HWE-Test	0.371	0.656*	0.298	0.437*	0.252
Exact Test	0.074	0.609*	0.081	0.367*	0.150
Likelihood Ratio Test	0.117	0.496*	0.073	0.512*	0.081
'r'	0.026	0.061	0.0	0.045	0.025
D17S79					
HWE-Test	0.187	0.084	0.300	0.748	0.234
Exact Test	0.115	0.520	0.267	0.546	0.619
Likelihood Ratio Test	0.210	0.093	0.382	0.708	0.477
'r'	0.032	0.042	0.033	0.002	0.0

'r' : Null Allele Frequency *Value with 'r'

Locus wise average heterozgygosity and G_{st} values for each of the four VNTR loci among Meitei, Naga, Kuki, Hmar and Manipuri Muslim populations are shown in table 2. The average heterozygosity values range from 0.685 for D4S139 in M.Muslim to 0.940 for D5S110 in Hmar.

Table 2: Average heterozygosity and Gst values for VNTR loci in the North-east populations of India.

Locus	*Meitei*	*Kuki*	*Naga*	*Hmar*	*M.Muslim*	*Gst*
D1S7	0.882	0.917	0.918	0.913	0.907	0.021
D4S139	0.910	0.890	0.897	0.906	0.685	0.030
D5S110	0.916	0.927	0.896	0.940	0.915	0.030
D17S79	0.796	0.815	0.833	0.828	0.795	0.038
Average	0.876	0.887	0.886	0.897	0.825	0.030

