AMERICAN JOURNAL OF HUMAN BIOLOGY 16:311–327 (2004)

Original Research Article

Genetic Structure and Affinity Among Eight Ethnic Populations of Eastern India: Based on 22 Polymorphic DNA Loci

V.K. KASHYAP,[1]* P. CHATTOPADHYAY,[1] R. DUTTA,[1] AND T.S. VASULU[2]
[1]*DNA Typing Unit, Central Forensic Science Laboratory, Calcutta, India*
[2]*AHGU, Indian Statistical Institute, Calcutta, India*

ABSTRACT The nature and extent of genetic variation at 22 polymorphic DNA loci, belonging to three distinct classes, especially, 12 STR loci (D3S1358, vWA, FGA, D5S818, D13S317, D7S820, D8S1179, D21S11, D18S51, HPRTB, F13B, LPL), four VNTR loci (D1S7, D4S139, D5S110, D17S79), and six coding loci (HLDQA1, LDLR, GYPA, HBGG, D7S8, GC) were investigated among eight population groups of West Bengal and Manipur regions of India. Of these, two groups from West Bengal belong to Caucasoid and six (one in WB and five in Manipur) belong to Mongoloid stock. Both STR and the expressed loci show wide diversity among the eight populations. For example, Manipur Muslims show differences in allele frequency when compared to four other regional populations. Similarly, Garo, one of the Mongoloid populations of West Bengal, differ in allele frequency from their counterparts in the Manipur region. Departure from Hardy-Weinberg expectations was observed at certain loci in a few populations (e.g., D21S1137 in Kayastha and Brahmin, HUM F13B in Meitei). Heterozygosity values were higher for Caucasoid than Mongoloid groups. The overall gene differentiation (G_{ST}) for STR loci is higher (5.3%) than for those at the expressed region (4.6%). The clustering pattern of the eight populations differs with respect to different classes of genetic markers used. The dendrograms based on six coding loci (HLDQA1, LDLR, GYPA, HBGG, D7S8, GC) differs from those based on STR and VNTR markers. Caucasoid and Mongoloid groups form different clusters and Manipur Muslims are distinct from others. The clustering pattern corresponded with the spatial and ethnic affiliations of the populations. Using different classes of DNA loci at the coding and noncoding region will help to better understand the influence of population structure variables on the genetic structure of populations. Am. J. Hum. Biol. 16:311–327, 2004.

The development of molecular genetic technology has renewed interest in investigating genomic diversity, phylogeny, and microevolution among human populations, which could provide more insight and clarity than what was possible with the traditional serological and biochemical genetic markers. DNA polymorphism of the human genome, especially VNTR and STR loci in the noncoding region, reveal large genetic variation with high levels of heterozygosity and mutation rates and are least influenced by selection processes as compared to genes at the coding loci (e.g., HLA DQA1 and 5 PM loci) in the expressed region. As such, VNTR and STR loci are more preferred than codons to study genetic diversity and microevolution. In the Indian context, some recent studies on DNA polymorphism have addressed the issue of genomic diversity (or affinity) with reference to its association with geographical, ethno-linguistic, and population structure in some populations. Most of the studies are based on highly polymorphic VNTR and STR loci (Papiha et al., 1996; Deka et al., 1999; Mukherjee et al., 1999, 2000; Dutta and Kashyap 2001a,b; Reddy et al., 2001) and HLA diversity is the basis of a few other studies (Balakrishnan et al., 1996; Pitchappan, 1986, 1988). In this regard, it is of interest to investigate the use of different sets of DNA polymorphism in reflecting genetic diversity. Do different sets of hyper-variable loci considered either separately or used collectively provide

similar results? How exactly do the findings based on DNA polymorphism differ or contradict those based on classical sero-genetic data?

In this study we report the genetic diversity among eight populations—three from West Bengal in the eastern region and five from Manipur in the northeastern region (Table 1) based on 22 loci belonging to three classes of DNA markers. These represent six PCR-based expressed coding loci (HLADQA1, LDLR, GYPA, HBGG, D7S8, GC), 12 STR loci, and four VNTR loci. These have been widely used in gene mapping, forensic studies, and information on population structure and population genetics of regional and global populations (Clark, 1987; Chakraborty, 1990; Edward et al., 1992; Jin and Chakraborty, 1995; Papiha et al., 1996; Robinson et al., 1996; Deka et al., 1999). This study seeks to evaluate the usefulness of DNA polymorphic loci at the coding region and with reference to STR and VNTR loci at the noncoding region in assessing the genetic affinities and diversity among the regional populations, and whether a combination of different sets of loci provides similar results than the separate use of either coding genes, or STR or VNTR—the tandem repeats of different classes.

Of all the regions, the northeast region in India is inhabited by mostly Mongoloid groups and is represented by several tribal populations. They have migrated from northern and eastern borders at different periods in the past. Although they share similar physical features, they speak different languages and show differences in cultural and biological characters. When compared with other regions, the genetic diversity and affinity of many of these groups have not been fully explored, due to relative isolation and accessibility and other practical problems. However, Roychoudhury (1992), based on serological data, observed that all the Mongoloid-affiliated populations show genetic similarity with respect to geographic proximity, no matter whether they originated from the same tribal group or linguistic family in the past. Manipur is one such region that represents several Mongoloid tribal groups, some of whom have adopted a caste structure, and it also represents other Caucasoid populations from other parts. The West Bengal region forms a link between the northeast and the rest of the country. The population is of a mosaic nature and represents a diverse population from different regions. There had been free movement and settlement of populations from West Bengal to the northeastern region and vice versa. Using genomic data, the study sought to examine the hypotheses that: 1) Mongoloid populations, especially Garo in West Bengal and Mongoloid groups in Manipur, show genetic similarity irrespective of their origin or language affiliation and their geographical contiguity. 2) Whether Caucasoid groups in West Bengal and in Manipur show genetic affinity. 3) Whether the pattern is consistent irrespective of the type of polymorphic loci used or whether a combination of loci of three different classes of DNA loci better explains the clarity of the genetic affinity of the diverse populations used in the study.

MATERIALS AND METHODS

Population

The study consists of eight populations, three from West Bengal and five from Manipur, northeast region, India. A total of 841 blood samples from unrelated and healthy individuals were collected for the study. Of these, 323 samples were from West Bengal, especially from two higher caste groups (Kayastha and Brahmin from the southern area) and the third from Garo, a Mongoloid-affiliated population from the northern region. Also, 518 samples were from different parts of the Manipur Valley in the northeast region, especially from Meitei, a caste and three tribal populations: Naga, Kuki, Hmar (all four are of Mongoloid ethnic origin), and from Manipuri Muslim a Caucasoid group. All 841 samples were used for analysis of nine STR Profiler Plus loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S17, D7S820). However, due to constraints on the availability of laboratory facilities, supporting chemicals, gene typing kits, probes, etc., only about 623 samples could be used for the analysis of three Monoplex STR loci, six coding loci (HLDQA1 and five coamplified loci) and for the four VNTR loci. The details of the sample size for the different set of loci, geographical location (Fig. 1), social status, and linguistic affiliation of the studied populations (Singh, 1998) are shown in Table 1. However, the allelic frequency data, genetic distance, and genetic structure of the four VNTR loci (D1S7, D4S139, D5S110, D17S79) and six polymorphic coding loci

TABLE 1. Studied population, sample size, location, and ethnic background (Singh, 1998)

Population	Sample size		Sample locations	Traditional occupation; sociocultural affiliation	Ethnic affiliation	Linguistic affiliation
	Monoplex STR Loci**	P.P.Loci, HLDQA1 & Coamp Loci*				
Kayastha	114	103	From the region surrounding Kolkata, districts of 24 Pgs (N) and 24 Pgs (S), Burdwan,	Agriculture, Clerical & Accounts jobs, Business. Upper Caste	Caucasoid	Indo-European
Brahmin	51	110	From the region surrounding Kolkata, districts of 24 Pgs (N) and 24 Pgs (S), Burdwan, Darjeeling	Priests, Business, Teaching, Astrology Upper Caste	Caucasoid	Indo-European
Garo	80	110	Darjeeling and Jalpaiguri, Coochbehar regions of West Bengal	Shifting Cultivators, Labourer. Tribe	Mangoloid	Tibeto-Burman
Meitei	102	105	Imphal, Churachandrapur Moirang, Ukhrul	Agriculture, Carpenters, Fisherman. Upper Caste	Mongoloid	Tibeto-Burman
Naga	76	106	Eastern districts of Manipur mainly Imphal and Ukhrul	Shifting Cultivators, Labourer Tribe	Mongoloid	Tibeto-Burman
Kuki	75	105	Churachandrapur Moirang, Ukhrul	Shifting Cultivators, Weaving, Blacksmith Tribe	Mongoloid	Tibeto-Burman
Hmar	60	101	Imphal and Churachandrapur districts of Manipur	Shifting Cultivators (Jhum) Tribe	Mongoloid	Tibeto-Burman
Muslim of Manipur	66	101	Imphal	Trade, Agricultural and wage labour. Religious Group	Mixed Population of Indo-European and Tibeto-Burman	Indo-European and Tibeto-Burman

*Sample size for Profiler Plus loci, and HLDQA1 and five coapmplified loci.
**Sample size for three monoplex STR loci.

