

RESEARCH ARTICLE

Deciphering diversity in populations of various linguistic and ethnic affiliations of different geographical regions of India: analysis based on 15 microsatellite markers

V. K. KASHYAP^{1*}, RICHA ASHMA¹, SONALI GAIKWAD¹,
B. N. SARKAR² and R. TRIVEDI¹

¹DNA Typing Unit, Central Forensic Science Laboratory, 30 Gorachand Road, Kolkata 700 014, India

²Anthropological Survey of India, 27 Jawaharlal Nehru Road, Kolkata 700 016, India

Abstract

The extent of genetic polymorphism at fifteen autosomal microsatellite markers in 54 ethnically, linguistically and geographically diverse human populations of India was studied to decipher intrapopulation diversity. The parameters used to quantify intrapopulation diversity were average allele diversity, average heterozygosity, allele range (base pairs), and number of alleles. Multilocus genotype frequencies calculated for selected populations were utilized for testing conformity with the assumption of Hardy–Weinberg equilibrium. The exact test values, after Bonferroni correction, showed significant deviation amongst Gowda (vWA, Penta E); Dhangar, Satnami and Gounder (D8S1179); Hmar (FGA); Kuki and Balti (vWA) groups. Relatively low number of alleles and allelic diversity (base-pairs size) had been observed in populations of central India as compared with southern and northern regions of the country. The communities of Indo-Caucasoid ethnic origin and Indo-European linguistic family (Kshatriya of Uttar Pradesh) showed highest allelic diversity, as well as rare alleles, not reported in any other Indian populations. Analysis based on average heterozygosity was also found to be lowest among the populations of central India (0.729) and highest among the populations from north (0.777) and west (0.784) regions of the country, having Indo-Caucasoid ethnic origin and Austro-Asiatic linguistic affiliation. The maximum power of discrimination (85%–89%) had been observed at loci FGA, Penta E, D18S51 and D21S11, suggested high intrapopulation diversity in India. Genetic diversity revealed by STR markers was consistent with the known demographic histories of populations. Thus, the present study clearly demonstrated that the intrapopulation diversity is not only present at the national level, but also within smaller geographical regions of the country. This is the first attempt to understand the extent of diversity within populations of India at such a large scale at genomic level.

[Kashyap V. K., Ashma R., Gaikwad S., Sarkar B. N. and Trivedi R. 2004 Deciphering diversity in populations of various linguistic and ethnic affiliations of different geographical regions of India: analysis based on 15 microsatellite markers. *J. Genet.* **83**, 49–63]

Introduction

Intrapopulation genetic diversity in India has always remained a point of interest for anthropologists and human geneticists, because of the rich cultural heritage (linguistic, ethnic, etc.), history of different waves of migration and impact of deep-seated caste system (or practice of strict

endogamy/marriage rule), which potentially make every population a unique, isolated and important group. Many systematic (mutation, selection, migration and admixture) and non-systematic evolutionary factors (such as size, density, distribution of population and marriage rules) affect population structure, diversity (important source of variation) and equilibrium. Effect of these systematic and non-systematic forces varies with geographical region and the population groups residing in the area. If a chromosome carrying a disease allele (certain variants of beta-thalas-

*For correspondence. E-mail: vkk2k@hotmail.com, dtubprd@giasc101.vsnl.net.in.

Keywords. microsatellites; Indian populations; heterozygosity; power of discrimination; Hardy–Weinberg equilibrium.

semia) is introduced into a population (as a result of migration or mutation), its transmission rate varies with ethnic group (level of admixture with the migrated group/ endogamy/ consanguinity). The high incidence of some genetic disease in certain ethnic groups may also influence the resistance and susceptibility towards various infectious diseases. Therefore it is important to understand genetic diversity in the Indian population.

The precision of studying intrapopulation diversity has been revolutionized by the molecular characterization of large number of polymorphic microsatellite markers in the human genome. Unlike protein markers, microsatellites have a large number of codominant alleles due to high mutation and recombination rates (~1/1000) (Weber and Wong 1993), and can easily be analysed by automated genotyping procedures (Lin *et al.* 1996). These features make microsatellite loci ideal for studying variability in human populations. Also, they are proved valuable tools for different anthropological studies like testing the hypothesis of the evolution of anatomically modern human (Relethford and Jorde 1999). Earlier studies have reported their utility in reconstructing genealogies in ethnically well defined populations (Calafell *et al.* 1998) and analysing sub-structure of global populations (Eller 1999). Although a considerable amount of information on polymorphism at microsatellite loci in humans is now available (Sambrook *et al.* 1989; Waye *et al.* 1989; CFSL 2002), such studies in India are confined to limited groups (Deka *et al.* 1995, 1999; Mountain *et al.* 1995; Papiha *et al.* 1996; Mukherjee *et al.* 1999; Panneerchelvam *et al.* 2001; Reddy *et al.* 2001; Agrawal *et al.* 2002a,b,c; Das *et al.* 2002). STR loci have also been used extensively in human identification for forensic purposes (Dutta and Kashyap 2001; Chattopadhyay *et al.* 2001; Tahir *et al.* 2000).

