ANNALS OF HUMAN BIOLOGY, 2003, VOL. 30, NO. 2, 117–131

Geographic contiguity and genetic affinity among five ethnic populations of Manipur, India: further molecular studies based on VNTR and STR loci

DUTTA RANJAN†, R. TRIVEDI†, T. S. VASULU‡ and V. K. KASHYAP†

† Central Forensic Science Laboratory, Calcutta, India
‡ AHGU, Indian Statistical Institute, Calcutta, India

Summary. *Background*: Apart from traditional markers studied among a few numerically small, geographically defined surveys among Mongoloid populations in northeastern parts of India, very little is known about their genomic diversity at the molecular level.
Primary objective: This study seeks to investigate how best the variable number tandem repeat (VNTR) and short tandem repeat (STR) loci together can detect the patterns of the genetic affinity among five geographically contiguous, linguistically and socio-culturally diverse Mongoloid-affiliated populations of Manipur in northeastern regions of India.
Subject and methods: Blood samples were collected from unrelated and randomly selected volunteers of five ethnic populations (Meitei, Kuki, Naga, Hmar and Manipuri Muslim) from different parts of the state. Allelic variation in four minisatellite loci (D1S7, D4S139, D5S110 and D17S79) and three STR loci (vWA, FESFPS and F13AO1) was studied.
Results: Average heterozygosity values among the five groups for the minisatellite range from 68% to 94%, while the hypervariable three STR loci were between 60% and 88%. In the populations, all the studied loci were highly polymorphic, with almost no departure from Hardy–Weinberg equilibrium. The gene differentiation for the VNTR loci was lower and moderate ($G_{st} = 0.030$) in comparison with microsatellites ($G_{st} = 0.043$). The neighbour-joining method of clustering based on both type of molecular markers reveals a close cluster for the tribal groups of Kuki, Naga and Hmar, while Manipur Muslim stand distinct in both the trees. The clustering pattern obtained from the combined DNA marker loci matches more closely the pattern from STR loci than that obtained from VNTR loci.
Conclusions: The results reinforce that using both VNTR and STR loci in detecting regional genetic affinity among the populations is more effective than using VNTR or STR independently, and also confirm the results obtained from the serological and electrophoretic data. However, the clustering pattern obtained from combined DNA markers is more in conformity with the pattern obtained by STR loci rather than with VNTR loci. Despite linguistic, geographical and cultural barriers, the populations show genetic affinity among the four populations except in the case of the migrant Manipur Muslim group.

1. Introduction

Indian populations represent several endogamous groups of varying size of castes, subcastes and tribes of diverse ethnic, linguistic, religious, racial, cultural stocks in different geographic regions of the country. This subdivided population structure and the nature of this vast diversity is of interest among population geneticists and anthropologists to investigate the extent and nature of genetic differentiation and its microevolutionary process. The empirical studies based on data mostly from serological and electrophoretic variants have revealed that, in general, geographic proximity has a greater influence on genetic structure of populations than socio-cultural affiliations among different regional castes. However the Mongoloid tribal groups in northeastern and eastern regions show greater genetic affinity than geographical proximity (Roychoudhury 1983, Malhotra and Vasulu 1993, Bhasin *et al.* 1994, Papiha 1996, Majumder 1998).

These findings can now be validated with the help of hypervariable loci at the genomic level. In this regard, variable number tandem repeat (VNTR) and short tandem repeat (STR) loci are found to be useful for investigating the population structure and population genetics of regional and global populations (Chakraborty 1990, Edwards *et al.* 1992, Balzas and Baird 1993, Jin and Chakraborty 1995, Papiha *et al.* 1996, Deka *et al.* 1995, Papiha *et al.* 1998). For example, studies on DNA polymorphism in some Indian populations have examined the genomic diversity and affinity and some have verified the results obtained from the traditional genetic markers. However, most of these studies are based on STR loci and very few have used VNTR loci (Karutha Pandian *et al.* 1995a,b, Papiha *et al.* 1996, Mukherjee *et al.* 1999, Mukherjee *et al.* 2000, Dutta and Kashyap 2001a,b, Reddy *et al.* 2001). In this regard it assumes significance to use both VNTR and STR loci together in assessing the genetic affinities among closely related regional populations. VNTR loci exhibit higher polymorphism and mutation rate but lesser resolution, and are used for identification and in understanding genetic relationship (e.g. Budowle *et al.* 1991a, Mastana 1999), whereas STR loci show relatively low mutation rates, but precise analytic power and are therefore preferred for detecting population substructure (Jin and Chakraborty 1995). Thus the use of both VNTR and STR loci reinforces the results of the analysis.

Our study concentrates on the genetic variation in the northeast region of India. Of all the regions, the northeast is inhabited by mostly Mongoloid groups and are represented by several tribal populations, those have migrated from eastern and southeast Asian regions and settled in different states. Though 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 are not fully explored due to relative isolation and accessibility and other practical problems. However Roychoudhury (1992), based on serological data on some tribal populations of Mongoloid origin of these region, 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'. Here we consider one such region, Manipur, the northeast-end part of India, close to Myanmar, which represents a unique population structure. It represents several tribal populations and some have adapted to the Hindu caste social structure whereas some have migrated from other parts of India. This possibly represents wide diversity—some are expected to show greater affinity to other Mongoloid regional populations and some might show mosaic genetic structure. Hence this situation might not support the observation of Roychoudhury (1992). The validation of the above hypothesis based on hypervariable polymorphic loci (VNTR and STR) at the genomic level is the main aim of the present study.

The study is based upon analysis of four VNTR and three STR loci in five ethnic population groups of Manipur region: a caste group (Meitei), 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 practise 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 practise consanguineous marriages. While Naga, Kuki and Hmar speak the local Manipur language, the Meities speak Hindi and Bengali more commonly (Singh 1998). The present study employs four minisa-

tellite, especially the loci D1S7, D4S139, D5S110 and D17S79, chromosome location: 1p33–35, 4q35, 5p and 17q21. These were well documented throughout the world as an effective tool for analysis of genetic variation (Chakraborty and Jin 1992, Papiha *et al.* 1996). The STR loci are characterized by tandomly repeated sequences 1–4 bp in length, highly polymorphic and abundant in the human genome. The STR system vWA, FESFPS and F13AO1 (chromosomal location: 12p12-pter, 15q25-pter and 6q24.3-p25.1) chosen for the study is a well characterized allelic systems (Hammond *et al.* 1994, Puers *et al.* 1993, and others). Our objectives of the study are: (1) To utilize both VNTR and STR loci independently and together in assessing the genetic affinities among the regional populations. (2) To verify the finding that Mongoloid populations show genetic similarity with respect to their geographic contiguity irrespective of their language affiliation or origin and that they are distinct from the Caucasoid population.