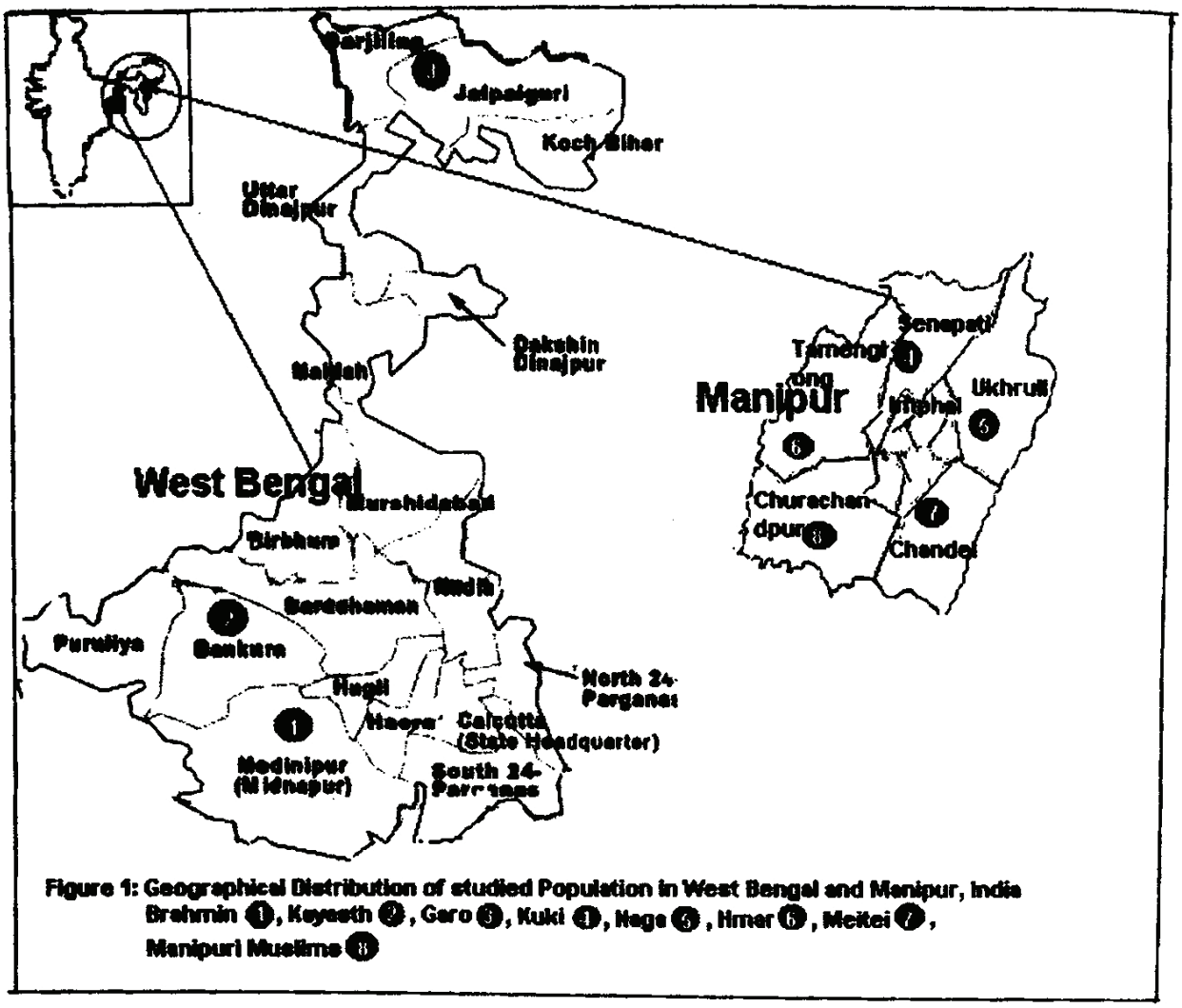


Fig. 1. Geographical location of the eight populations in West Bengal and in Manipur regions, India.

were presented earlier (Dutta et al., 2001a; Chattopadhyay et al., 2002) and some results are utilized here for comparison.

DNA Extraction

The samples were collected in EDTA-containing vacutainers, aliquoted (700 μL), and stored at –20°C prior to DNA extraction. DNA was extracted from aliquoted blood samples following a phenol-chloroform method (Comey et al., 1994). Briefly, the samples were incubated in 50 mM Tris-HCl, 150 mM NaCl, and 100 mM EDTA Na_2, with the addition of SDS (1.25%) and 0.03 mg/mL proteinase K, precipitated with absolute ethanol after two extractions with phenol:chloroform:isoamyl alcohol (24:1), respectively. The quantity of DNA in each sample was estimated using the slot-blot procedure described by Waye et al. (1989) and by Hoefer's DyNA Quant 200 Flurometer.

PCR amplification and typing

HLDQA1 and five polymorphic loci. Fifteen ng of DNA samples was amplified in a total volume of 50 ml reaction mixture in one multiplex PCR reaction and typed for genotyping of the six codons (HLA DQA1 and five coamplified loci) by using the Amplitype PM+DQA1 PCR Amplification and Typing Kit (Perkin Elmer, Boston, MA) according to the manufacturer's protocol (Amplitype: User Guide, 1998, and Cetus, Norwalk, CT). The amplified PCR products were hybridized to DNA probe strips and specifically bound amplified DNA was visualized upon

the enzymatic conversion of a colorless substance into a blue precipitate.

HPRTB, F13, and LPL. Three different monoplex PCR reactions were carried out to analyze the HPRTB, F13B, and LPL loci. Twenty-five ng of DNA was amplified in a total volume of 25 ml reaction mixture using a commercially available kit (Promega, Madison, WI). The amplified product was separated on 4% denaturing polyacrylamide gels. Following electrophoresis, the gels were stained using a silver staining kit from Promega. The repeat sizes were determined with respect to the standard HPRTB, F13B, and LPL allelic ladders supplied with the Gene Print Kit (Promega). Alleles at these loci have been designated by the size (in basepairs) of their PCR product and their corresponding repeat numbers.

Profiler plus STR loci. A total of 2.5 ng of extracted DNA was amplified for N STR loci using a commercial kit (Perkin Elmer) and run on an ABI 377 automated DNA sequencer. The resultant data analysis and allele designation was carried out using Genescan and Genotyper software. The details of VNTR and six polymorphic loci analyses were described previously (Dutta and Kashyap, 2001a; Chattopadhyay, 2002).

Statistical analysis

The allele frequency for each genotype in the sample set was calculated by the gene count method. Possible divergence from Hardy-Weinberg expectations (HWE) was determined on the basis of the likelihood ratio test (Weir, 1992) and the exact test (Guo and Thomson, 1992). The level of significance of the test, where applicable, was determined by shuffling 2,000 times for each test (Chakraborty et al., 1993). Average heterozygosity, genetic diversity, and genetic distance were estimated with the help of Genetic Distance and Phylogenetic Analysis Software (Nei, 1973, 1987; Nei et al., 1983) and the phylogenetic trees were drawn by unweighted group-method with arithmetic mean (UPGMA) (Sneath and Sokal, 1973). Both standard genetic distance and DA distance and dendrograms were calculated for the 12 STR loci and 18 loci. Since the VNTR loci lack the precision of scoring the alleles because of poor resolution on acrylamide gels and a limitation in the differentiation of the observed and real alleles, which cannot reflect the accurate information through UPGMA clustering, the bootstrap method (Felsenstein, 1985) was applied to test the stability of the topology of the UPGMA dendrogram. A replication of 1,000 times of the distance matrix was repeated at random with replacement. Each dendrogram was compared with each cluster with the original topology to examine the stability of the topology. The frequencies (percentages) of all the clusters are shown on the nodal points in the dendrogram. The high frequency of a cluster reflects its higher stability.