Indian populations have received much attention from anthropologists, archaeologists, human geneticists and historians because of their enormous sociocultural, geographical, linguistic and biological diversity. Anthropological and archaeological records provide evidence that people from different ethnic groups, cultures and languages had migrated from different directions to India during the prehistoric era. The main corridors of this genetic influx were: (i) the Maracan coast and Baluchistan in the west and (ii) Southeast Asia and south China through India's north-eastern border (Fairservis 1971). These immigrants slowly amalgamated with the local populations of northwest region (Malhotra 1978; Papiha *et al.* 1996) and gradually changed the genetic profile of a major part of India. This contributed significantly to the present day gene pool of the subcontinent (Allchin and Allchin 1982; Bhasin *et al.* 1994; Gadgil *et al.* 1998; Singh 1998). The diverse populations in India can be broadly classified phenotypically into four ethnic classes: Australoid, Negrito, Mongoloid, and Caucasoid. The last ethnic group is spread the over entire country, with specific concentration in the northern

regions. Australoid group is mostly confined to western and southern states. The Negrito element is restricted to the Andaman Islands, and Siddis, the only Negroid population in India, a migrant group from Africa, reside in Karnataka, Gujarat and Andhra Pradesh. Periodic influx of various ethnic stocks also introduced diverse languages that blended with dialect of indigenous people. Impact of linguistic mingling is evident from four major groups of languages spoken by the Indians: Dravidian, Austro-Asiatic, Tibeto-Burman and Indo-European. Majority of mainland populations of southern India are Dravidian speakers whereas those in northern India are Indo-European speakers. Austro-Asiatic languages are used exclusively by tribal groups, e.g. Korkus, Mundas, Santhals, Khasis, Nicobarese, Oraon, etc. as they are believed to have migrated to the eastern region of the subcontinent in primordial time (Gadgil *et al.* 1998). Perhaps no other region in the world harbors such a rich array of ethnic and linguistic elements as India. The contemporary population of India is an agglomeration of thousands of groups, so it is of immense interest to quantify the genetic diversity in the country and to relate the observed patterns of the variation with cultural, linguistic and demographic histories of the selected groups.

With this background and rationale, the present study was carried out, across 15 STR loci over 3600 individuals belonging to 54 different communities. Interpretation of enormous genomic data generated for intrapopulation diversity study was handled by utilizing various bioinformatics tools. Communities were classified on the basis of their ethnic origin, culture, language, and geographical location. Diversity statistics such as average heterozygosity, allele diversity per locus, unique alleles per locus, power of discrimination, etc. were used to quantify the intrapopulation variation/diversity (Jin *et al.* 2000).

Materials and methods

Populations

Fiftyfour diverse Indian communities were selected for the present study based on prevalence of endogamy, population size and their anthropological significance (CFSL 2002). Populations were categorized on criteria of ethnic origins, linguistic affiliations and geographical regions. Ethnic classification (appendix 1) comprised of three defined categories: Indo-Caucasoid (27), Mongoloid (10) and Australoid (17). Geographical cluster included: north (15), west (4), east (17), south (14) and central (4) regions. Linguistic division consists of Dravidian (15), Indo-European (26), Tibeto-Burman (9) speaking populations, along with four tribal groups representing the Austro-Asiatic speakers. Thus, the communities selected for the present study represent a broad spectrum of ethnic, linguistic and geographical diversity of India.

STR markers

The selected STR markers are tetra- and penta-nucleotide (TPOX, D3S1358, FGA, D5S818, CSF1PO, D7S820, D8S1179, TH01, vWA, D13S317, D16S539, D18S51, D21S11, Penta D and Penta E) located on 13 different chromosomes (table 1). CSF1PO and D5S818 loci on chromosome 5, and Penta D and D21S11 located on chromosome 21 are unlinked, which makes them ideal for analysing intrapopulation genetic diversity. Out of a total of 54 populations, fortyone populations were analysed for both nine and 15 co-amplified microsatellite loci. Thirteen populations, of which majority (eight) were Tibeto-Burman speakers, four Indo-European and one Dravidian speaking population, were analysed for nine STR loci.

DNA isolation

One ml blood from each of 3647 unrelated consenting donors belonging to different groups was obtained by venipuncture in EDTA-coated vacutainers. Community, health status and family disease histories were recorded on blood donor cards of the DNA Typing Unit, CFSL, Kolkata. Genomic DNA was extracted from peripheral blood leukocytes using organic methods (Sambrook *et al.* 1989). Quality and quantity assessment of the isolated DNA was performed using 0.8% agarose yield gel (Waye *et al.* 1989).

Amplification and genotyping

Amplification was achieved with 1 to 10 ng of DNA in 25 µl reaction mixture containing 0.5 unit of *Taq* polymerase, fluorescent labeled specific primers and buffers, obtained for 15 co-amplified STR loci from Power Plex™16 System (Promega Corporation, Madison, USA), and nine co-amplified STR loci from AmpF1STR® Profiler Plus™

(Applied Biosystems, USA). Polymerase chain reaction was carried out in PE2400 PCR machine (Perkin Elmer Biosystems, Palo Alto, USA) following the manufacturer's protocols. Electrophoresis (denaturing polyacrylamide gel) of amplified product for genotyping was done using the ABI Prism™ 377 DNA sequencer.

Statistical analysis

Since the STR loci are autosomal co-dominant, allele frequencies of individual loci were calculated from the number of each genotype in the sample set by the method of gene count. Three tests were used for testing conformity with expected Hardy-Weinberg proportion of genotype frequencies, viz. homozygosity test (Nei and Roychoudhary 1974; Nei 1978; Chakraborty *et al.* 1998), log-likelihood ratio method (Weir 1992), and exact test for multiallelic loci (Guo and Thompson 1992). Bonferroni correction was applied to find out the significant departures from HW equilibrium (Hochberg 1988). Average heterozygosity (Nei 1973) and power of discrimination (Graber *et al.* 1983) at different microsatellites for each population were computed for estimating intrapopulation variations among selected groups across the 15 microsatellite loci.