2. Materials and methods

2.1. Populations sampled

Collection of blood samples was carried out from several regions of the state of Manipur. The participants of the study were collected randomly from Meitei and Manipuri Muslims from Imphal city, while Kuki, Naga and Hmar tribal were from the hills bordering the sate from the adjoining regions of Assam and Arunachal Pradesh. The Naga participants were from Tankul and Kabui, the two major subgroups of the population.

Blood samples were obtained by venipuncture in EDTA-coated vacutainers from unrelated, volunteered donors. Samples from the respective communities were stored in ice containers during transportation from the field of collection to CFSL, Calcutta. The collected samples Kuki (no. 75), Naga (no. 78), Meitei (no. 128), Hmar (no. 60) and Manipuri Muslim (no. 65) were subjected for testing. DNA isolation 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).

2.2. VNTR analysis

VNTR loci in the study were analysed as per the method of Jeffreys *et al.* (1985).

2.2.1. Digestion, electrophoresis and hybridization. High molecular weight DNA (200 ng) 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 (Gibco BRL Gaithersburg, MD, USA) specific for loci D1S7, D4S139, D5S110 and D17S79, respectively. The membranes were finally treated with lumiphos reagent in the dark before exposing them to films for development of profiles. Developed profiles were compared with respect to the 22 kb ladder and their sizes were ascertained with computer-assisted image analysis software (BIORAD 'Diversity Database', Hercules, CA). Approximate molecular weight (base pair) sizing estimates for each sample was obtained. Following development of profiles, membranes were stripped and rehybridized with subsequent probes. The band size (range) of VNTR loci (D1S7, D4S139, D5S110 and D17S79) are: 0892-4821, 1353-8452, 0964-8452 and 1078-2862, respectively.

2.3. STR analysis

The three microsatellites vWA, FESEPS and F13A01 were analysed by amplifying the product analyser detection by silver stain method (Bassam 1991).

2.3.1. Multiplex PCR amplification and typing. All the microsatellite were analysed by co-amplification using 25 μL final reaction volume containing 10 ng of sample DNA and 0.75 units of TaqDNA polymerase with STR buffer and primers specific for the multiplex. The primers and buffers were obtained from Promega Corporation, Madison, Wisconsin, USA. PCR amplification was carried out at PE 2400 model PCR (PE Biosystems, Foster City) following the manufacturer's protocols. The samples were prepared through denatured PAGE. The gel was silver-stained and allele designation of sample fragments was carried out in comparison with the standard allelic ladders supplied by the manufacturer.

2.4. Statistical analysis

The size estimates of VNTR 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 Budowle *et al.* (1991b). The unbiased and expected estimate of the VNTR profiles and number of genotype in the sample set was used for determination of the expected heterozygosity (Edwards *et al.* 1992). Alleles for STR loci were scored by a manual method in relation to allele ladders provided by primer kits. The expected number of distinct homozygous and heterozygous classes and their standard errors (SE) was calculated for the STRs according to the method described by Chakraborty *et al.* (1991).

Possible divergence from Hardy–Weinberg expectation (HWE) was determined by calculating the unbiased estimates of the expected homozygotes and heterozygote frequencies (Nei and Roychoudhury 1974, Nei 1978), the likelihood ratio test (Chakraborty *et al.* 1991, Weir 1992) and exact test (Guo and Thompson 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 1993). An interclass correlation criterion (Karlin *et al.* 1981) for two-locus association was used for detection of the disequilibrium between the set of VNTR and STR loci. In addition, to determine the linkage disequilibrium between the set of VNTR and STR loci, variance (S_k^2) of the number of heterozygous 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) were computed according to methods described by Nei *et al.* (1983). Neighbour-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 bootstrapping method with 500 replications.

3. Results

Distribution of frequency occurring bins at the VNTR (figure 1) and STR (figure 2) loci among the five Manipuri populations show that the most frequent fragments in both the DNA markers occurring either in the same or adjacent bins or alleles. We have estimated the accordance of these loci with predictions of Hardy–Weinberg

equation. HWE-homozygosity test, likelihood ratio test and exact test were computed to observe possible departure from HWE (table 1). The data for the HWE showed deviation from the HWE expectations (values less than 0.050) at the D5S110 locus for the Kuki and Hmar population. An apparent departure from HWE at some of the loci showing the deviations were re-analysed following presence of non-detectable allele frequency *r* (Chakraborty *et al.* 1994). The re-analysed data was in conformity with HWE expectations. The value of *r* was between 0 and 4% among the 20 population–VNTR loci combinations. These values were consistent with the frequency of non-detectable alleles observed among other populations using VNTR markers (FBI Laboratory Division 1993, Chakraborty *et al.* 1994, Sovinsky *et al.* 1996, Dutta and Kashyap 2001a,b). No such deviations were encountered in the analysis of the STR data among the five populations for the three loci studied for confirmation of HWE (table 1).

Independence of alleles at any of the pairwise comparison of the loci was calculated among the five populations (data not shown) for the four VNTR and three STR loci. No association among the alleles among all the studied loci in the selected populations were observed. 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). Random data sets generated by shuffling the data 2000 times were used to determine the upper and lower boundaries within 95% CI. The S_k^2 values did not show any association across the four VNTR and three STR loci among the population groups.

The absence of association across the studied minisatellite and microsatellite markers prompted us to calculate the probability of match, suggesting the uniqueness of the profile obtained using these markers among the Indian populations. The probability of match (P_M) for the four VNTRs obtained was between 1.7×10^{-5} and 3.2×10^{-7}. On the other hand, combined P_M and for the three STRs was between 1.6×10^{4} and 6.5×10^{-5}. The cumulative P_M obtained using all the seven hypervariable loci was between 4.54×10^{-9} and 2.0×10^{-11}. The match probabilities were similar to our earlier observations using these markers for Indian populations (Dutta and Kashyap 2001a,b).

Locus-wise heterozygosities and G_{st} values for each of the four VNTR and three STR loci among Meitei, Naga, Kuki, Hmar and Manipuri Muslim populations are shown in table 2. Higher average heterozygosity values (between 82% and 89%) were obtained at the minisatellite repeat systems, in comparison with the microsatellite markers (between 71% and 75%). The average G_{st} value for the four minisatellite markers (0.03) was lower than the microsatellite loci (0.043).