RESULTS

Variation in 12 STR loci

Distribution of allele frequency at 12 STR (nine profiler plus loci and three monoplex) loci (Fig. 2) shows that most common alleles were found at either similar or adjacent alleles. The data on allelic frequencies were published previously (Dutta et al., 2001a). Despite the wide range of the allelic variation in the nine profiler plus loci, the figure shows a discernable pattern that reflects the geographical and ethnic affiliation (and population structure) of the three West Bengal and five Manipur populations. In general, a few allelic frequencies of the nine loci show high values among the two West Bengal populations and low frequency in the Manipur populations (e.g., allele 16 of D3S1358, allele 15 of vWA, alleles 20 and 22 of FGA, allele 32.2 of D21S11), while a few other alleles show the opposite trend of greater frequency among Mongoloid populations (e.g., allele 17 of D3S1358, allele 17 of vWA, allele 24 of FGA, allele 29 of D21S11). Meitei, a caste population of Manipur, show intermediary frequency between the regional and the two caste populations of West Bengal, whereas Garo, a Mongoloid group from West Bengal, show similar trends, reflecting the influence of gene flow due to geographic proximity. Manipur Muslims, a migrant community of non-Mongoloid origin in Manipur, show some distinct allelic frequency in some loci, thus reflecting the influence of population structure variables. For example, it shows a high (percentage) frequency for alleles 15 (44.4%) and 16 (30.5%) at the D3S1358 locus, for alleles 9 (30.5%) and 12 (22.2%) at D18S51, for alleles 9 (38.8%) and 11 (25.0%)

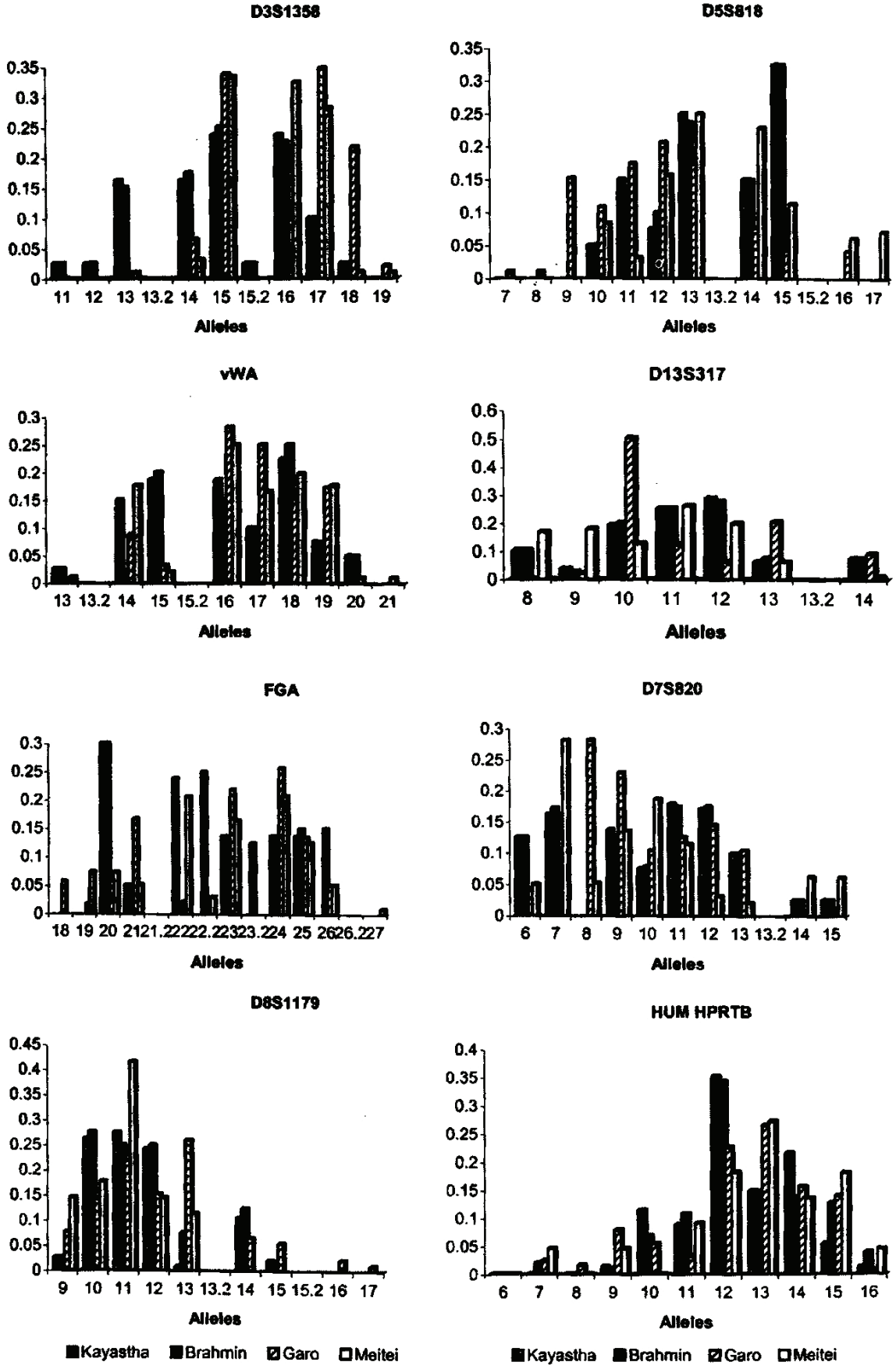


Fig. 2. Allelic frequencies for 12 STR loci in three West Bengal and five Manipur populations.

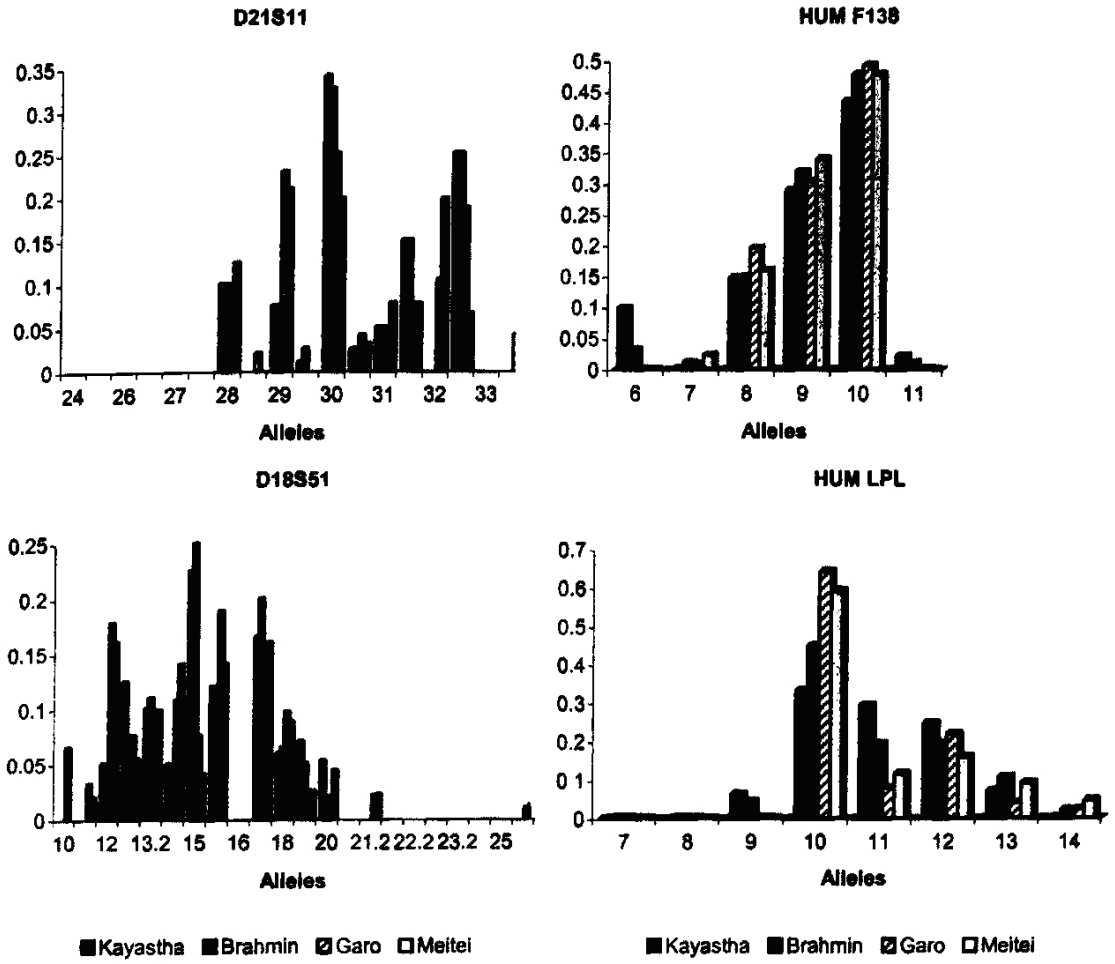


Fig. 2. *Continued.*

at D13S317 loci, and the absence of several other alleles at other loci.

Similar trends are also observed at three STR loci: HUM PRTB, F13B, LPL (Fig. 3a). The West Bengal group shows the highest frequency in allele 12 (0.226 to 0.351), whereas the Manipur group shows the highest frequency for allele 13 (0.272 to 0.400). Alleles 7, 8, 15, and 16 at HUMHPRTB were absent in all Manipur populations and alleles 13 and 14 at LPL were also absent in all, except in Meitei. Manipur Muslim shows a low frequency (11.1%) of allele 10 of F13B loci, while allele 11 is absent in the Mongoloid group. VNTR loci also showed similar trends (Dutta et al., 2001a).

Variation in six polymorphic coding loci

In the case of the HLADQA1 locus (Fig. 3b), the two caste populations, Kayastha and Brahmin, and three Manipur tribal populations, Naga, Kuki, and Hmar, show similar frequencies (Chattopadhyay, et al., 2002). Allele 2 is the most frequent in Kayastha and Brahmin, whereas allele 4.2/4.3 is the most frequent in the Manipur populations. The M.Muslim group shows high frequency (16.6%) for allele 1.3, lower frequency (2.7%) for allele 3, and the least (5.5%) for allele 1.2. In the case of five coapmplified coding loci (Fig. 2b), the Kuki population shows the lowest frequency for alleles HBGG (A), D7S8 (A), and GC (A). The Naga and Muslim groups from Manipur show the highest frequency of D7S8 (A) and GC (C) alleles, respectively.