Results and discussion

Microsatellites, being highly polymorphic and most suitable markers for population genetic studies (Destro-Bisoli *et al.* 2000) have been used in the present analysis for revealing intrapopulation diversity of the Indian populations. The details of allele frequencies at selected loci and their adherence with expectations of Hardy-Weinberg equilibrium of the studied populations reported in literature

Table 1. Characteristics of the fifteen autosomal microsatellite loci.

STR locus	Chromosomal location	Repeat motif (5'-3')	No. of alleles in ladder	Size range of allelic ladder (base pairs)
1 TPOX	2p23-2pter	AATG	8	262-290
2 D3S1358*	3p	TCTA Complex	9	115-147
3 FGA*	4q28	TTTC Complex (19)	29	322-444
4 D5S818*	5q23.3-32	AGAT	10	119-155
5 CSF1PO	5q33.3-34	AGAT	10	321-357
6 D7S820*	7q11.21-22	GATA	9	215-247
7 D8S1179	8q	TCTA Complex (19)	12	203-247
8 TH01	11p15.5	AATG(19)	9	156-195
9 vWA	12p12-pter	TCTA Complex (19)	15	123-171
10 D13S317*	13q22-q31	TATC	9	169-201
11 D16S539	16q24-qter	GATA	9	264-304
12 D18S51*	18q21.3	AGAA(19)	22	290-366
13 D21S11*	21q11-21q21	TCTA Complex(19)	26	203-259
14 PENTA E	15q	AAAGA	20	379-474
15 PENTA D	21q	AAAGA	15	376-449

*Common to both 15 and nine co-amplified systems.

(Dutta *et al.* 2000; Chattopadhyay *et al.* 2001; Kashyap *et al.* 2001; Ashma and Kashyap 2002a,b; Gaikwad and Kashyap 2002; Kashyap *et al.* 2002; Rajkumar and Kashyap 2002; Sahoo and Kashyap 2002a,b; Sarkar and Kashyap 2002; Sitalaximi *et al.* 2002; Tandon *et al.* 2002; Trivedi *et al.* 2002). The value of exact test, the most deducible among the three HWE test parameters (log likelihood ratio, exact test and homozygosity test), was calculated for 54 selected communities, out of which 20 significant departures from equilibrium (out of 810 locus-by-population combinations) after Bonferroni correction (appendix 2) were observed. Deviation among these 20 combinations, namely Gowda (vWA and Penta E), Dhangar, Satnami and Gounder (D8S1179); Hmar (FGA); Kuki, Balti (vWA), could be either due to excessive homozygosity or heterozygosity, attributed to practice of strict endogamy or admixture with the neighboring populations, respectively.

To understand the genetic structure and intrapopulation diversity in the selected communities group-wise average allele frequency was calculated (data not shown). Maximum variation was observed among the Indo-Caucasoid ethnic group. Allele 5 (389 bp) at Penta D was found to be present exclusively in Maratha, Chitpavan brahmin and Dhangar communities of Maharashtra. Out of 54 populations, unique allele 27.2 (217 bp) at D21S11 locus was observed in Dheria Gond of Madhya Pradesh (Australoid ethnic group). Buddhist (north region), Lepcha and Bhutia (east region) communities of Mongoloid ethnic origin and Tibeto-Burman linguistic family showed a distinct allele 23.2 (201 bp) at D21S11 and Penta E (451 bp and 479 bp) loci. Similarly, among the linguistic groups Dravidians had maximum number of unique alleles present at D5S818 (115 bp), D18S51 (290 bp), FGA (386 bp), D21S11 (217 bp) loci, followed by Indo-Europeans where alleles 14 (294 bp) and 14.2 (320 bp) at TPOX and D18S51 loci, respectively, were exclusive to the populations. The unique alleles 23.2 (201 bp) and 25 (479 bp) were recorded in Tibeto-Burmans at the highly polymorphic loci D21S11 and Penta E, respectively. Numbers of unique alleles were high in north and south Indian populations, followed by east, west and central regions of the country. In the northern region, 201 bp allele (D21S11), 203 bp allele (D8S1179), allele 171 bp (vWA) and alleles 320 bp and 370 bp (D18S51) were found, exclusively whereas in the south, 386 bp allele at FGA, 115 bp allele at D5S818 and allele 7 (286 bp) at D18S51 loci were exceptionally common. In the east zone, higher repeat number alleles were found to be rare, viz. allele 479 bp (Penta E), allele 159 bp (D5S818) and allele 251 bp (D7S820). Evolution of these unique alleles in certain ethnic groups might be responsible for affecting resistance or susceptibility towards infectious diseases. Although in some of the populations the frequencies of these unique alleles were below 5% level of significance, they provide a signature of population expansion in terms of intrapopulation diversity.

To further reveal diversity within 54 endogamous communities, allele range (number of observed alleles) and most frequent allele per locus were calculated using average allele frequency data (appendix 3). The observed number of alleles ranged from eight to 11 at eight loci (TPOX, D13S317, D16S539, CSF1PO, TH01, D7S820, D3S1358 and D5S818), followed by 12 to 14 alleles at three loci (vWA, D8S1179 and Penta D). More than 20 alleles were observed at FGA, Penta E, D18S51, and D21S11 loci indicating high level of polymorphism across these four loci amongst Indian populations. However, the frequencies of extreme (low base-pair size and high base-pair size) alleles across studied populations were statistically insignificant (at 5% level), indicating uniqueness and gradual evolution in the populations. With respect to distribution of alleles at the 15 loci, large variability existed, which clearly suggested significantly high intrapopulation diversity among the studied Indian populations. The age of an allele can be predicted from its frequencies under the infinite allele model and the stepwise mutation model, which suggests that the most common allele is the oldest (Watterson and Guess 1977). Allele 11 at seven loci with varying base-pair sizes [135 bp (D5S818), 185 bp (D13S317), 235 bp (D7S820), 282 bp (TPOX), 288 bp (D16S539), 409 bp (Penta E) and 417 bp (Penta D)] was found to be common (varying frequency) across the 54 communities, which suggested inter-population similarity in Indians. However, no common pattern of allele distribution had been observed at the remaining loci. This could be due to the effect of genetic influx and admixture due to the different socio-cultural practices prevalent in the populations.