The matrix of genetic distance D_A for both VNTR and STR loci for the studied population groups is shown in table 3. In the case of VNTR the Meitei, Naga, Kuki and Hmar populations did not show any significant distance between them and the distance exhibited by these groups with the VNTRs were consistent with the observations of the microsatellite markers. Based on the distance matrix, neighbour-joining trees were constructed depicting the affinities between the populations (figure 3) at the VNTR and STR loci. In the neighbour-joining plot using VNTR markers, population groups of Brahmin, Kayastha and Garo, from West Bengal (Dutta and Kashyap 2001a) and three northwestern population samples of Hindu, Sikh and Punjabi (Sovinsky *et al.* 1996) were included. Similarly for the neighbour-joining plot using the STR markers, data from the Brahmin, Kayastha and Garo were included for comparison with other Indian populations (Dutta and Kashyap

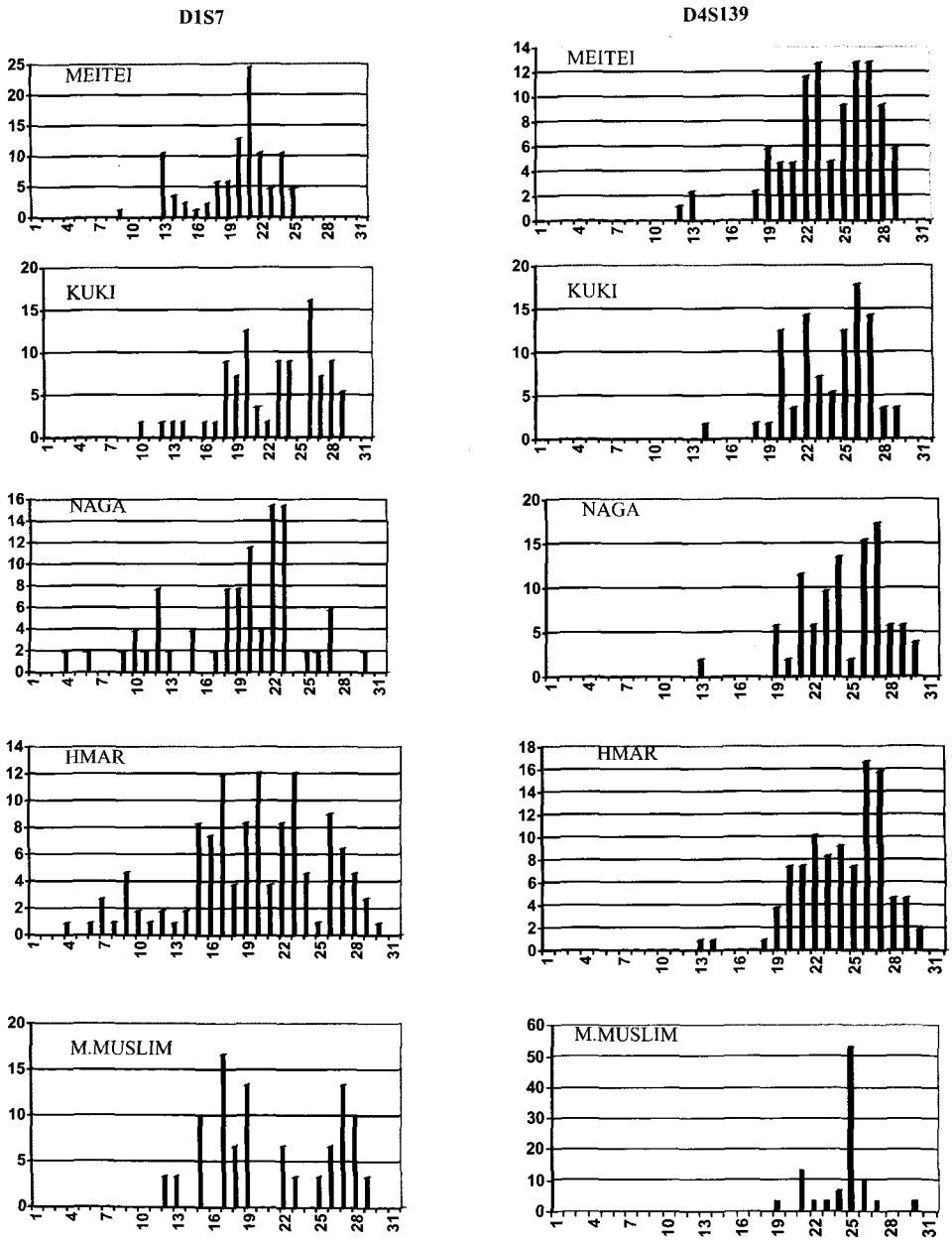

Figure 1. Fixed bin allele distribution at four VNTR loci in five Manipuri populations.

2001b). The data obtained in the current study was, however, used in plotting the third dendrogram incorporating all the seven hypervariable loci. The plot using the seven loci was similar to the plot obtained using STRs. In the VNTR plot, three distinct branches were observed. The northeastern Indian populations were separated from the other Indian populations at 94% of bootstraps. The Indo-Caucasoid populations of Brahmin, Kayastha, Hindi, Punjabi and Sikh populations formed the second branching with a bootstrap value of 87%. The close association between the