The allelic frequencies were examined for deviation from Hardy-Weinberg expectation by the H-W exact test (Table 2). The results show no detectable deviation from H-W expectation for the six coding loci (HLD QA1 and five coapmplified loci (of PM Kit)), but in the case of (Profiler Plus) STR loci, deviation from HWE is noticed, for example,

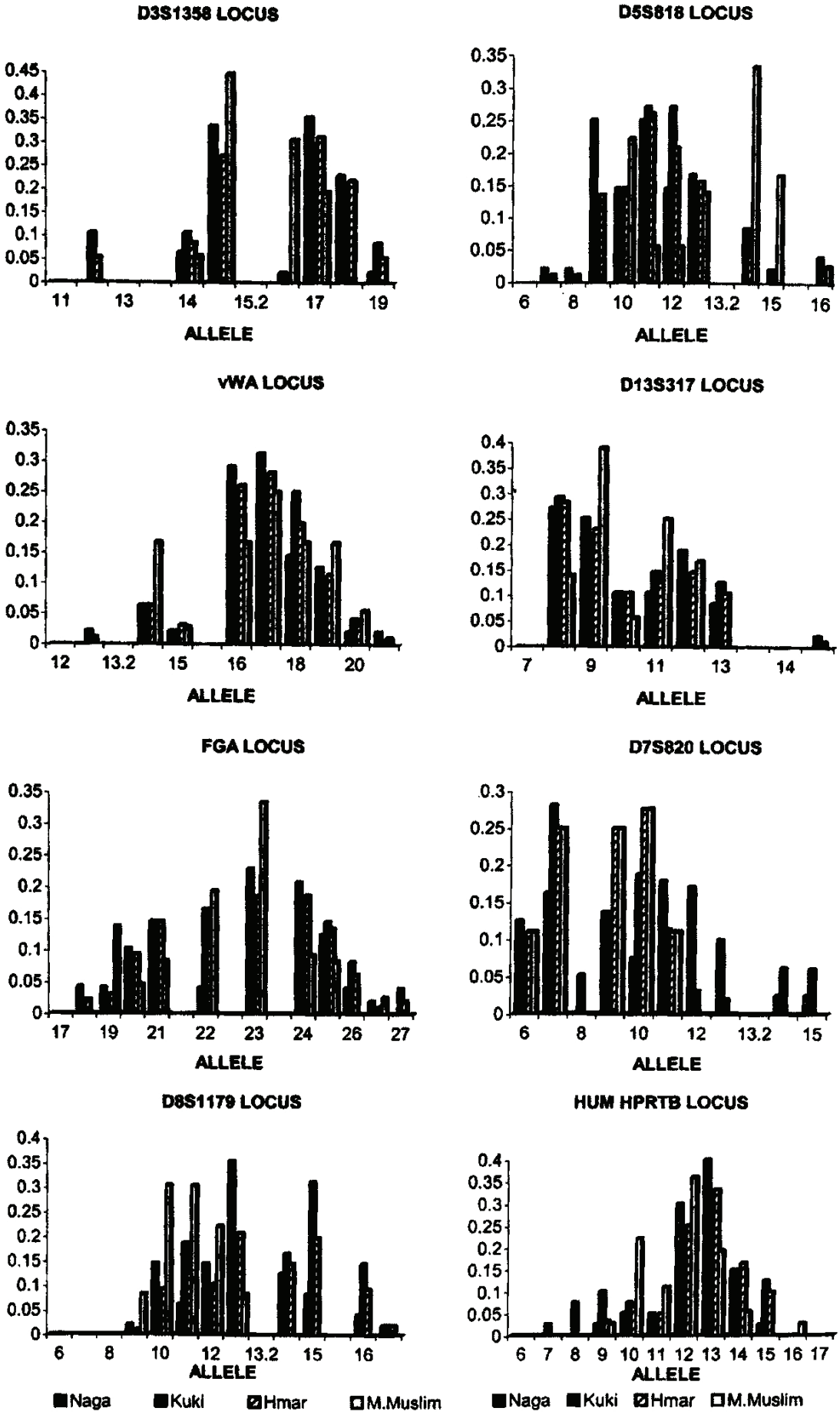


Fig. 2. *Continued.*

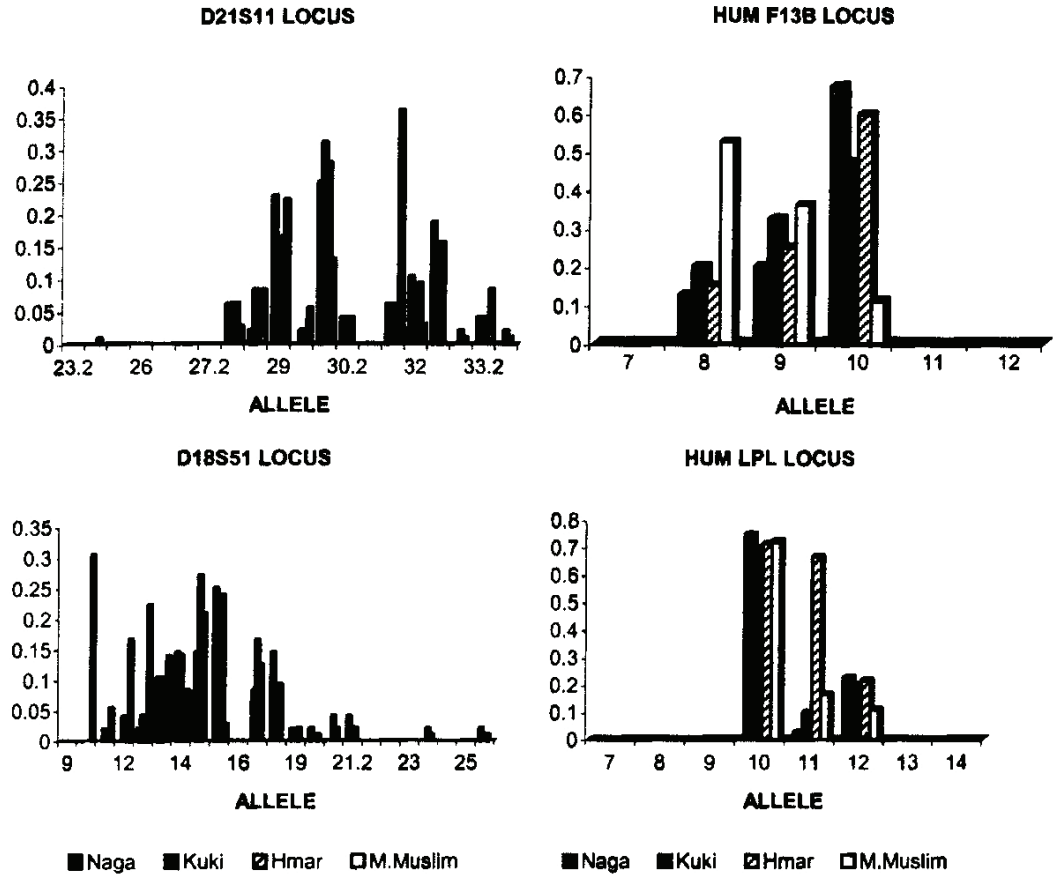


Fig. 2. *Continued.*

in Kayastha for D8S1179, D21S11, D13S317, and D7S820 loci; in Brahmin for D13S317 and D21S11; in Meitei for vWA and D7S820 loci; in Garo for D8S1179 loci; and in Kuki for vWA and FGA loci. Except for HUM F13B loci in the Meitei population, the other two Monoplex STR loci are in Hardy-Weinberg equilibrium for the studied populations.

Gene diversity

Average heterozygosity and G_{ST} estimates for each of the 22 loci of the three class of DNA markers, in both the expressed and nonexpressed regions, of the human genome were calculated for the eight populations (Table 3). In the case of the STR loci, D7S820 depicts the least variation among the populations (0.66 to 0.77), whereas both D3S1358 and FGA show a higher range of average heterozygosity values. In the case of six coding loci, the HLDQA1 locus and alleles in GC loci maintain higher average heterozygosity values. LDLR and D7S8 alleles (in the PM system) show the least variation. Between the different classes of DNA markers employed in the study, VNTR and STR demonstrate higher average heterozygosity than the six coding loci. The results indicate the influence of population structure variables. For example, the two caste populations from West Bengal, Kayastha and Brahmin, show lower values of average heterozygosity at most of the 12 STR alleles than the Garo and the five Manipur populations. The M.Muslim population shows higher average heterozygosity values at the loci: vWA, D5S818, D7S820 in the STR system, and in four of the five codon loci (LDLR, GYPA, D7S8, and HLADQA1). The Naga population shows higher average heterozygosity at the HLDQA1 locus and at GYPA and D7S8 loci. Average heterozygosity values for the combined 12 STR loci, 18 DNA loci (12 STRs and six coding loci), and at 22 DNA loci collectively (12 STR, six coding loci, and four VNTR) for the studied populations are shown in Table 4. The values for STR loci reflect higher average heterozygosity than for either the combined 18 DNA or 22 DNA loci. This could be due to the lower heterozygosity estimates obtained in the case of six coding loci.