The range of allele diversity (base pair size) at the 15 loci, a reliable measure of intragroup diversity, showed minimum diversity (as evident from the length of the arrow) in central region whereas north and west Indian communities had moderately high variation (figure 1a,b). In the four linguistic groups (figure 2), the patterns indicate least variation in Tibeto-Burman speakers and maximum diversity among Indo-European linguistic family. Highest range of allele variation among the ethnic groups was observed in Indo-Caucasoids followed by Australoids (figure 3). Maximum diversity was observed in north Indian populations having Caucasoid ethnic origin and Indo-European linguistic affiliation. Group-wise study, as well as history of peopling of India, clearly indicate that the northern region of the country had paved an easy route for immigration and turned the region into a melting pot of genes of various groups of immigrants. This has also introduced few variants of beta-thalassemia genetic disorder in some of the admixed Indian groups. A similar diversity pattern was observed when population-wise analysis was carried out. Allele diversity was observed to be highest (168 alleles) in Thakur caste group of Uttar Pradesh and lowest (105 alleles) in Reddy of Andhra Pradesh as compared to rest of the studied populations. The high degree

of diversity in Thakur (Kshatriya), a warrior group of Indo-Caucasoid ethnic origin could be due to extensive gene flow during different historical wars and invasions (Singh 1998). Conjugal relationships with local women of sub-

jugated territory (irrespective of caste and community) and fraternal polyandry during war periods in the past were quite frequent and had social acceptance because of the high social status of the Kshatriya group in the Hindu caste

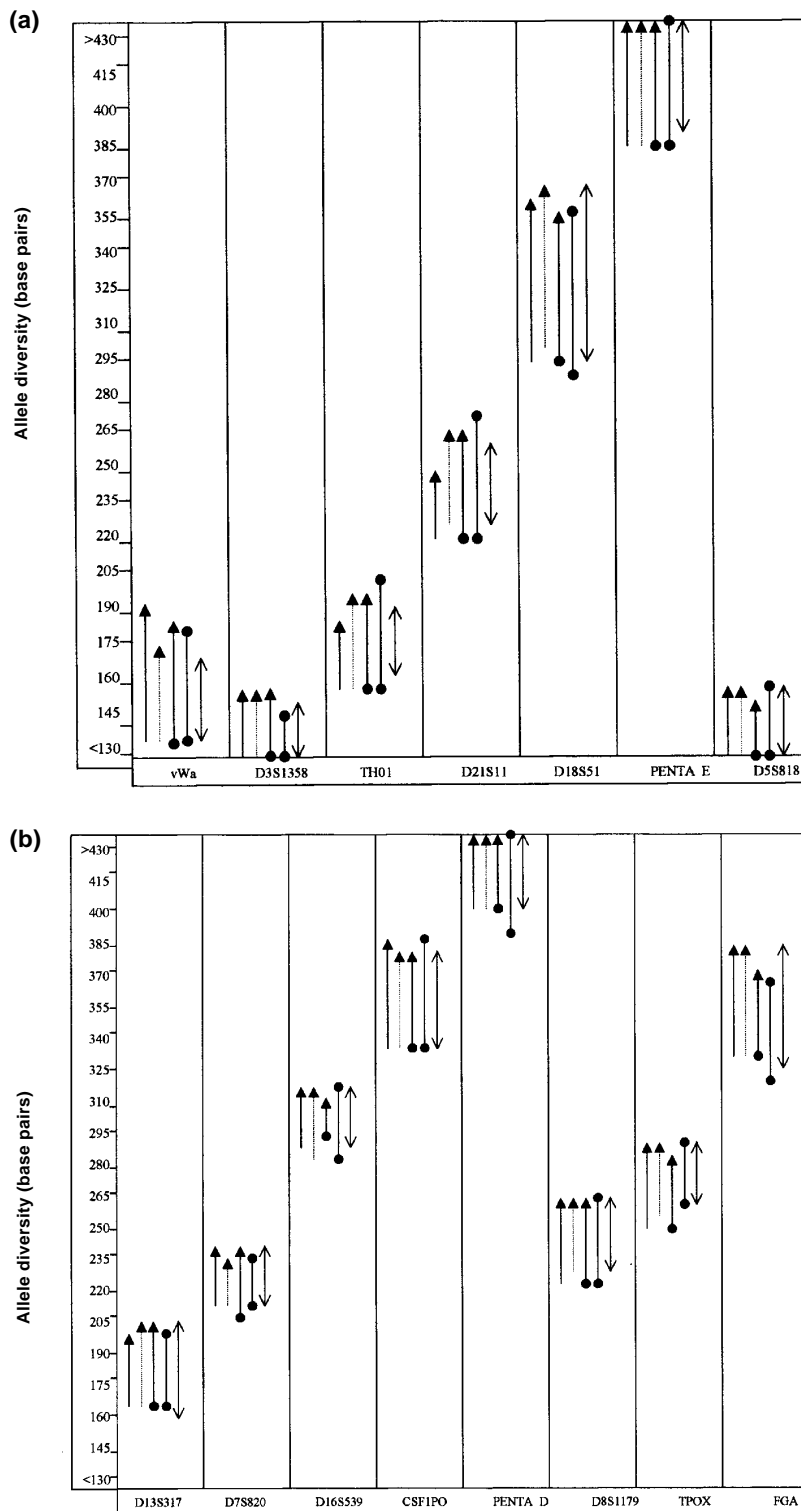


Figure 1. (a and b) Allele diversity (base pairs) at 15 STR loci in five major geographical regions of India. →, North; ····→, west; ●→, east; ●—●, south; ↔, central.