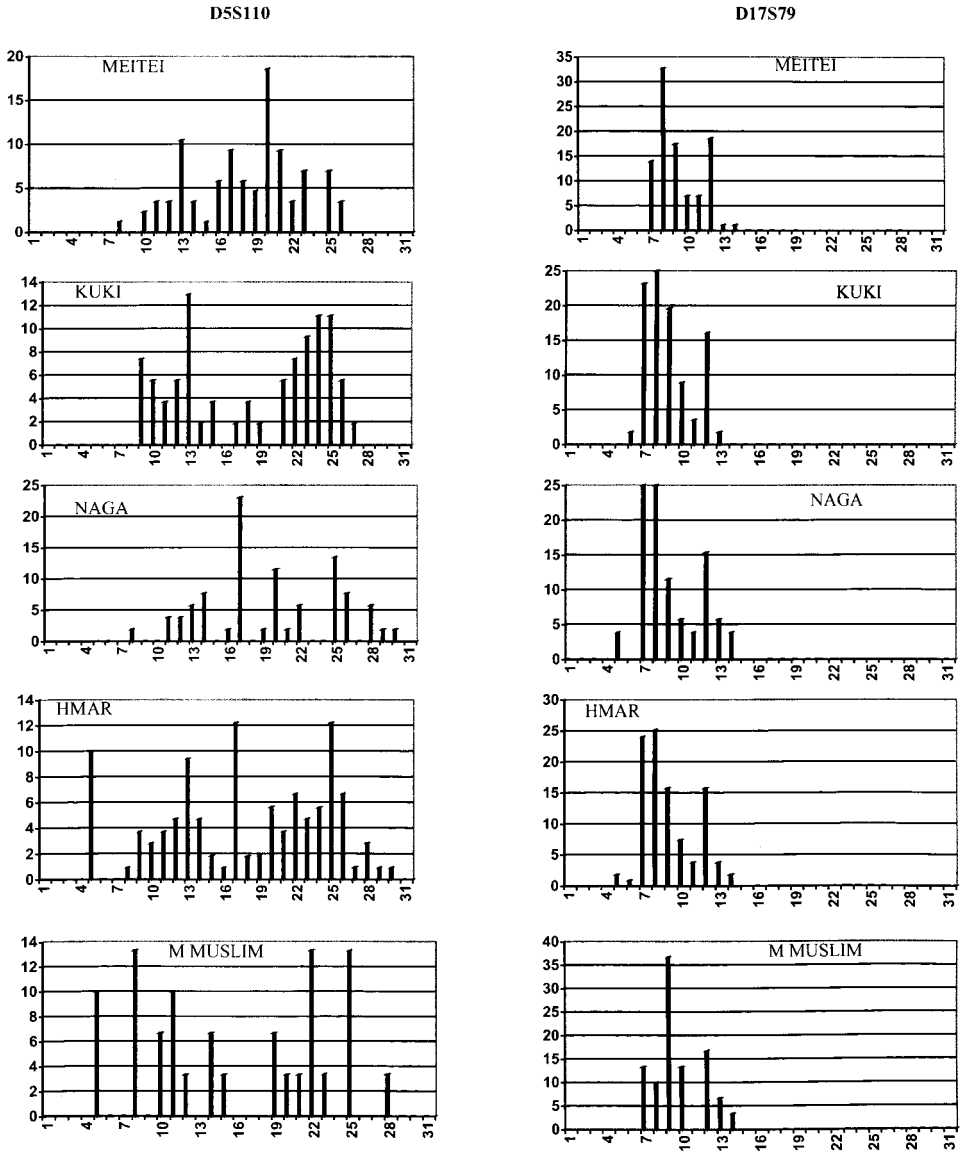

Figure 1. Continued.

Punjabi and Sikh populations of India was reflected in the 100% bootstrap value. The Manipuri Muslim formed the third complete separate branch. The pattern was similar in case of the STRs. Once again the northeastern Indian populations were found separated from the other Indian (Caucasoid populations) in 97% of the bootstraps. The Meitei was found separated from the combined branch of the northeastern Indian tribal populations (Naga, Kuki, Hmar) in 96% of the bootstraps (85% in the case of RFLP markers). The STRs are more effective in the detection of variation between the Mongoloid and the Caucasoid population samples of the study since they were separated in 100% of bootstraps. The STRs also show the complete separation of the Manipuri Muslim with other populations.

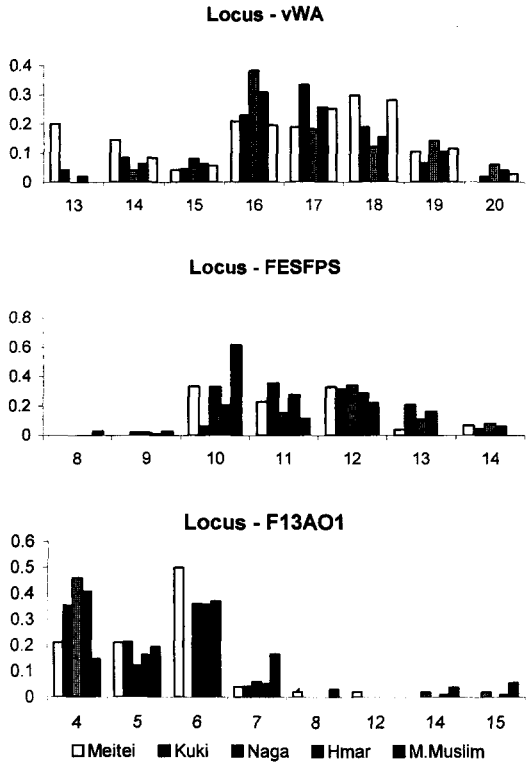

Figure 2. Allelic frequencies at three STR loci in five Manipur populations.

4. Discussion

The study among the five geographically contiguous and culturally and linguistically diverse Manipur population groups show distinct allelic differences in both VNTR and STR for some specific groups, suggesting the influence of population structure variables. The occurrence of the frequent allele was in accordance with other reports (FBI Laboratory Division 1993, Sovinsky *et al.* 1996, Dutta and Kashyap 2001a,b). One possible explanation of such a distribution could be the evolutionary antiquity of alleles, where the most common alleles are usually the oldest, while the wider distribution of the low-frequency alleles reflects different rates and types of mutations in populations (Waterson and Guess 1977).

The analyses of the VNTR systems in the five populations have shown record of all the alleles in 6 out of the 20 population–locus combinations. At loci D5S110, significant effect of r was observed. The frequency of non-detectable allele r, however, did not affect the suitability of the other loci to be in HWE. However, one could argue that the excess homozygosity values could also lead to a substantial increase in the value of r yet provide satisfactory HWE values. As we have analysed the difference between observed and calculated homozygosity and the values were found to be within acceptable limits (data not shown), the cause for the deviation could be attributed to the presence of non-detectable frequency r. The independence of alleles across pairwise combination of the loci, test for association across all the loci and probability of match values document for the inclusion of these VNTR and STR loci for the purpose of human identification among Indian populations.

Table 1. Hardy–Weinberg equilibrium test for the VNTR loci and STR loci in Manipur population groups.

Locus		n =	Meitei 128	Kuki 75	Naga 78	Hmar 60	M.Muslim 68
VNTR							
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	0.02
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
STR							
HUMvWA:	HWE test		0.477	0.455	0.749	0.205	0.984
	Likelihood ratio test		0.091	0.760	0.537	0.130	0.193
	Exact test		0.131	0.809	0.596	0.630	0.209
FESFPS:	HWE test		0.336	0.814	0.316	0.696	0.949
	Likelihood ratio test		0.329	0.308	0.586	0.222	0.182
	Exact test		0.323	0.408	0.649	0.283	0.214
F13AO1:	HWE test		0.430	0.209	0.831	0.280	0.834
	Likelihood ratio test		0.897	0.161	0.954	0.530	0.543
	Exact test		0.976	0.288	0.956	0.195	0.383

r = null allele frequency.
* Value with r.