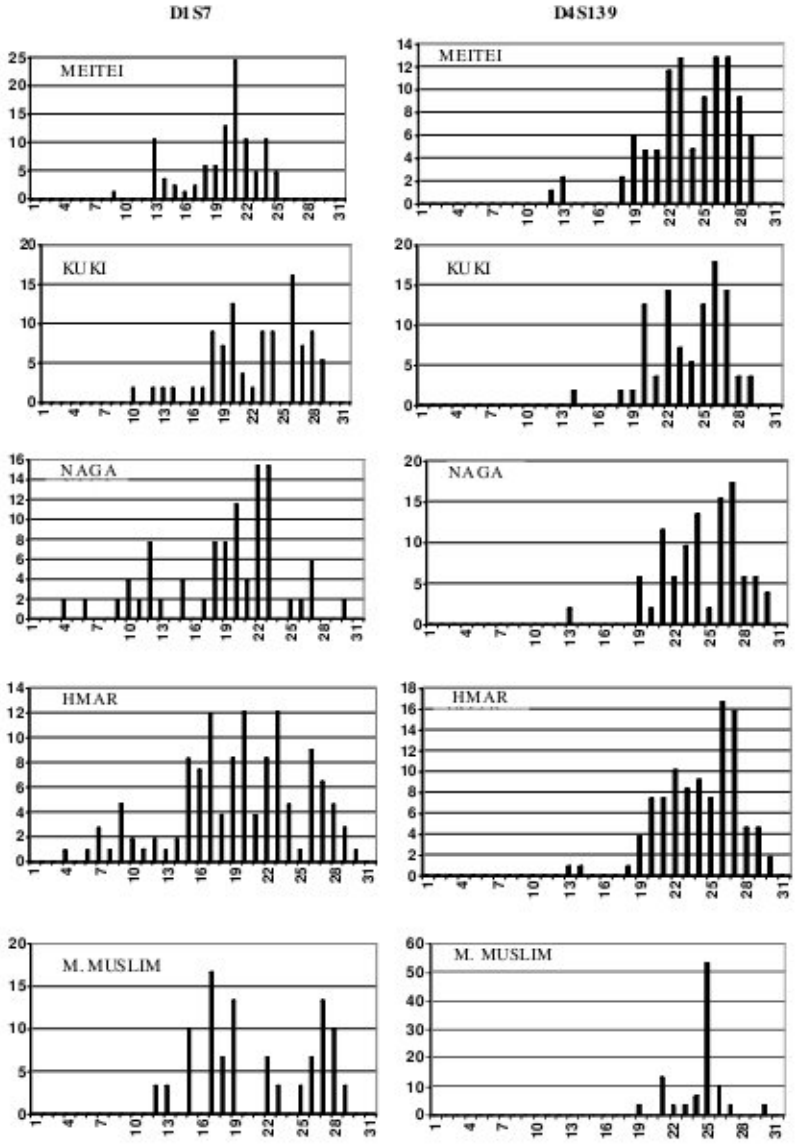

Fig. 1A. Fixed Bin Allele frequency data at the four VNTR loci

The over all average heterozygosity for the four VNTR loci range between 82.5% (M.Muslim) to 89.7% (Hmar). The average G_{st} values for the four minisatellite markers were found to be 0.030. The matrix of genetic distance D_A among the studied population groups is shown in table 3. Although the Meitei, Naga, Kuki and Hmar populations did not show any significant distance between them as compared to the distance exhibited by these groups with the Manipuri Muslims. Based on the distance matrix, neighbor-joining trees were constructed depicting the affinities between the populations (Fig. 2) at the VNTR loci. In the NJ plotting VNTR markers, population groups of Brahmin, Kayastha and Garo from West Bengal, India (Dutta and Kashyap 2001a) and north-western populations (Slovinsky et al. 1996) were included. The dendrogram shows three distinct