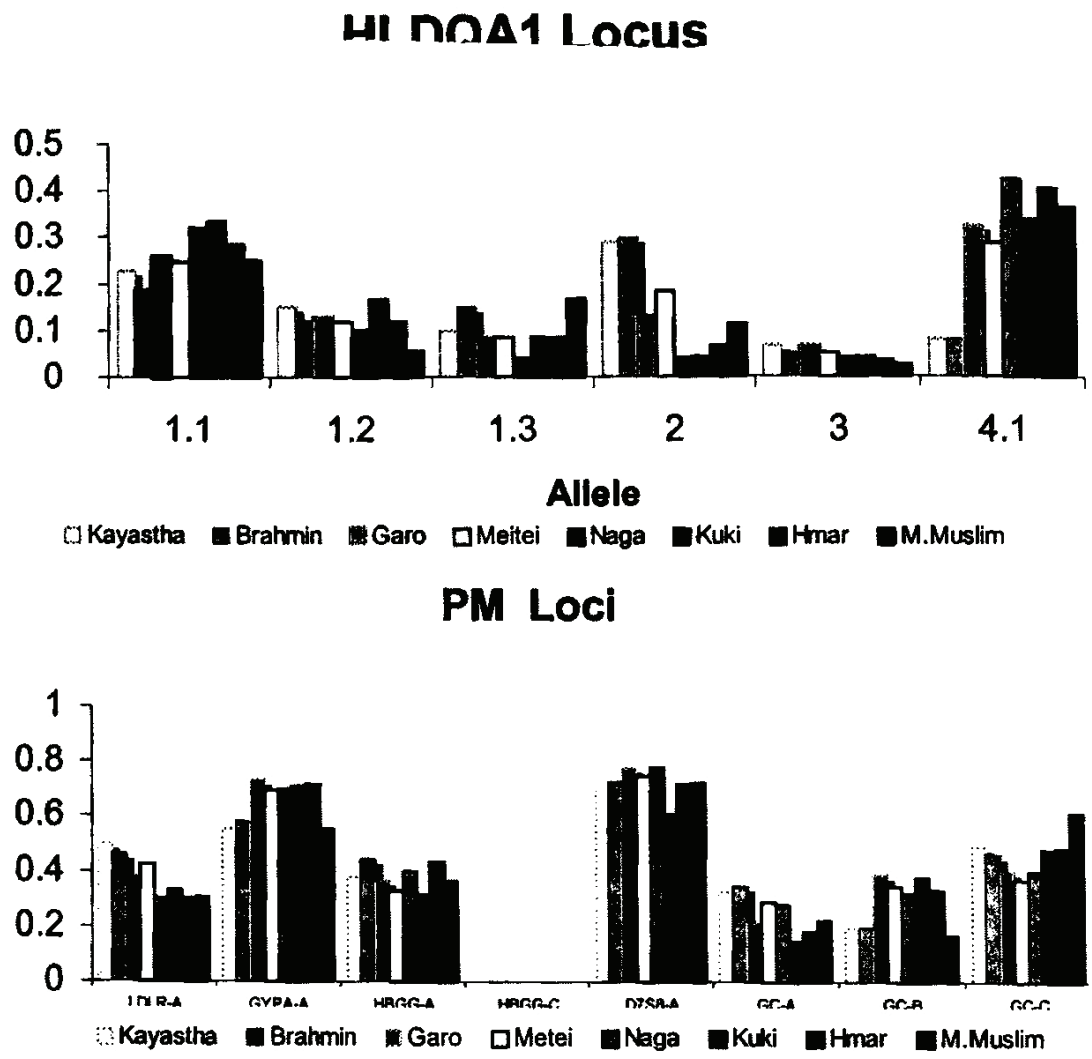


Fig. 3. Allelic frequencies for six coding loci (HLDQA1, five coapmplified loci) and in eight populations from West Bengal and Manipur.

The estimates G_{ST}, which reflect the extent of genetic differentiation for the studied loci (Table 3) show the least differentiation at VNTR, six coding loci than STR loci. Different alleles in each of the three types of DNA loci used for the study also show differences in the extent of gene diversity among the populations. Among STR loci, vWA depicts the least differentiation (2.3%), while D5S818 shows the highest (6.7%). In the case of the PM loci system HBGG allele shows the least values (0.7%), possibly suggesting greater stability than other alleles. Overall, the extent of gene differentiation between the populations for the combined 12 STR loci is 5% and for the combined 18 DNA and 22 DNA loci it is about 4% (Table 4).

Based on the 12 STR and 18 polymorphic loci collectively "standard genetic distance" and "genetic distance D_A" were computed to investigate the genetic affinity between the eight populations. Since the results for two distance matrices were consistent, we have shown "genetic distance D_A" values in the lower triangular for 18 DNA loci and upper triangular matrices for 12 STR loci (Table 5). In the case of the 12 STR loci, both Kayastha and Brahmin, the two caste populations from the West Bengal region, show the least value and Garo and M.Muslim populations show the highest value. Garo, a Mongoloid group from West Bengal, shows close affinity with the Naga population of Manipur. Among the Manipur populations, the Meitei caste group shows affinity with Naga, whereas Hmar and M.Muslim are distinctly different from the others. At the level of 18 polymorphic loci, which represent both the coding and noncod-

TABLE 2. Test of Hardy-Weinberg equilibrium (PH-W Exact test) for the 12 STR D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, HPRTB, F13B, and LPL and for the six coding (HLADQA1 and five coapmplified loci) and four VNTR loci (D1S7, D4S139, D5S110, and D17S79) in three West Bengal and in five populations of Manipur

	West Bengal			Manipur				
Locus	Kayastha	Brahmin	Garo	Meitei	Naga	Kuki	Hmar	MMuslim
STR Loci								
D3S1358	0.425	0.870	0.835	0.781	0.152	0.612	0.386	0.347
VWA	0.500	0.305	0.118	0.005	0.235	0.042	0.127	0.111
FGA	0.095	0.188	0.136	0.496	0.140	0.054	0.049	0.552
D8S1179	0.005	0.281	0.002	0.271	0.717	0.663	0.168	0.289
D21S11	0.002	0.006	0.119	0.710	0.166	0.208	0.187	0.603
D18S51	0.443	0.211	0.217	0.198	0.250	0.250	0.250	0.337
D5S818	0.912	0.988	0.117	0.583	0.708	0.856	0.730	0.383
D13S317	0.070	0.011	0.325	0.411	0.050	0.291	0.166	0.112
D7S820	0.050	0.166	0.214	0.064	0.434	0.171	0.129	0.813
HUM HPRTB	0.425	0.870	0.835	0.781	0.152	0.612	0.386	0.347
HUM F13B	0.500	0.305	0.118	0.005	0.235	0.042	0.127	0.111
HUM LPL	0.095	0.188	0.136	0.496	0.140	0.054	0.049	0.552
6 Coding Loci								
HLADQA1	0.175	0.274	0.068	0.092	0.074	0.864	0.214	0.228
LDLR	0.271	1.000	0.469	0.842	0.615	1.000	0.364	1.000
GYPA	0.095	0.774	0.567	0.824	0.632	1.000	1.000	0.337
HBGG	0.434	1.000	0.627	0.523	1.000	1.000	0.310	1.000
D7S8	0.045	0.077	0.731	0.118	1.000	1.000	1.000	0.253
Gc	0.198	0.120	0.348	0.183	0.788	0.398	1.000	0.065
4 VNTR Loci								
D1S7	0.242	0.206	0.289	0.087	0.246	0.184	0.085	0.111
D4S139	0.156	0.244	0.776	0.077	0.467	0.185	0.552	0.540
D5S110	0.957	0.579	0.134	0.074	0.081	0.609	0.367	0.150
D17S79	0.606	0.114	0.334	0.115	0.267	0.520	0.546	0.619

ing regions of the DNA, the "genetic distance D_A" (lower triangular matrix), Kayastha and Brahmin show the least values, and Garo is closer to Naga populations than others. Both Hmar and M.Muslim maintain the highest distance values with other regional populations.

A comparison of the pattern of clustering of the studied populations, based on genetic distance matrices, for each of the three class of DNA polymorphism (coding loci, VNTR, and STR) (Chattopadhyay et al., 2002; Dutta et al., 2001a) is considered separately (Fig. 4). The striking common features and differences in the pattern of clustering are conspicuous. In all three classes of DNA markers both Kayastha and Brahmin form a close cluster and are separated from the other six populations. The four tribal populations, Naga, Kuki, Garo, and Hmar, form a separate cluster in the case of dendrograms based on VNTR and STR loci. In contrast, in the case of six coding loci, the Kuki population is separate from the cluster, whereas Meitei, a compatriot caste Mongoloid group, is clustered with three compatriot tribes (Fig. 4b). One more significant difference is that in the case of 12 STR and VNTR loci, M.Muslim is separate from the other seven populations, whereas it is clustered with Meitei in Profiler Plus loci and in six coding loci M.Muslim clusters with the local three tribal populations of Manipur.