Table 2. Average heterozygosity and G_{st} values for VNTR and STR loci in the northeast populations of India.

Locus	Meitei	Kuki	Naga	Hmar	M. Muslim	G_{st}
VNTR						
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
STR						
vWA	0.664	0.707	0.643	0.683	0.779	0.046
FESFPS	0.729	0.733	0.737	0.779	0.566	0.067
F13AO1	0.807	0.795	0.781	0.809	0.807	0.017
Average	0.733	0.745	0.720	0.757	0.717	0.043

Both the VNTR and STR systems have shown higher G_{st} values in comparison with the classical markers (Papiha 1996). Although we have not analysed the data for population sub-structuring using the existing models of mutation, it could be said that the effect of population sub-structuring is less pronounced at the VNTR loci in comparison with the protein or blood group markers. This is different from the observation of Chakraborty *et al.* (1996), which showed that the VNTR and protein

Table 3. Genetic distance (D_A) based on VNTR (lower diagonal) and STR loci (upper diagonal) among the Manipur populations.

Population	Meitei	Kuki	Naga	Hmar	M. Muslim
Meitei	–	0.0310	0.0275	0.0215	0.0377
Kuki	0.179	–	0.0306	0.0075	0.0698
Naga	0.160	0.234	–	0.0095	0.0502
Hmar	0.107	0.062	0.072	–	0.0514
M. Muslim	0.323	0.270	0.282	0.208	–

loci have almost the same G_{st} values and hence have some effect of population substructuring. The higher G_{st} values of the microsatellite could, however, be attributed to the lesser susceptibility of these loci towards mutations. Among the STR markers, a difference in the G_{st} values across the populations were observed (0.046 at vWA, 0.0067 at FESFPS and 0.017 at F13AO1). We are unsure of the reason for this variation, which could be due to factors such as inbreeding, natural selection, etc., operating at the loci.

The distance matrices show large values for the Manipuri Muslim group as compared with others. The clustering pattern depicts separation of populations, which corresponds to their socio-cultural and ethnic differences and not with their geographic proximity. Manipur Muslims are distinct from other populations and stand out separately from the rest. Although the type and number of genetic systems used are important in detecting the genetic relationship, the results do not support the findings of Mukherjee *et al.* (1999). They studied four STR loci in eight diverse populations from West Bengal (three), Orissa (three) and Uttar Pradesh (two) where they found greater geographic influence on genetic similarity than socio-cultural proximity. Comparison with other Indian populations gives much more clarity to the clustering pattern. For example, Garo, a Mongoloid population from West Bengal gets clustered with the northeast Mongoloid groups. And the northeast Indian populations form separate cluster from the other Indian populations. This observation further supports the findings based on a classical gene frequency study (Roychoudhury 1992) that Mongoloid populations show genetic similarity irrespective of their geographical, socio-cultural and linguistic backgrounds. These findings suggest the importance of population structure variables in studying the genetic similarity among the Indian populations. Possibly, owing to their geographical barriers, socio-cultural and ethnic differences have remained isolated from the other Indian populations.

One of the observations of the study is that the clustering pattern of the five populations based on VNTR, STR and the combined DNA markers (STR and VNTR) agree in general. However, the clustering pattern obtained from combined DNA markers are more in conformity with the pattern obtained by STR loci rather than with VNTR loci. If these findings are confirmed by other studies, this possibly suggests the preference of STR loci over VNTR for investigating the genetic affinities among the population groups. This might also suggest the possibility of significant association (e.g. linkage, etc.) between VNTR and STR loci. However, we are not aware of such studies in this regard. But such association between VNTR and STR loci is expected to reveal further insights into the nature of relationship and its importance in the genetic study of populations.