hierarchy. The study based upon microsatellites provides strong evidence in support of the history of origin of the caste group. The low diversity in Reddy, a Dravidian speaking population, is attributed to negligible gene flow in the group because cross-cousin marriages of both the types (with father's sister's daughter and with mother's brother's daughter, and also between maternal uncle and niece) widely practiced in the group. Low microsatellite diversity in the Reddy community was thus in accordance with socio-cultural norms.

The average heterozygosity (H_0) at 15 STR loci amongst 54 populations was further calculated to get a clear image of intrapopulation diversity (appendix 4). Value of H_0 was highly informative in understanding the antiquity and origin of population. Locus wise average heterozygosity of various ethnic and linguistic groups inhabiting diverse geographical regions of India is presented in table 2. The observed H_0 values ranged between 0.729 (central) and 0.784 (west) when five geographical regions were analysed, while it varied between 0.739 (Dravidian) and 0.797 (Austro-

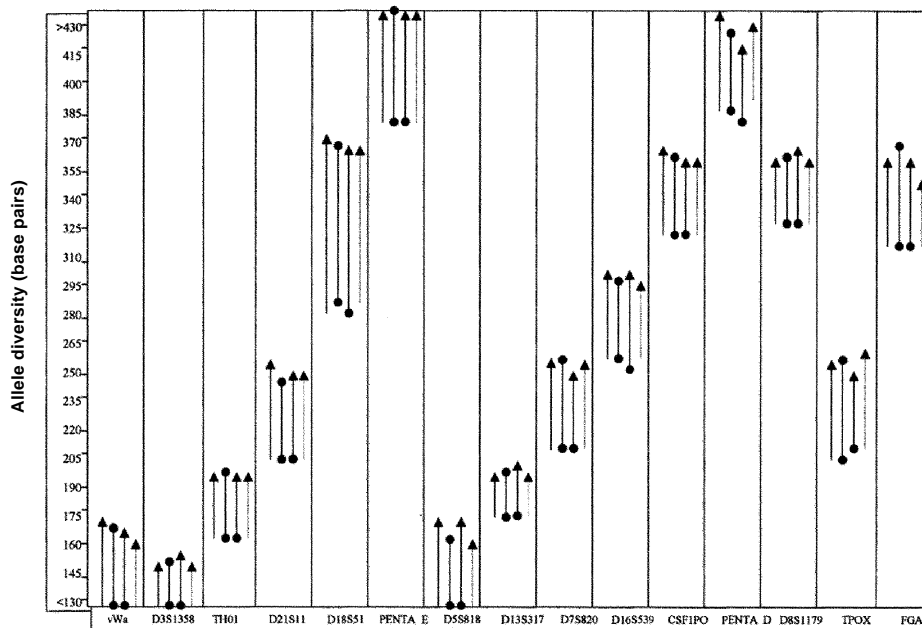


Figure 2. Allele diversity (base pairs) at 15 STR loci in four linguistic groups of India. →, Indo-Aryan; ●, Dravidian; ●→, Austro-Asiatic; ···→, Tibeto-Burman.

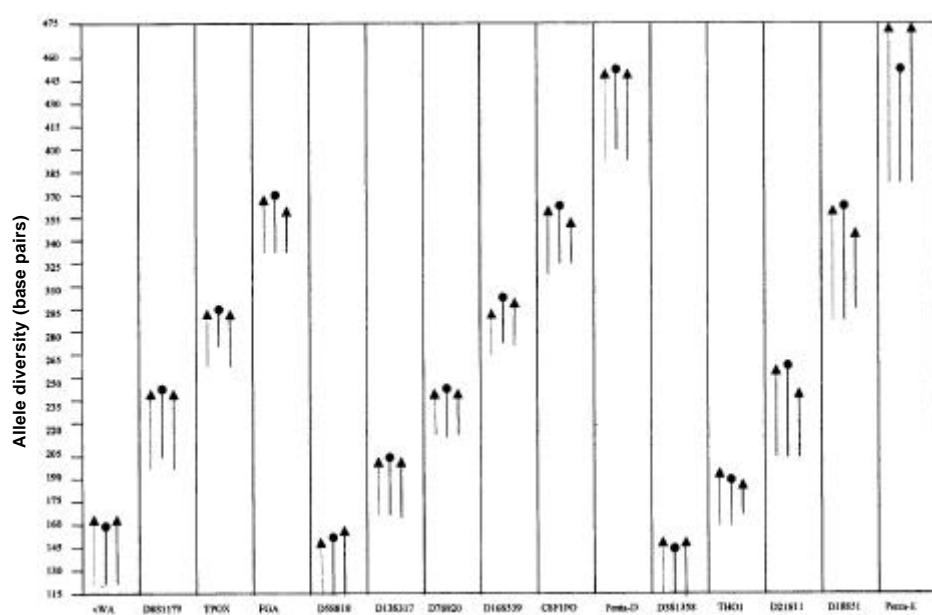


Figure 3. Allele diversity (base pairs) at 15 STR loci among the Australoid, Caucasoid and Mongoloid ethnic groups of India. ●, Australoid; →, Mongoloid; ···→, Caucasoid.