Based on the distance matrices for 12 STR loci and 18 DNA loci, two separate dendrograms depicting the different levels of clustering of related populations were drawn by following the UPGMA clustering method (Figs. 4a, 5). Both dendrograms show almost a similar clustering pattern. The difference is that the clustering based on 12 STR shows higher bootstrap values, suggesting more consistency than for the clustering based on combined STR and coding (18 DNA) loci. In the case of 18 DNA loci, the tree shows two main clusters. Manipur Muslims stand apart and are distinct from other populations. The smaller cluster has three populations where Kayastha and Brahmin form a close cluster and Meitei is distinctly related. The Manipur populations Naga, Hmar, and Kuki form another close cluster and Garo, a West Bengal Mongoloid population, is distinctly related. The bootstrap values indicate the

TABLE 3. Average heterozygosity for 12 STR loci (nine Profiler plus and three monplex STR loci), six coding (HLDQA, PM loci) and four VNTR loci in three West Bengal and five Manipur populations

	West Bengal			Manipur					
Locus N	Kayastha 114	Brahmin 51	Garo 80	Meitei 102	Naga 76	Kuki 66	Hmar 60	MMuslim 66	G_{ST}
12 STR Loci									
D3S1358	0.695	0.700	0.631	0.771	0.584	0.875	0.730	0.500	0.0648
vWA	0.650	0.600	0.798	0.875	0.667	0.834	0.750	0.889	0.0234
FGA	0.841	0.850	0.826	0.896	0.792	0.542	0.667	0.834	0.0342
D8S1179	0.650	0.600	0.674	0.688	0.792	0.792	0.792	0.667	0.0618
D21S11	0.600	0.650	0.888	0.788	0.885	0.775	0.580	0.811	0.0425
D18S51	0.700	0.680	0.778	0.882	0.869	0.895	0.700	0.712	0.0490
D5S818	0.795	0.800	0.783	0.813	0.792	0.834	0.813	0.889	0.0679
D13S317	0.700	0.688	0.797	0.759	0.700	0.821	0.781	0.701	0.0577
D7S820	0.725	0.700	0.696	0.750	0.667	0.709	0.688	0.778	0.0615
HUM HPRTB	0.695	0.700	0.631	0.771	0.584	0.875	0.730	0.500	0.0289
HUM F13B	0.650	0.600	0.798	0.875	0.667	0.834	0.750	0.889	0.0654
HUM LPL	0.841	0.850	0.826	0.896	0.792	0.542	0.667	0.834	0.0530
Average	0.711	0.688	0.760	0.813	0.732	0.777	0.721	0.750	0.0508
SE	0.022	0.026	0.023	0.020	0.029	0.034	0.018	0.040	0.0044
Six coding Loci									
HLADQA1	0.833	0.843	0.713	0.784	0.560	0.750	0.633	0.778	0.0400
LDLR	0.447	0.490	0.425	0.480	0.360	0.416	0.333	0.500	0.0226
GYPA	0.578	0.529	0.362	0.411	0.360	0.416	0.433	0.666	0.0234
HBGG	0.508	0.490	0.500	0.480	0.480	0.458	0.600	0.500	0.0078
D7S8	0.508	0.509	0.375	0.441	0.360	0.458	0.433	0.555	0.0134
GC	0.719	0.764	0.612	0.705	0.680	0.708	0.633	0.388	0.0253
Average	0.598	0.604	0.497	0.550	0.466	0.534	0.510	0.564	0.0221
SE	0.060	0.064	0.057	0.063	0.054	0.062	0.052	0.056	0.0045
Four VNTR Loci									
D1S7	0.959	0.945	0.931	0.882	0.918	0.917	0.913	0.907	0.0235
D4S139	0.827	0.897	0.902	0.910	0.897	0.890	0.906	0.685	0.0308
D5S110	0.948	0.891	0.936	0.916	0.896	0.927	0.940	0.915	0.0304
D17S79	0.731	0.685	0.830	0.796	0.833	0.815	0.828	0.795	0.0374
Average	0.882	0.854	0.899	0.876	0.886	0.887	0.897	0.825	0.0305
SE	0.052	0.057	0.024	0.027	0.018	0.025	0.024	0.054	0.0028

TABLE 4. Estimates of average heterozygosity and G_{ST} based on 12 STR Loci, 18 DNA loci (12 STRs and six codons) and 22 DNA loci (12 STRs, six coding and four VNTRs) in West Bengal and Manipur populations

	12 STR DNA loci*		18 DNA Loci**		22 DNA loci***	
Population and G_{ST}	Ave. Htz.	SE	Ave. Htz.	SE	Ave. Htz.	SE
1. Kayastha	0.7987	0.0147	0.7183	0.0345	0.7289	0.0583
2. Brahmin	0.7939	0.0169	0.7159	0.0347	0.7177	0.0515
3. Naga	0.7322	0.0422	0.6573	0.0421	0.7279	0.0836
4. Kuki	0.7808	0.0336	0.6943	0.0404	0.7689	0.0745
5. Hmar	0.7689	0.0382	0.6851	0.0420	0.6958	0.0874
6. Garo	0.7640	0.0301	0.6830	0.0403	0.7394	0.0742
7. Meitei	0.7851	0.0264	0.7017	0.0396	0.7114	0.0789
8. M.Muslim	0.7302	0.0326	0.6600	0.0364	0.7209	0.0553
Average G_{ST}	0.0505	0.0044	0.0439	0.0046	0.0392	0.0038

Ave. Htz.: Average Heterozygosity, SE: Standard Error.
*12 STR: Nine Profiler Plus and three monoplex STR.
**12 STR and six codons—HLDQA1 and 5 5 coapmplified loci.
***12 STR, six codons and four VNTR loci.

consistency of the clustering pattern. Except for Meitei (38%), the clustering of the other populations is more consistent, as they show higher bootstrap values. In the case of 12 STR loci the dendrogram shows (Fig. 4a) two main clusters. The least differentiated populations are Naga and Hmar, followed by Garo and Kuki, which form one main sub-

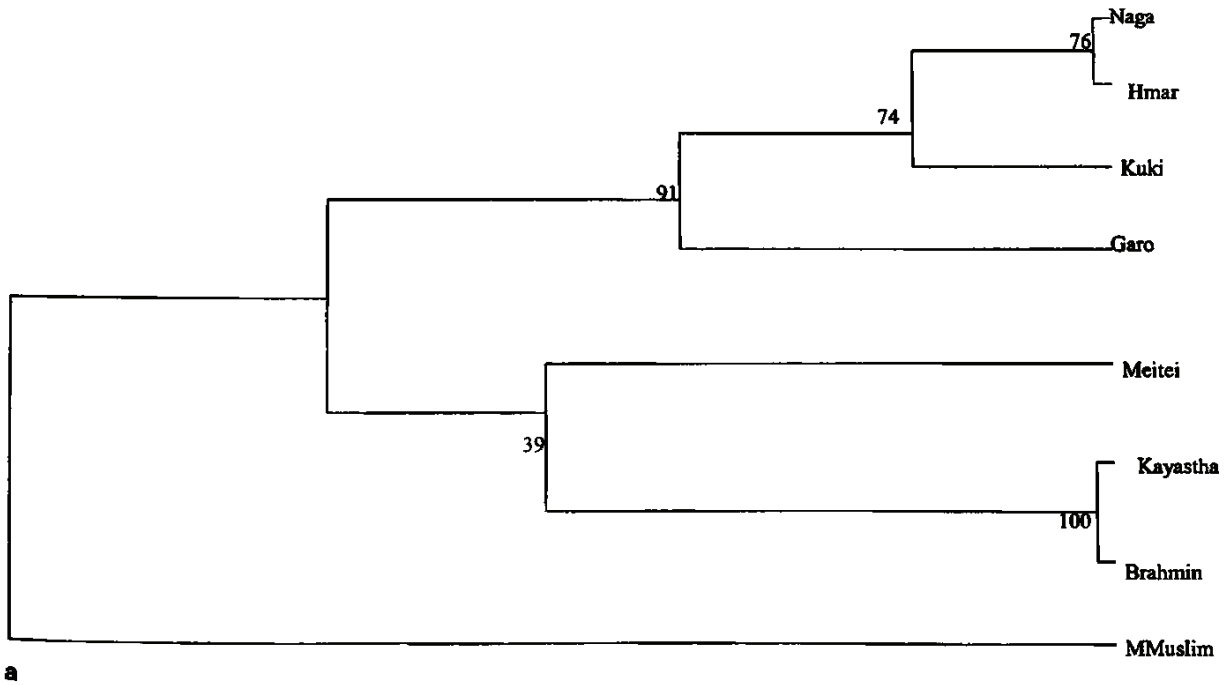
TABLE 5. Genetic distance (D_A) based on 12 STR loci (upper diagonal) and based on 18 DNA loci (9 profiler + 3 STR + 6 HLDQA1) (lower diagonal) between three West Bengal and five Manipur populations

	1	2	3	4	5	6	7	8
1. Kayastha		0.0083	0.1828	0.1646	0.1530	0.1605	0.1148	0.1848
2. Brahmin	0.0062		0.1851	0.1634	0.1549	0.1547	0.1037	0.1878
3. Garo	0.1325	0.1346		0.0843	0.0230	0.0852	0.1574	0.2351
4. Metei	0.1235	0.1247	0.0602		0.0261	0.1128	0.1483	0.2273
5. Naga	0.1121	0.1138	0.0165	0.0197		0.0787	0.1310	0.2129
6. Kuki	0.1138	0.1105	0.0586	0.0791	0.0538		0.1215	0.2148
7. Hmar	0.0810	0.0741	0.1078	0.1047	0.0901	0.0817		0.1347
8. M.Muslim	0.1305	0.1326	0.1611	0.1580	0.1450	0.1475	0.0941	

cluster. The Meitei stands apart from the four neighboring populations. The second cluster includes Kayastha and Brahmin from West Bengal, which form a separate cluster distinctly apart from the four Manipur populations. The Muslim population is distinct from all the populations and is separate from the others.