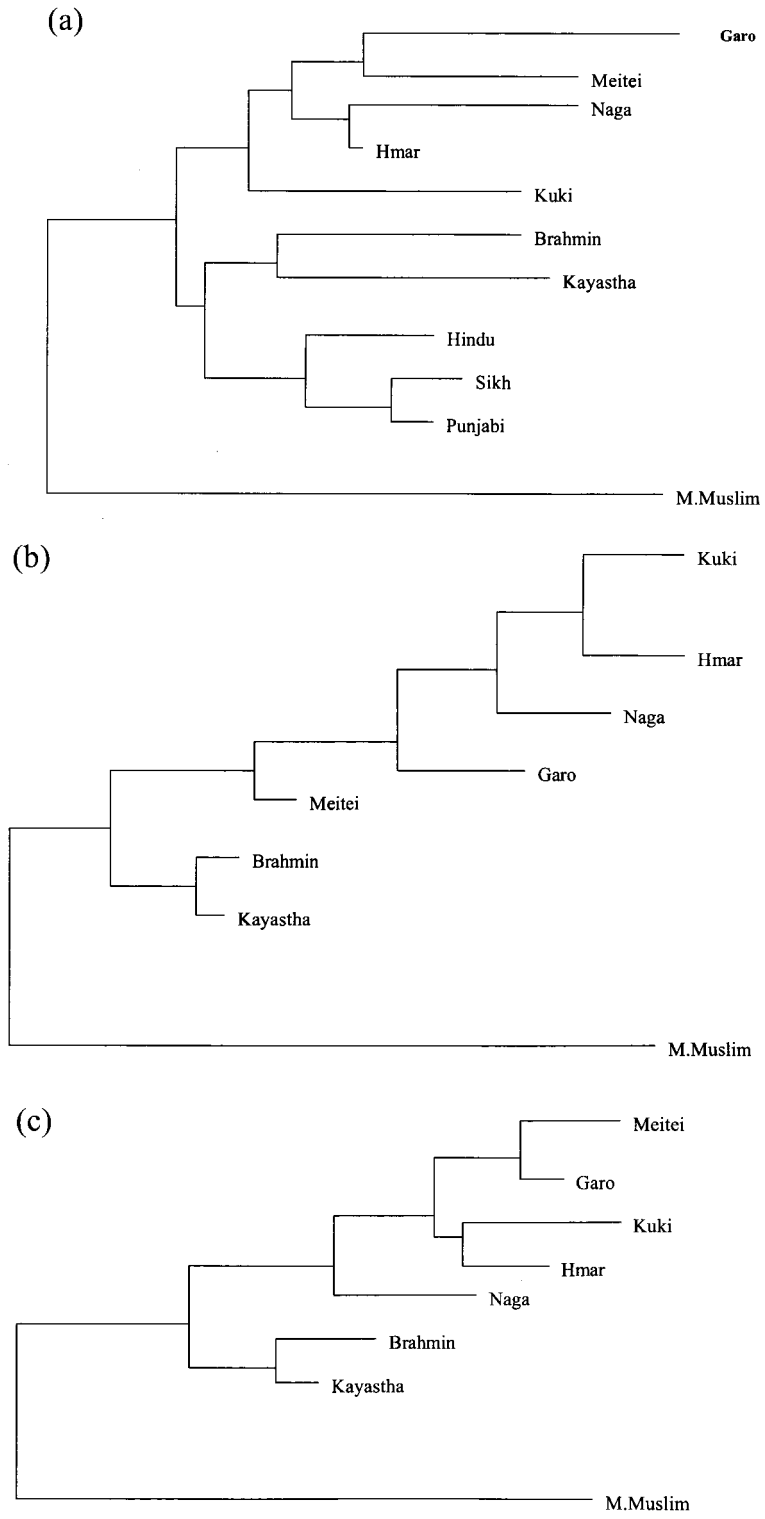

Figure 3. Neighbour-joining trees of clustering of five Manipur population samples based on (a) the four VNTR loci, (b) the three STR loci and (c) the seven DNA (VNTR + STR) markers.

Acknowledgements

We thank all the participants for their willingness and cooperation for providing their blood samples. The first author acknowledges the Bureau of Police Research & Development for the fellowship to him. The technical assistance of the staff of DNA typing unit of CFSL, Kolkata has been a great support in communicating this paper.

References

BALZAS, L., and BAIRD, M., 1993, Comparison of VNTR DNA profile frequencies in a large number of populations. *Procedings of the 45th Annual Meeting of the American Academy of Forensic Science*, West Conshohocken, ASTM, PA, USA, pp. 62–63.

BASSAM, A., 1991, Fast and sensitive silver staining of DNA in polyacrylamide gels. *Annals of Biochemistry*, **196,** 80–83.

BHASIN, M. K., WALTER, H., and DANKER-HOPFE, H., 1994, *People of India: Investigation of Biological variability in Ecological, Ethnic, Economic and Linguistic Groups* (Delhi: Kamla-Raj Enterprises).

BROWN, A. D. H., FELDMAN, M. W., and NEVO, E., 1980, Multilocus structure of natural populations of *Horeum spontaneum*. *Genetics*, **96,** 523–536.

BUDOWLE, B., CHAKRABORTY, R., GUISTI, A. M., EISENBERG, A. J., and ALLEN, R. C., 1991a, Analysis of the VNTR locus D1S80 by PCR followed by high-resolution PAGE. *American Journal of Human Genetics*, **48,** 137–144.

BUDOWLE, B., GUISTI, A. M., and WAYE, J. S., 1991b, Fixed bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci for use in forensic comparisons. *American Journal of Human Genetics*, **48,** 841–855.

CHAKRABORTY, R., 1990, Genetic profile of cosmopolitan populations. Effect of hidden subdivisions. *Anthropology Anzeiger*, **48,** 313–331.

CHAKRABORTY, R, 1993, A class of population genetic questions formulated as the generalized occupancy problem. *Genetics*, **134,** 953–958.

CHAKRABORTY, R., and JIN, L., 1992, Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. *Human Genetics*, **88,** 267–272.

CHAKRABORTY, R., JIN, L., and SHONG, Y., 1991, On allele frequency compilation from DNA typing data. *International Journal of Legal Medicine*, **106,** 103–106.

CHAKRABORTY, R., ZHONG, Y., JIN, L., and BUDWOLE, B., 1994, Nondetectability of restriction fragments and independence of DNA fragments size within and between loci in RFLP typing of DNA. *American Journal of Human Genetics*, **55,** 391–401.

CHAKRABORTY, R., STIVENS, D. N., and ZHONG, L., 1996, Estimation of mutation rates from parentage exclusion data. Application to STR and VNTR. *Mutation Research*, **54,** 41–48.

DEKA, R., SHRIVER, M. D., YU, L. M., FERREL, R. E., and 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., and KASHYAP, V. K., 2001a, Genetic variation observed in four VNTR loci—D1S7, D4S139, D5S110 and D17S79 in three population groups of eastern India. *Journal of Genetics*, **80,** 23–30.

DUTTA, R., and KASHYAP, V. K., 2001b, Genetic variation observed at three tetrameric short tandem repeat loci—HUMTHO1, TPX and CSF1PO—in five ethnic population groups of north-eastern India. *American Journal of Human Biology*, **13,** 23–29.

EDWARDS, A., HAMMOND, H., JIN, L., CHASKEY, C. T., and CHAKRABORTY, R. 1992, Genetic variation at five trimeric and terameric repeat loci in four human population groups. *Genomics*, **12,** 241–253.

GUO, S. W., and THOMPSON, E. A., 1992, Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48,** 361–372.

HAMMOND, H. A., JIN, L., ZHONG, Y., CASKEY, C. T., and CHAKRABORTY, R., 1994, Evaluation of thirteen STR loci for use in personal identification application. *American Journal of Human Genetics*, **55,** 175–194.

JEFFREYS, A. J., WILSON, V., and THEIN, S. L., 1985, Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314,** 67–74.

JIN, L., and CHAKRABORTY, R., 1991, Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74,** 274–285.

JIN, L., and CHAKRABORTY, R., 1995, Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity*, **74,** 274–283.

KARLIN, S., CAMERON, E. C., and WILLIAMS, P. T., 1981, Sibling and parent–offspring correlation estimation with variable family size. *Proceedings of the National Academy of Sciences*, **78,** 2664–2668.

KARUTHA PANDIAN, S., KUMAR, S., DHARMALINGAM, K., and 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., and DAMODARAN, C., 1995b, Allele frequency distribution for the variable number tandem repeat locus D1OS28 in Tamil Nadu (south India) population. *Electrophoresis*, **16,** 1689–1692.

FBI (FEDERAL BUREAU OF INVESTIGATION) LABORATORY DIVISION, 1993, VNTR population data: a worldwide study. Forensic Science Research Center, FBI Academy, Quantico, VA.

MAJUMDER, P. P., 1998, People of India: biological diversity and affinities. *Evolutionary Anthropology*, **6,** 100–110.