Table 2. Locuswise average heterozygosity amongst geographical region, ethnic and linguistic groups of India.

	vWA	D18S51	D8S1179	FGA	D5S818	D13S317	D7S820	D3S1358	D21S11	D16S539	TPOX	TH01	Penta D	CSF1PO	Penta E	Average
Geographical regions																
North	0.798	0.803	0.840	0.826	0.728	0.779	0.736	0.742	0.826	0.780	0.647	0.758	0.774	0.739	0.876	0.777
South	0.708	0.843	0.824	0.779	0.682	0.759	0.756	0.644	0.815	0.790	0.677	0.685	0.780	0.632	0.876	0.750
West	0.759	0.849	0.844	0.815	0.791	0.796	0.788	0.741	0.846	0.804	0.651	0.690	0.836	0.716	0.837	0.784
East	0.795	0.798	0.756	0.782	0.781	0.766	0.735	0.713	0.781	0.774	0.647	0.758	0.804	0.739	0.909	0.769
Central	0.774	0.777	0.795	0.783	0.712	0.756	0.688	0.690	0.812	0.741	0.538	0.716	0.735	0.651	0.766	0.729
Linguistic groups																
Indo-european	0.787	0.807	0.798	0.824	0.741	0.779	0.751	0.723	0.811	0.789	0.666	0.749	0.784	0.723	0.868	0.773
Dravidian	0.702	0.833	0.825	0.760	0.684	0.748	0.738	0.646	0.806	0.782	0.654	0.672	0.771	0.612	0.852	0.739
Tibeto-burman	0.787	0.792	0.788	0.717	0.778	0.775	0.711	0.741	0.783	0.744	0.600	0.683	0.780	0.752	0.882	0.754
Austro-asiatic	0.847	0.832	0.838	0.910	0.790	0.765	0.755	0.686	0.873	0.763	0.660	0.765	0.834	0.728	0.910	0.797
Ethnic groups																
Caucasoid	0.784	0.805	0.802	0.821	0.739	0.780	0.751	0.721	0.812	0.789	0.666	0.749	0.784	0.723	0.868	0.773
Mongoloid	0.796	0.801	0.778	0.735	0.781	0.773	0.715	0.744	0.783	0.744	0.601	0.680	0.780	0.739	0.882	0.755
Australoid	0.734	0.833	0.828	0.794	0.708	0.752	0.742	0.655	0.821	0.777	0.656	0.694	0.786	0.639	0.866	0.752

Asiatic) in the linguistic groups. The range of average heterozygosity found in the three ethnic groups varied between 0.750 (Mongoloid) and 0.773 (Indo-Caucasoid). Highest average heterozygosity was observed amongst populations of Caucasoid origin, occupying north and west regions. This further corroborates our observation of high allele diversity in north Indian population of Caucasoid ethnic origin and Indo-European linguistic affiliation and supports the history of population influx in the north region of India.

Power of discrimination (PD) is the average number of people one would have to survey before finding similar genotype pattern amongst randomly selected individuals. This is one of the most common but vital parameters used for individual identification and population genetic study. Among the studied populations, its value ranged from 0.540 (Lepcha) at TPOX locus to 0.976 (Kuki) at D18S51 locus (appendix 5). The tetra-nucleotide repeat loci, like TPOX (67.4%), CSF1PO (70.3%), TH01 (74.8%), D16S53 (75.1%), D7S820 (76.5%), D3S1358 (78.1%), D5S818 (79.1%), D13S317 (80.5%), vWA (81.7%) and D8S1179 (82.4%) exhibit relatively low power of discrimination, thus suggesting lowest polymorphism, and least degree of informativeness among the selected STR loci. More than 20 alleles were observed at Penta E (22), D18S51 (23), D21S11 (23) and FGA (23) loci, providing maximum discriminatory power to these four markers. The PD across the highly polymorphic loci Penta E (89.4%), D18S51 (84.6%), D21S11 (84.3%) and FGA (85.3%) clearly illustrates very high intrapopulation diversity in the studied populations.

The statistical parameters (number of alleles, allele size variation) analysed across the 15 STR loci to decipher intrapopulation diversity amongst studied populations clearly suggest that the populations of Indo-Caucasoid ethnic origin and Indo-European linguistic affiliation of the northern and western region of the country are the most diverse groups. This observation is consistent with the demographic and migration history of India. Allele diversity and average

heterozygosity also support the same pattern of genetic diversity in the Indian population. The Kshatriya caste of Uttar Pradesh (north India) having Indo-Caucasoid ethnic origin and Indo-European linguistic affiliation denotes maximum number of alleles, including some rare alleles that were not observed earlier in any of the Indian populations so far studied. Similarly, the Reddy community of Andhra Pradesh (south India), having Australoid ethnic origin and Dravidian linguistic affiliation, has minimum diversity in terms of number of alleles. Analysis based upon average heterozygosities further supports the results of high inter-population diversity in the Indian communities. The database and our findings in 54 diverse populations of India based upon allele distribution pattern across 15 STR loci demonstrate the potential use of these highly polymorphic loci in population genetic research as well as in determination of biological relatedness of individuals. Our results also suggest that the micro-evolutionary forces that influence the genotype have operated differently on various ethnic groups and in different geographical regions. The microsatellite data can be used for screening the association of STR alleles with predisposition to disease, which will be helpful in planning health strategies of the country.