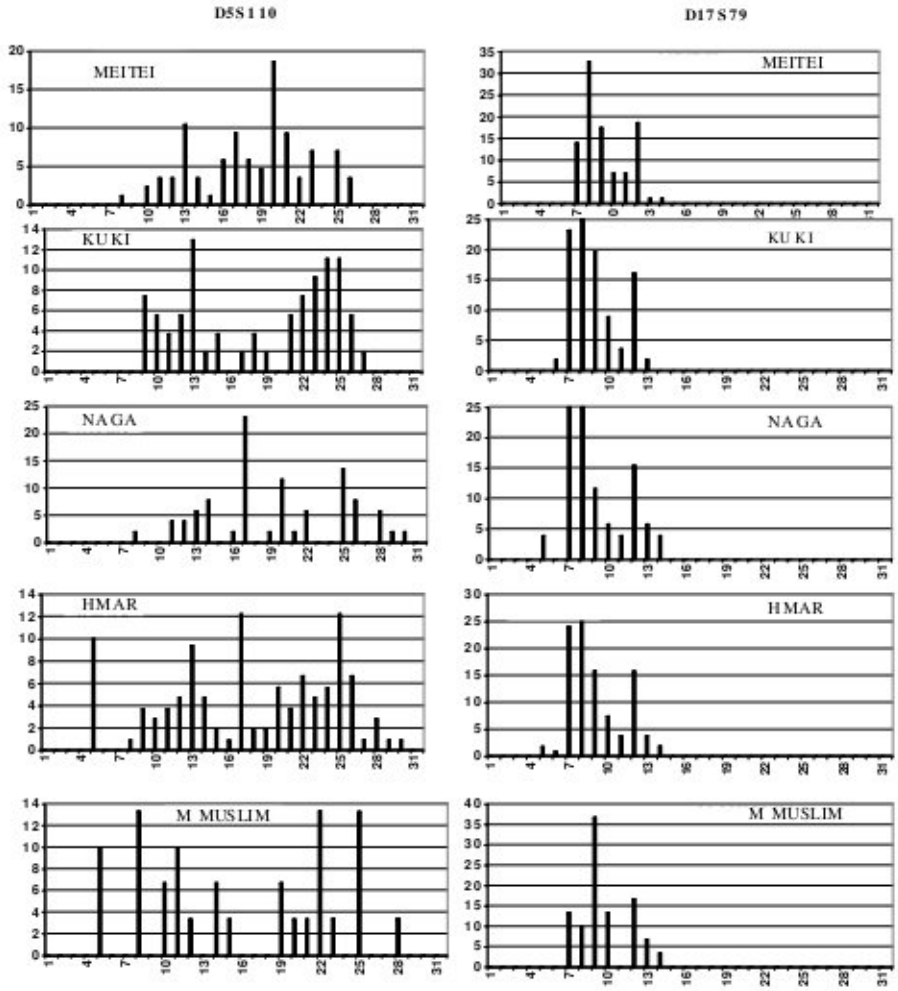

Fig. 1B. Fixed Bin allele frequency data at the four VNTR loci

clusters. The three Manipur populations of the study and Garo (from West Bengal) formed a separate cluster and it is separated from the other Indian populations at 94% of all the trees by the method of bootstrap. The five Indo-Caucasoid populations—Brahmin, Kayastha, Hindu, Punjabi and Sikh—formed a separate cluster with a bootstrap value of 87%. The Manipuri Muslim is distinctly different from other regional populations including the four neighboring Manipur populations.

DISCUSSION

Despite allelic diversity among the populations, frequent allele bins were found in close proximity to each other for the VNTR loci. The occurrence of the frequent alleles was in accordance with other reports (Slovinsky et al. 1996; Dutta and Kashyap 2001a,b). One possible explanation of such a distribution could be the evolutionary antiquity of alleles, where usually the most common alleles are the oldest, while the wider distribution of the low-frequency alleles reflect different rates and types of mutations in populations (Waterson and Guess 1997). It is interesting to find the lack of significant HWE result for the three tribal groups where they practice consanguineous marriage and high

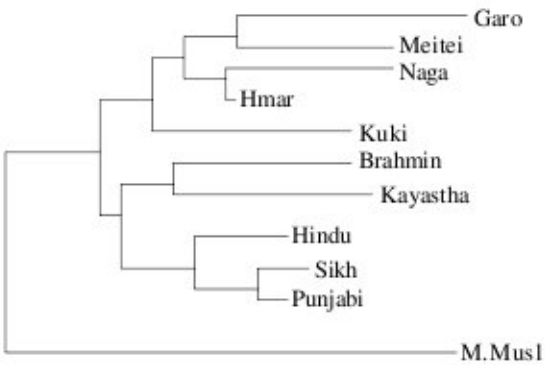

Fig 2. Neighbor-joining network depicting genetic affinities between population groups of India at the four VNTR loci

levels of endogamy. This could be due to random selection of the samples and the large size of the populations.

The distance matrices show large distance values for Manipuri Muslim group with others (Table 3). The clustering pattern shows separation of populations, which corresponds with their socio-cultural and ethnic differences and not

Table 3: Genetic distance (DA) among the studied populations

Population	*Meitei*	*Kuki*	*Naga*	*Hmar*
Meitei	–			
Kuki	0.179			
Naga	0.160	0.234	–	
Hmar	0.107	0.062	0.072	–
M.Muslim	0.323	0.270	0.282	0.208

according to their geographic proximity. Manipuri Muslims are a migrant community and are Caucasoid group in contrast to the neighboring Mongoloid groups. The clustering shows that they are distinct from both the other regional Caucasoid populations and also from neighboring Mongoloid populations and standout separately from the rest. This reflects the influence of populations structure variables, especially marriage pattern, migration that need to be investigated for understanding the genetic diversity of the populations. In comparison to caste groups the tribal population and the Muslim community were relatively isolated since several generations after their migration to northeastern hilly terrain (Singh 1998). Thus though Manipuri Muslim populations live along with Meitei and other tribal populations in the same geographical region but show wide genetic diversity. This suggests that the geographic contiguity is less influential factor of genetic affinity between diverse ethnic groups in Manipur. However in another study by Mukherjee et al. (1999) (based on 4 STR loci in eight diverse populations from West Bengal (3), Orissa (3) and Uttara Pradesh (2)) they found greater geographic influence on genetic similarity than sociocultural proximity. This is because of the marriage pattern and possible gene flow between the neighboring castes in other parts of India. Comparison with other Indian populations gives much more clarity of the clustering pattern in this regard. For example, Garo a Mongoloid population from West Bengal gets clustered with the northeast Manipur (Mongoloid) groups. Meitei a middle caste Mongoloid population clusters along with other Mongoloid tribal groups. The other regional caste populations form a separate cluster irrespective of their caste hierarchy. The results of our study is supported by the findings based on classical genetic markers.