MALHOTRA, K. C., and VASULU, T. S., 1993, Structure of human populations in India. In *Human Population Genetics: a Centennial Tribute to J. B. S. Haldane*, edited by P. P. Majumder (New York: Plenum), pp. 207–233.

MANIATIS, T., FRITSCH, E. P., and SAMBROOK, J., 1982, *A Laboratory Manual* (New York: Cold Spring Harbor Laboratory Publication), pp. 458–459.

MASTANA, S. S., 1999, Genetic analysis of the D1S80 locus in five north Indian populations. *Annals of Human Biology*, **26,** 405–411.

MUKHERJEE N, MAJUMDER, P. P, ROY, M., DEY, B., CHAKRABORTY, M., and BANERJEE, S., 1999, Genomic variation at four short tandem repeat loci in eight population groups of India. *Human Biology*, **71,** 439–446.

MUKHERJEE, N., MITRA, M., CHAKRABORTY, M., and MAJUMDER, P. P., 2000, Congruence of genomic and ethnolinguistic affinities among five tribal populations of Madhya Pradesh (India). *Journal of Genetics*, **79,** 41–46.

NEI, M., and ROYCHOUDHURY, A. K., 1974, Sampling variances in heterozygosity and genetic distance. *Genetics*, **76,** 379–390.

NEI, M., 1973, Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70,** 3321–3323.

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

NEI, M., TAJIMA, F., and TATENO, Y., 1983, Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *Journal for Molecular Evolution*, **19,** 153–170.

PAPIHA, S. S., CALDERON, R., and SERTEDAKI, A.. 1998, Study of three hypervariable DNA loci in three Euproean populations. *Annals of Human Biology*, **25,** 29–41.

PAPIHA, S. S., MASTANA, S. S., PURANDARE, C. A., JAYASEKAHARA, R., and CHAKRABORTY, R., 1996, Population genetic study of three VNTR loci (D2S44, D7S22 and D12S11) in five ethnically defined populations of the Indian subcontinent. *Human Biology*, **68,** 819–835.

PAPIHA, S. S., 1996, Genetic variation in India. *Human Biology*, **68,** 607–628.

PUERS, C., LINS, A. M., SPRECHER, C. J., BRINKMANN, B., and SCHUMM, J. W., 1993, Analysis of polymorphic short tandem repeat loci using well-characterized allelic ladders. *Proceedings of the 4th International Symposium of Human Identification*, Puers Promega Corporation, Madison, WI.

REDDY, B. M., SUN, G., RODRIGUEZLUIS, J., CRAWFORD, M. C. H., HEMAM, N. S., and DEKA, R., 2001, Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Human Biology*, **73,** 175–190.

ROYCHOUDHURY, A. K., 1983, Genetic polymorphism in human populations of India. In *Peoples of India: Some Genetical Aspects*, edited by G. V. Satyavathi (New Delhi: Indian Council of Medical Research), pp. 1–29.

ROYCHOUDHURY, A. K., 1992, Genetic relationships of the populations in eastern India. *Annals of Human Biology*, **19,** 489–501.

SAHA, N., and TAY, J. S. H. 1990, Genetic studies among Nagas and Hmars of eastern India. *American Journal of Physical Anthropology*, **82,** 101–112.

SAITOU, N., and NEI, M., 1987, The neighbor joining method. A new method for reconstructing phyllogenetic tress. *Molecular Biology and Evolution*, **4,** 406–425.

SOVINSKY, S. M., BAIRD, L. S., BUDOWLE, B., CARUSO, J. F., CHEEMA, D. P. S., DUNCAN, G. T., HAMBY, P. P., MASIBAY, A. S., SHARMA, V. K., and TAHIR, M. A., 1996, Development of deoxiribonucleic acid (DNA) RFLP database for Punjabis in East Punjab, India. *Forensic Science International*, **79,** 187–198.

SINGH, K. S., 1998, *People of India*, National Series (IV–VI), Anthropological Survey of India (Delhi: Oxford University Press).

WATTERSON, G. A., and GUESS, H. A., 1977, Is the most frequent allele the oldest? *Theoretical Population Biology*, **11,** 141–160.

WEIR, B. S., 1992, Independence of VNTR alleles defined as fixed bins. *Genetics*, **130,** 873–887.

WAYE, J. S., PRESLEY, L., BUDWOLE, B., SHUTTER, G. G., and FOURNEY, R. M., 1989, A simple method for quantifying human genomic DNA in forensic specimen extracts. *Biotechniques*, **7,** 852–855.

Address for correspondence: Dr V. K. Kashyap, Central Forensic Science Laboratory, 30, Gorachand Road, Calcutta 700014, India. Email:vkk2k@hotmail.com

Zusammenfassung. *Hintergrund*: Abgesehen von einigen traditionellen Markern, welche an einigen zahlenmäßig kleinen, geografisch abgegrenzten Erhebungen von mongoliden Populationen im Nordosten Indiens studiert wurden, ist sehr wenig über ihre genetische Verschiedenheit auf molekularer Ebene bekannt.
Primäres Ziel: Die vorliegende Studie versucht zu untersuchen, wie die geeignetsten variable number tandem repeat (VNTR)- und short tandem repeat (STR)- Loci gemeinsam die Muster der genetischen Verwandtschaft zwischen 5 geografisch angrenzenden, aber sprachlich und soziokulturell unterschiedlichen, der mongoliden Gruppe angehörigen Populationen in Manipur (nordöstliche Region von Indien) herausgefunden werden können.
Material und Methoden: Es wurden Blutproben von nicht verwandten und zufällig ausgewählten Freiwilligen von 5 ethnischen Populationen (Meitei, Kuki, Naga, Hmar und Manipur- Moslems) aus 5 unterschiedlichen Landesteilen gesammelt. Es wurde die Allelenvariation von 4 Minisatelliten-Loci (D1S7, D4S139, D5S110 und D17S79) und 3 STR-Loci (vWA, FESFPS und F13AO1) untersucht.
Ergebnisse: Die durchschnittlichen Heterozygotenfrequenzen zwischen den 5 Gruppen für die Mikrosatelliten erstreckten sich zwischen 68 und 94%, währenddessen die Hypervariabilität der 3 STR-Loci zwischen 60 und 88% schwankte. In den Populationen waren alle untersuchten Loci hoch polymorph- mit nahezu keiner Abweichung vom Hardy-Weinberg-Equilibrium. Die Gendifferenzierung für die VNTR-Loci war niedriger und moderat ($G_{st} = 0.030$) im Vergleich zu den Mikrosatelliten ($G_{st} = 0.043$). Die neighbour-joining-Methode, welche auf beiden Typen der molekularen Marker basiert, enthüllt ein dichtes Cluster bei den Stammesgruppen der Kuki, Naga und Hmar, während die Manipur-Moslems in beiden Stammbäumen gesondert stehen. Die Muster des Clusters, welche von den kombinierten DNA-Marker-Loci erhalten wurden, sind bei den Mustern der STR-Loci dichter als bei den VNTR-Loci.
Schlußfolgerungen: Die Resultate dieser Studie untermauern, daß die Untersuchung sowohl der VNTR als auch der STR zur Bestimmung der genetischen Verwandtschaft zwischen den Populationen effektiver ist als die voneinander unabhängige Bestimmung der VNTR oder der STR. Außerdem werden die Ergebnisse durch die serologischen und elektrophoretischen Daten unterstützt.
Dennoch sind die Cluster-Muster, welche von den kombinierten DNA-Markern erhalten wurden, mit den Mustern der STR-Loci eher konform als mit den VNTR-Loci. Ungeachtet der sprachlichen, geografischen und kulturellen Barrieren zeigen die 4 Populationen mit Ausnahme der umgesiedelten Manipur-Moslems eine genetische Verwandtschaft.