Acknowledgements

This research was supported by a grant to Central Forensic Science Laboratory, Kolkata under Ninth Plan of Government of India. This study would not have been possible without the volunteer participation of the subjects. We sincerely thank the research fellows Ms S. Sahoo, Ms R. Rajkumar, Ms N. Sarkar, Ms Sitalaximi T., Mr S. Guha, Mr B. Maity, Mr M. Tandon, Drs R. Dutta and P. Chattopadhyay of CFSL, Kolkata for their active cooperation. We thank the two anonymous reviewers for their comments, which helped in improved presentation of the manuscript. This study is a part of a long-term project on 'Validation of highly polymorphic loci and data on alleles for Indian population', being carried out by the DNA Typing Laboratory of the Central Forensic Science Laboratory, Kolkata.

Appendix 1. Ethnic classification of fifty-four Indian populations.

Population (<i>n</i> = sample size)	Abbr. used	Region	Linguistic affiliation	Traditional occupation	Demographic history	Microsatellite studies (reference)
AUSTRALOID						
1 Sakanupakshollu (<i>n</i> = 30)	SP	Andhra Pradesh	Dravidian		Migrated from adjacent areas	Kashyap <i>et al.</i> 2002
2 Reddy (<i>n</i> = 30)	RC	Andhra Pradesh	Dravidian	Cultivator	Original inhabitant	Kashyap <i>et al.</i> 2002
3 Golla (<i>n</i> = 65)	GO	Andhra Pradesh	Dravidian	Pastoral	Original inhabitant	Reddy <i>et al.</i> 2001
4 Irula (<i>n</i> = 54)	IR	Tamil Nadu	Dravidian	Hunter-gatherer	Original inhabitant	Sitalaximi <i>et al.</i> 2002
5 Chakkliyar (<i>n</i> = 49)	CK	Tamil Nadu	Dravidian	Washerman	Migrated from Andhra Pradesh	Sitalaximi <i>et al.</i> 2002
6 Gounder (<i>n</i> = 56)	GU	Tamil Nadu	Dravidian	Agriculture	Original inhabitant	Sitalaximi <i>et al.</i> 2002
7 Vanniyar (<i>n</i> = 87)	VA	Tamil Nadu	Dravidian	Oil pressing community	Original inhabitant	Sitalaximi <i>et al.</i> 2002
8 Pallar (<i>n</i> = 33)	PA	Tamil Nadu	Dravidian	Agriculture	Original inhabitant	Sitalaximi <i>et al.</i> 2002
9 Pariyar (<i>n</i> = 21)	PR	Tamil Nadu	Dravidian	Agriculture	Original inhabitant	Sitalaximi <i>et al.</i> 2002
10 Kallar (<i>n</i> = 101)	TK	Tamil Nadu	Dravidian	Feudal chieftains	Original inhabitant	Sitalaximi <i>et al.</i> 2002
11 Gowda (<i>n</i> = 56)	GW	Karnataka	Dravidian	Agriculture	Migrated from neighboring areas	Rajkumar and Kashyap 2002
12 Karnataka Muslim (<i>n</i> = 65)	MU	Karnataka	Dravidian	Agriculture	Migrant	Rajkumar and Kashyap 2002
13 Juang (<i>n</i> = 50)	JU	Orissa	Austro- Asiatic	Shifting cultiva- tion	Original inhabitant	Sahoo and Kashyap 2002
14 Saora (<i>n</i> = 35)	SO	Orissa	Austro- Asiatic	Shifting cultiva- tion	Original inhabitant	Sahoo and Kashyap 2002
15 UP Kurmi (<i>n</i> = 45)	KM	Uttar Pradesh	Austro- Asiatic	Agriculture	Migrated from Bihar	Tandon <i>et al.</i> 2002
16 Bihar Kurmi (<i>n</i> = 50)	KU	Bihar	Austro- Asiatic	Agriculture	Original inhabitant	Ashma and Kashyap 2002
17 Dheria Gond (<i>n</i> = 35)	DG	Madhya Pradesh	Dravidian	Agriculture	Original inhabitant	Sarkar and Kashyap 2002
MONGOLOID						
18 Budhist (<i>n</i> = 156)	BD	Ladakh, Jammu & Kashmir	Tibeto- Burman	Priesthood	Migrated from Tibet	Trivedi <i>et al.</i> 2002
19 Argon (<i>n</i> = 51)	AR	Ladakh, Jammu & Kashmir	Tibeto- Burman	Trade and com- merce	Mixed population	Trivedi <i>et al.</i> 2002
20 Nepali (<i>n</i> = 110)	NE	Sikkim	Tibeto- Burman	Agriculture	Migrated from adjacent areas	Kashyap <i>et al.</i> 2002
21 Lepcha (<i>n</i> = 48)	LP	Sikkim	Tibeto- Burman	Hunter-gatherer, cultivator	Migrated from adjacent areas	Kashyap <i>et al.</i> 2002
22 Bhutiya (<i>n</i> = 75)	BH	Sikkim	Tibeto- Burman	Terrace cultiva- tion	Migrated from Tibet	Kashyap <i>et al.</i> 2002
23 Naga (<i>n</i> = 106)	NG	Manipur	Tibeto- Burman	Shifting cultiva- tion	Original inhabitant	Chattopadhyaya <i>et al.</i> 2001
24 Kuki (<i>n</i> = 105)	KK	Manipur	Tibeto- Burman	Shifting cultiva- tion	Original inhabitant	Chattopadhyaya <i>et al.</i> 2001
25 Hmar (<i>n</i> = 101)	HM	Manipur	Tibeto- Burman	Shifting cultiva- tion	Migrated from adjacent areas	Chattopadhyaya <i>et al.</i> 2001
26 Garo (<i>n</i> = 110)	GR	West Bengal	Tibeto- Burman	Shifting cultiva- tion	Migrated from Garo hill	Chattopadhyaya <i>et al.</i> 2001
27 Meitei (<i>n</i> = 105)	ME	West Bengal	Indo- European	Agriculture	Original inhabitant	Chattopadhyaya <i>et al.</i> 2001

contd . . .