ACKNOWLEDGEMENTS

The work was performed with financial assistance to CFSL, Calcutta under plan scheme. The first author thanks the BPR and D for providing a JRF for the research work, L. Fimate, RIMS, Imphal, Manipur for providing assistance in the collection of blood samples.

REFERENCES

Bhasin MK, Walter H 2001. *Genetics of Castes and Tribes of India.* Delhi: Kamla-Raj Enterprises.

Bhasin MK, Walter H, Danker-Hopfe H 1994. *People of India. An Investigation of Biological Variability in Ecological, Ethno-economic and Liguistic Groups.* Delhi: Kamla-Raj Enterprises.

Brown ADH, Feldman MW, Nevo E 1980. Multilocus structure of naturalpopulations of Horeum spontaneum. *Genetics,* **96:** 523-536.

Budwole B, Guisti AM, Waye JS 1991. Fixed bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci for use in forensic comparisons. *Am J Hum Genet,* **48:** 841-855.

Chakraborty R 1990. Genetic profile of cosmopolitan populations. Effect of hidden subdivisions. *Anthropol Anz,* **48:** 313-331.

Chakraborty R, Walter H, Mukherjee BN, Malhotra KC, Sauber P, Banerjee S, Roy M 1986. Gene differentiation among ten endogamous groups of West Bengal, India. *Am J Phys Anthrop,* **71:** 295-309.

Chakraborty R, Smouse PE, Neel V 1988. Population amalgamation and genetic variation: Observations on artificially agglomerated tribal populations of Central and South America. *Am J Hum Genet,* **43:** 709-725.

Chakraborty R, Zhong Y, Jin L, Budwole B 1994. Nondetectability of restriction fragments and independence of DNA fragments size within and between loci in RFLP typing of DNA. *Am J Hum Genet,* **55:** 391-401.

Dani AH 1960. *Prehistory and Protohistory of Eastern India.* Calcutta: Firma KL Mukhopadhyay.

Das BM, Deka R 1985. Population studies in Assam: ABO blood groups, Hemoglobin D and G-6-PD deficiency. *Anthrop Anz,* **43:** 81-86.

Das BM, Walter H, Gilbert K, Lindenberg P, Malhotra KC, Mukherjee BN, Deka R, Chakraborty R 1987. Genetic variation of five blood group polymorphisms in ten population groups of Assam, India. *Int J Anthrop,* **2:** 325-340.

Deka R, Reddy AP, Mukherjee BN, Walter H, Malhotra KC 1988. Hemoglobin E distribution in ten endogamous population groups of Assam, India. *Hum Hered,* **38:** 261-266.

Deka R, Shriver MD, Yu LM, Ferrel RE, Chakraborty R 1995. Intra and inter-population diversity at short tandem repeat loci in diverse population of the world. *Electrophoresis,* **16:** 1659-1664.

Dutta R, Kashyap VK 2001a. Genetic variation observed in four VNTR loci-D1S7, D4S139, D5S110 and D17S79 in three population groups of eastern India. *J Genet,* **80:** 23-30.

Dutta R, Kashyap VK, 2001b. Genetic variation observed at three tetrameric short tandem repeat loci-HUMTHO1, TPX and CSF1PO-in five ethnic population groups of Northeastern India. *Am J Hum Biol,* **13:** 23-29.

Edwards A, Hammond H, Jin L, Chaskey CT, Chakraborty R 1992. Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. *Genomics,* **12:** 241-253.

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

Jin L, Chakraborty R 1991. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity,* **74:** 274-285.