DISCUSSION

The results of the present study based on three classes of DNA polymorphism support the other studies based on classical genetic markers, especially that Caucasoid-affiliated populations (Brahmin, Kayastha, etc.) are distinct from the Mongoloid population. Further, the Mongoloid-affiliated populations of West Bengal depict close affinity with those in the northeast region (Roychoudhury, 1992). Allele frequencies indicate distinct differences between populations, possibly suggesting the influence of population structure variables, especially admixture and migration. The Mongoloid-affiliated population Garo in West Bengal, in general, displays intermediary frequencies in some of the alleles of STR loci. One of the interesting aspects of the results concerns the Muslim population in Manipur. Unfortunately, there is little information about the history or origin of the population. The fieldwork enquiry suggests that they are a migrant group from a nearby region of



a

Fig. 4. Genetic affinities between eight populations from West Bengal and Manipur regions based 12 STR loci (a), six coding (HLDQA1 and five coapmplified) loci (b), and four VNTR loci (c) by D_A distance and UPGMA clustering methods.

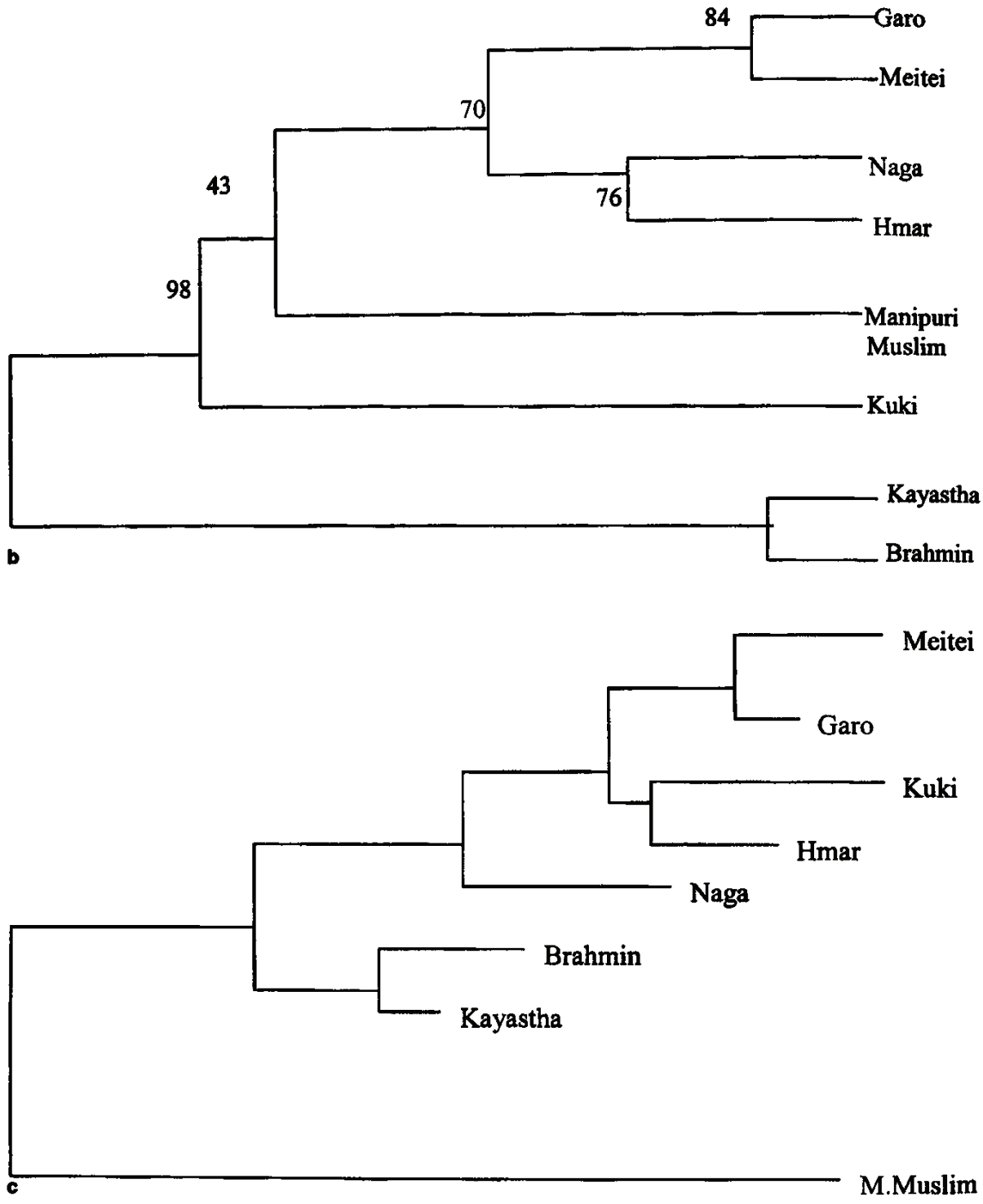


Fig. 4. *Continued.*

Assam in the past. Their physical features are different from the local Mongoloid ethnic features. They have remained as a distinct ethnic group and maintained their cultural norms and marriage regulations, possibly with the least admixture. This possibly accounts for some of the allelic differences observed in the study. STR loci reveal some distinct features of Manipur Muslim, possibly suggesting that they are genetically different from other local populations. The striking features are: the least frequency

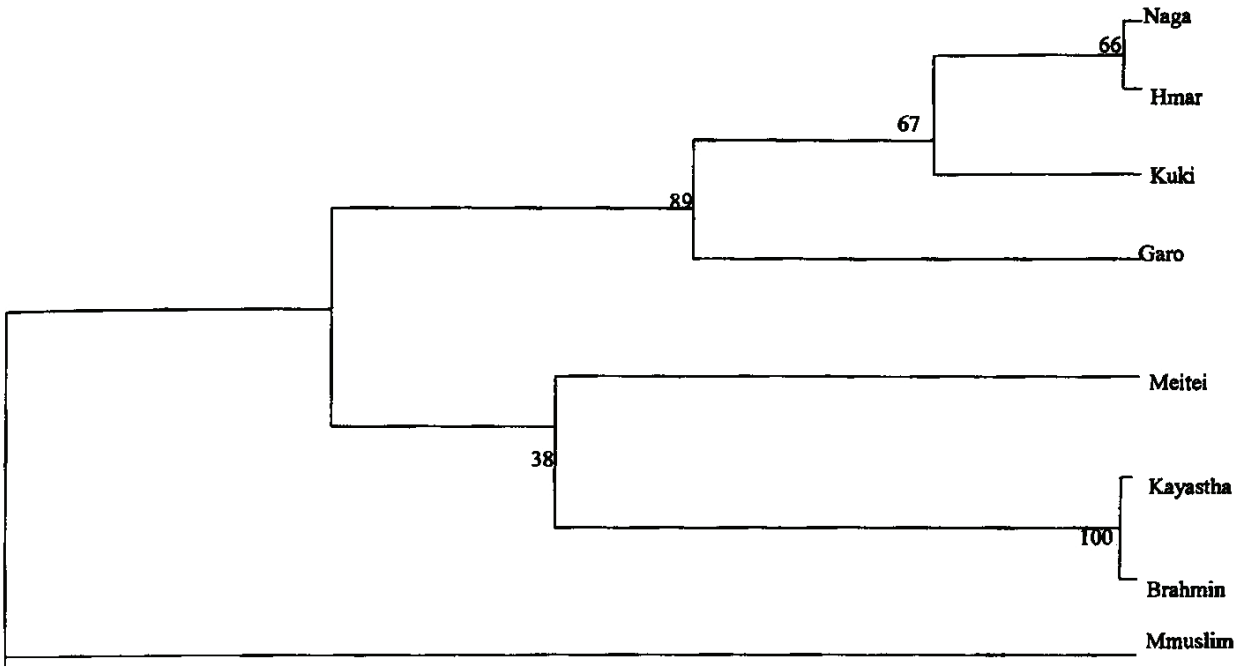


Fig. 5. Genetic affinities between eight populations from West Bengal and Manipur regions based on 18 DNA markers (nine Profile Plus STR and three STR and six HLDQA1 and five coapmplified loci) by D_A distance and UPGMA clustering methods.