Résumé. *Arrière-plan*: Hormis les marqueurs traditionnels étudiés lors d'enquêtes locales portant sur de petits effectifs, la diversité moléculaire du génome des populations mongoloïdes du nord-est de l'Inde est très peu connue.
Objectif premier: Cette étude cherche à évaluer dans quelle mesure les loci présentant des répétitions de tandem en nombre variable (VNTR) et les loci à répétitions de tandem courtes (STR) peuvent révéler ensemble les affinités génétiques de cinq populations mongoloïdes, géographiquement contiguës mais socio-culturellement diverses, de l'état de Manipur dans le nord-est de l'Inde.
Sujets et méthodes: Des échantillons sanguins ont été prélevés chez des volontaires pris au hasard et non apparentés entre eux, de cinq groupes ethniques (Meitei, Kuki, Naga, Hmar et musulmans Manipuri) de différentes parties de l'état. On a étudié la variation allélique de quatre loci minisatellites (D1S7, D4S139, D5S110 and D17S79) et de trois loci STR (vWA, FESFPS and F13AO1)
Résultats: Les valeurs moyennes d'hétérozygosité dans les cinq groupes pour le minisatellite, s'étendent de 68% à 94% et de 60% à 88% pour les trois loci hypervariables. Tous les loci étudiés sont hautement polymorphes et ne présentent presque pas d'écart aux proportions de Hardy-Weinberg. La différenciation génétique pour les loci VNTR est plus basse et modérée ($G_{st} = 0.030$) en comparaison avec les microsatellites ($G_{st} = 0.043$). La méthode d'agrégation de proche en proche appliquée aux deux types de marqueurs moléculaires révèle l'existence d'un étroit cluster rassemblant les groupes tribaux Kuki, Naga et Hmar, alors que les deux analyses maintiennent les musulmans de Manipur à part. Le mode d'agrégation obtenu à partir des marqueurs de DNA combinés est beaucoup plus proche du mode que décrit les loci STR que de celui des loci VNTR.
Conclusions: Les résultats renforcent l'idée qu'il est beaucoup plus efficace d'utiliser conjointement les loci VNTR et STR pour détecter l'affinité génétique de populations régionales, que de les utiliser séparément. Par ailleurs, ils confirment ceux qui ont été obtenus à partir de données sérologiques et électrophorétiques. Cependant, le mode d'agrégation obtenu à partir des marqueurs d'ADN est plus conforme à celui que fournissent les marqueurs STR que VNTR. En dépit de barrières géographiques, linguistiques et culturelles, quatre des populations étudiées manifestent des affinités qui les distinguent de la cinquième, population musulmane immigrée à Manipur.

Resumen. *Antecedentes*: Además de los marcadores tradicionales estudiados en algunas investigaciones realizadas en poblaciones mongoloides del noreste de la India, limitadas tanto numérica como geográficamente, se sabe muy poco de la diversidad genética de dichas poblaciones a nivel molecular.
Objetivo principal: Este estudio pretende analizar cómo el uso conjunto de los loci para el numero variable de repeticiones en tándem (VNTR) y para repeticiones cortas en tándem (STR) puede mejorar la detección de patrones de afinidad genética entre cinco poblaciones con afinidad mongoloide de Manipur, situadas en

regiones del noreste de la India, geográficamente próximas pero lingüistica y socioculturalmente diversas.
Muestra y métodos: Se recogieron muestras sanguíneas de voluntarios seleccionados aleatoriamente y no emparentados de 5 grupos étnicos (Meitei, Kuki, Naga, Hmar y Musulmanes Manipuri) de diferentes zonas del estado. Se estudió la variación alélica de 4 loci minisatélites (D1S7, D4S139, D5S110 y D17S79) y 3 loci STR (vWA, FESFPS y F13AO1).
Resultados: Los valores de heterocigosidad media entre los 5 grupos para los minisatélites oscilaban entre el 68% y el 94%, mientras que los tres loci STR hipervariables estaban entre el 60% y el 88%. En las 5 poblaciones, todos los loci estudiados fueron altamente polimórficos, sin desviarse apenas del equilibrio Hardy-Weinberg. La diferenciación génica de los loci VNTR resultó ser menor y moderada (G_{st} = 0.030) en comparación con la de los microsatélites (G_{st} = 0.043). El método de agrupamiento del vecino más próximo ("neighbour-joining") basado en ambos tipos de marcadores moleculares revela un estrecho agrupamiento entre los grupos tribales de Kuki, Naga y Hmar, mientras que los musulmanes de Manipur se mantienen separados en ambos árboles. El patrón de agrupamiento obtenido a partir de la combinación de loci marcadores del DNA se corresponde en mayor medida con el patrón de los loci STR que con el obtenido a partir de loci VNTR.
Conclusiones: Los resultados apoyan que el uso conjunto de los loci VNTR y STR para la detección de afinidad genética regional entre poblaciones es más efectivo que el empleo de VNTR o STR de forma independiente, y también confirma los resultados obtenidos a partir de datos serológicos y electroforéticos. Sin embargo, el patrón de agrupamiento obtenido a partir de marcadores combinados de DNA se corresponde más con el patrón obtenido mediante loci STR que con loci VNTR. A pesar de las barreras lingüísticas, geográficas y culturales, existe afinidad genética entre 4 de las poblaciones estudiadas, con la excepción del grupo emigrante Musulmán de Manipur.