Karlin S, Caeron EC, Williams PT 1981. Sibling and parent-offspring correlation estimation with variable family size. *Proc Natnl Acad Sc,* **78:** 2664-2668.

Karutha Pandian S, Kumar S, Dharmalingam K, Damodaran C 1995a. DNA profiling of the VNTR locus D2S44 in the population of Madras City and in the tribal Kotas of the Nilgiri Hills, south India. *Gene Geography,* **9:** 65-90.

Karutha Pandian S, Kumar S, Dharmalingam K, Damodaran C, 1995b. Allele frequency distribution for the variable number tandem repeat locus D1OS28 in Tamil Nadu (south India) population. *Electrophoresis,* **16:** 1689-1692.

Maniatis T, Fritsch EP, Sambrook J 1982. *A Laboratory Mannual.* New York: Cold Spring Harbour Laboratory Publications. Pp 458-459.

Mukherjee N, Majumder PP, Roy M, Dey B, Chakraborty M, Banerjee S 1999. Genomic variation at four short tandem repeat loci in eight population groups of India. *Hum Biol,* **71:** 439-446.

Mukherjee N, Mitra M, Chakraborty M, Majumder PP, 2000. Congruence of genomic and ethnolinguistic affinities among five tribal populations of Madhya Pradesh (India). *J Genet,* **79:** 41-46.

Nei M 1973. Analysis of genetic diversity in subdivided populations. *Proc Natnl Acad Sc USA,* **70:** 3321-3323.

Nei M, Roychoudhury AK 1974. Sampling variances in heterozygosity and genetic distance. *Genetics,* **76,** 379-390.

Nei M 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* **76:** 583-590.

Nei M, TaJina F, Tateno Y 1983. Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *J Molec Evol,* **19:** 153-170.

Papiha SS, Mastana SS, Purandare CA, Jayasekahara R, Chakraborty R 1996. Population genetic study of three VNTR loci (D2S44, D7S22 and D12S11) in five ethnically defined populations of the Indian subcontinent. *Hum Biol,* **68:** 819-835.

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

Rapson EJ 1955. People and languages. In: EJ Rapson (Ed.): *Cambridge History of India.* Vol.1. *Ancient India.* Delhi: S Chand and Co. pp.33-57.

Reddy BM, Dutta R, Langstieh BT, Kashyap VK 2001. Diversity at three tetrameric STR loci in a substructured Golla caste population in southern Andhra Pradesh, in comparison to other Indian populations. *Int J Hum Genet,* **1:** 101-108.

Roy choudhury AK 1992. Genetic relationships of the populations in eastern India. *Ann Hum Biol,* **19:** 489-501.

Saha N, Tay JSH 1990. Genetic studies among Nagas and Hmars of eastern India. *Am J Phys Anthrop,* **82:** 101-112.

Saitou N, Nei M. 1987. The neighbor joining method. A new method for reconstructing phylogenetic tress. *Mole Biol Evol,* **4:** 406-425.

Sovinsky SM, Baird LS, Budwole B, Caruso JF, Cheema DPS, Duncan GT, Hamby PP, Masibay AS, Sharma VK, Tahir MA 1996. Development of Deoxyribonucleic acid (DNA) RFLP database for Punjabis in East Punjab, India. *Foren Sci In,* **79:** 187-198.

Singh KS 1998. *People of India. National Series (IV-VI).* Anthropological Survey of India. Delhi: Oxford University Press.

Walter H 1986. Genetic differentiation process among the people of India. *Int J Anthrop,* **1**; 297.

Watterson GA, Guess HA 1997. Is the most frequent allele the oldest? *Theor Popul Biol,* **11:** 141-160.

Waye JS, Presley L, Budwole B, Shutler GG, Fourney RM 1989. A simple method for quantifying human genomic DNA in forensic specimen extracts. *Biotechniques.* **7:** 852-855.

Weir BS 1992. Independence of VNTR alleles defined as fixed bins. *Genetics,* **130:** 873-887.