(0.028) of allele 16 and absence of allele 15 at HPRTB locus and the lowest frequency (0.111) for allele 10 at F13B locus. Similar results have also been found by other studies in different ethnic populations in other regions of the continent. Bamshad et al. (1996) investigated the influence of population structure on genetic variation among four castes and inferred that the effects of social structure on mtDNA variation are much greater than those based on traditional markers. Balakrishnan et al. (1996) investigated the affinities of the Iyer, a Brahmin population of Tamil Nadu, based on HLA haplotypes and found that they differ from other Brahmin populations and possibly that their ancestors migrated from Central Asia or the Eurasian steppes. Papiha et al. (1996) found that the G_{ST} values based on VNTR loci are within the range found for classical markers studied from different populations of the Indian subcontinent. Mukherjee et al. (1999), studying the variation on four STR loci among eight population groups, indicated that geographic proximity has a greater influence than sociocultural affinity.

The dendrograms based on STR and VNTR loci for the studied eight populations demonstrate similar pattern of clustering: Caucasoid and Mongoloid groups are distinctly apart from each other, thus validating the usefulness of polymorphic loci for investigating the genetic affinities of regional populations. The NJ-tree, Manipur Muslim constitutes a separate group from the two subclusters. VNTR and STR sequences are extremely polymorphic and distributed in nonexpressed regions and show high mutation rates. As such, these two classes of loci are expected to show better resolution of genetic affinities and origin of the closely related regional and continental populations. However, the clustering pattern based on 12 STR and six coding loci (Fig. 4b) differs from the clustering obtained by the VNTR (Fig. 4c). Higher bootstrap values (92% in the case of Profiler Plus) suggest that the cluster is consistent, and thus it differs from VNTR and STR clustering patterns. The results of the clustering patterns suggest the importance of selecting different sets of DNA polymorphism for consistency and validation of the genetic affinities between populations. This also suggests that the expressed and nonexpressed regions might reflect different micro-evolutionary aspects. Coding genes HLDQA1 and other 5-coapmplified loci with low mutation rates, but with greater stability and selection pressures, might indicate the influence of population variables when compared to

VNTR and STR loci. However, the clustering pattern based on combined 12 STR loci and 18 DNA loci (combined STR and coding loci) provide similar clustering. This could be due the preponderance of STR and VNTR loci possibly masking the influence of those in the expressed region. The clustering pattern based on STR and VNTR loci corresponds with the geographic distribution of the Manipur Mongoloid populations. Garo M.Muslim population do not follow the geographic regional contiguity but concur with the historical, cultural, and linguistic affiliation. In another recent study based on a different set of STR loci among eight regional populations from a wider geographic region that include Orissa, Uttara Pradesh, and West Bengal, suggested poor affinity with sociocultural proximity (Mukherjee et al., 1999). The variance in results for different classes of polymorphic DNA loci indeed could be due to several reasons; for example, the nature of populations studied, and the nature of population structure, especially the influence of gene flow. The selection, type, and number of STR (or VNTR) loci considered for the study is important in this regard. Although it is expected that the findings of the study may not differ much on STR loci selected, it might differ with the class or type of loci selected, in the expressed and nonexpressed region.

Although there are some reports on STR and VNTR loci diversity in other parts of the world (Deka et al., 1995) and also among Indian populations, they were based on different sets of loci (for example, Balakrishnan et al., 1996; Mukherjee et al., 1999). Our results display higher levels of average heterozygosity for the two Caucasoid populations (73%) than the Mongoloid populations. Higher levels of heterozygosity for different sets of DNA polymorphic data in the Caucasoid population was also recently reported (for example, Martinovic et al., 1999). But the observed heterozygosity levels (53–66%) for the studied DNA loci for the Mongoloid samples of the present study are similar to the values reported for Cambodian (69.4) and Filipino (68.6%) at nine autosomal microsatellite loci (Parra et al., 1999). But the values observed in our study are below the reported values for other Mongoloid populations. For example, Robinson et al. (1996) reported an average heterozygosity of 76.3% (based on three STR loci) for immigrant Vietnamese and Hongkong Chinese in Australia. Although the heterozygous levels differ with respect to the type of loci used for the study, it also reflects the influence of population structure variables, especially marriage pattern and migration, which need to be investigated to understand the genetic diversity of the populations. The coefficient of gene differentiation G_{ST}, which is a measure of the relative magnitude of a genetic contribution that can be attributed to subdivision, differs between the types of DNA marker. The average G_{ST} coefficient for the STR loci is 5.3% for the studied populations, and for the combined 18 (STR and coding loci) and 22 loci it is about 4%. Overall, the estimated G_{ST} values of the study are similar to other studied populations of other Indian and global populations. The genetic differentiation for different sets of DNA loci was found consistent with classical markers (Das et al., 1987; Deka et al., 1988; Robinson et al., 1996; Das et al., 1996, etc.), although the G_{ST} estimated values in the present study were lower that other such studies on Mongoloid populations.

The available historical records suggest that different Mongoloid populations, especially in Manipur and other nearby states in the northeast region, are believed to be migrants from Myanmar (Burma) and adopted different lifestyles in their pursuit of survival in their adopted region in the recent past. Recent studies based on STR loci also exhibit clustering of northeastern populations with Chinese and other neighboring Mongoloid populations (Reddy et al., 2001) and form a separate cluster in comparison with other Indian populations (Dutta and Kashyap, 2001a, 2001b). The genetic similarity possibly suggests a common genetic affinity and the least genetic differentiation irrespective of cultural differences among them, but show distinct differences with the Caucasoid populations of the region.

ACKNOWLEDGMENTS

We thank Dr. R. Trivedi for providing logistic support in preparing the manuscript, and the laboratory and administrative help for completion of the work provided by the Director, Central Forensic Science Laboratory, Kolkata, India.

LITERATURE CITED

Balakrishanan K, Pitchappan RM, Suzuki K, Sankar Kumar U, Santha Kumari R, Tokunaga K. 1996. HLA affinities of Iyers, a Brahmin population of Tamil Nadu. Hum Biol 68:523–537.

Bamshad M, Farley AE, Crawford MH, Cann RL, Busi BR, Naidu JM, Jorde LB. 1996. Mt DNA variaton in caste populations of Andhra Pradesh, India. Hum Biol 68:1–26.

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

Chakraborty R, Jin L, Zhong Y. 1993. On allele frequency compilation from DNA typing Data. Int J Leg Med 106: 103–106.

Chattopadhyay P, Vasulu TS, Kashyap VK. 2002. Genetic affinity among eight ethnic populations of West Bengal and Manipur, India: a study based on six polymorphic functional loci (HLDQA1, LDLR, GYPA, HBGG, D7S8 and GC). Int J Hum Genet 2: 233–242.

Clark AG. 1987. Neutrality tests of highly polymorphic restriction fragment length polymorphisms. Am J Hum Genet 41:948–956.

Comey CT, Kooms BW, Presley KW, Smerick JB, Sobieralski CA, Stanley DM, Baechtel FS. 1994. DNA extraction strategies for amplified fragment length polymorphisms analysis. J Forensic Sci 39:1254–1269.

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 Anthropol 2:325–340.

Das KC, Malhotra KC, Mukherjee BN, Wlater H, Majumder PP, Papiha SS. 1996. Population structure and genetic differentiation among 16 tribal populations of Central India. Hum Biol 68:679–705.

Deka R, Reddy AP, Mukherjee BN, Das BM, Banerjee S, Roy M, Dey B, Malhotra KC, Walter H. 1988. Hemoglobin E distribution in ten endogamous populations 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.

Deka R, Shriver MD, Yu LM, Heidrecich EM, Jin L, Zhong Y, McGarvey ST, Agarwal SS. 1999. Genetic variation at twenty three microsatellite loci in sixteen human populations. J Genet 78:99–121.

Dutta R, Kashyap VK. 2001a. Genetic variation at minisatellite loci D1S7, D4S139, D5S110 and D17S79 among 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.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

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

Jin L, Chakraborty R. 1995. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. Heredity 74:274–285.

Martinovic I, Brac L, Furac I, Janicijevic B, Kubat M, Pericic M, Vidoric B, Rudan P. 1999. STR polymorphism in the population of the island of Hvar. Hum Biol 71:341–352.

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 Natl Acad Sci USA 70:3321–3323.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

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

Papiha SS. 1996. Genetic variation in India. Hum Biol 68:607–628.

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.

Parra E, Saha N, Soemantri AG, Mcgarvey ST, Hundrieser J, Shriver MD, Deka R. 1999. Genetic variation at nine autosomal microsatellite loci in Asian and Pacific population. Hum Biol 71:757–779.

Pitchappan RM. 1986. HLA and India. In: HLA in Asia-Oceania. Aizawa M, editor. Sapporo, Japan: Hokkaido University Press. p 535–538.

Pitchappan RM. 1988. Founder effect explains the distribution of the HLA-A1-B17 but not the absence of the A1-B 8 haplotype in India. J Genet 67:101–111.

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.

Robinson S, Gutowski S, Van Oorschot RAH, Fripp Y, Mitchell J. 1996. Genetic diversity among selected ethnic subpopulations of Australia: evidence from three highly polymorphic DNA loci. Hum Biol 68: 489–508.

RoyChoudhury AK. 1992. Genetic relationships of the populations in eastern India. Ann Hum Biol 19: 489–501.

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

Sneath PHA, Sokal RR. 1973. Numerical taxonomy. San Francisco: WH Freeman.

Way JS, Presley L Budwole. 1989. A simple method for quantifying human genomic DNA in forensic specimen extracts. Biotechinques 7:852–855.

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