Appendix 1. (contd) Ethnic classification of fifty-four Indian populations.

Population (<i>n</i> = sample size)	Abbr. used	Region	Linguistic affiliation	Traditional occupation	Demographic history	Microsatellite studies (reference)
CAUCASOID						
28 Drokpa (<i>n</i> = 38)	DR	Ladakh, Jammu & Kashmir	Indo-European	Trade and agriculture	–	Trivedi <i>et al.</i> 2002
29 Baltis (<i>n</i> = 67)	BA	Ladakh, Jammu & Kashmir	Indo-European	Trade and agriculture	Migrated from Balistan	Trivedi <i>et al.</i> 2002
30 Manipuri Muslim (<i>n</i> = 101)	MM	Manipur	Indo-European	Agriculture	Migrated from adjacent areas	Dutta <i>et al.</i> 2000
31 Bengali Brahmin (<i>n</i> = 110)	WB	West Bengal	Indo-European	Priesthood	Migrated from adjacent areas	Dutta <i>et al.</i> 2000
32 Bengali Kayastha (<i>n</i> = 103)	KW	West Bengal	Indo-European	Service	Original inhabitant	Dutta <i>et al.</i> 2000
33 Oriya Brahmin (<i>n</i> = 57)	OB	Orissa	Indo-European	Priesthood	Migrated from adjacent areas	Sahoo and Kashyap 2002
34 Khandayat (<i>n</i> = 62)	KH	Orissa	Indo-European	Agriculture	Original inhabitant	Sahoo and Kashyap 2002
35 Karan (<i>n</i> = 62)	KA	Orissa	Indo-European	Traditional scribes, agriculture	Original inhabitant	Sahoo and Kashyap 2002
36 Gope (<i>n</i> = 60)	GP	Orissa	Indo-European	Milkman	Original inhabitant	Sahoo and Kashyap 2002
37 Bihar Brahmin (<i>n</i> = 59)	BR	Bihar	Indo-European	Priesthood	Migrated from adjacent areas	Ashma and Kashyap 2002
38 Bhumihar (<i>n</i> = 65)	BU	Bihar	Indo-European	Agriculturist	Descendant from the Brahman	Ashma and Kashyap 2002
39 Rajput (<i>n</i> = 58)	RA	Bihar	Indo-European	Warrior, agriculture	Migrated from adjacent areas	Ashma and Kashyap 2002
40 Kayastha (<i>n</i> = 53)	BK	Bihar	Indo-European	Scribes, agriculture	Migrated from adjacent areas	Ashma and Kashyap 2002
41 Yadav (<i>n</i> = 44)	YA	Bihar	Indo-European	Animal husbandry	Original inhabitant	Ashma and Kashyap 2002
42 Baniya (<i>n</i> = 45)	BY	Bihar	Indo-European	Moneylenders and business	Migrated from adjacent areas	Ashma and Kashyap 2002
43 Iyengar Brahmin (<i>n</i> = 65)	IB	Karnataka	Dravidian	Priesthood	Migrated from adjacent areas	Rajkumar and Kashyap 2002
44 Lingayat (<i>n</i> = 98)	LI	Karnataka	Dravidian	Cultivation	Migrated from adjacent areas	Rajkumar and Kashyap 2002
45 Desasth Brahmin (<i>n</i> = 107)	DB	Maharashtra	Indo-European	Priesthood	Original inhabitants	Gaikwad and Kashyap 2002
46 Marathas (<i>n</i> = 102)	MA	Maharashtra	Indo-European	Warrior	Kshtriyas, agglomerated group	Gaikwad and Kashyap 2002
47 Dhangar (<i>n</i> = 80)	DH	Maharashtra	Indo-European	Shepherds	Migrated from adjacent areas	Gaikwad and Kashyap 2002
48 Chitpavan Brahmin (<i>n</i> = 70)	CB	Maharashtra	Indo-European	Priesthood	Migrated from adjacent areas	Gaikwad and Kashyap 2002
49 Agharia (<i>n</i> = 70)	AG	Madhya Pradesh	Indo-European	Agriculture	Migrated from Agra	Sarkar and Kashyap 2002
50 Satnami (<i>n</i> = 50)	SA	Madhya Pradesh	Indo-European	Traditional leather worker	Oldest resident of Chattisgarh	Sarkar and Kashyap 2002
51 Teli (<i>n</i> = 50)	TE	Madhya Pradesh	Indo-European	Oil pressing community	Migrated from adjacent areas	Sarkar and Kashyap 2002
52 Khatri (<i>n</i> = 47)	KT	Uttar Pradesh	Indo-European	Businessmen	Migrated from north-west India	Tandon <i>et al.</i> 2002
53 UP Jat (<i>n</i> = 48)	JA	Uttar Pradesh	Indo-European	Agriculture	Migrated from Punjab & Rajasthan	Tandon <i>et al.</i> 2002
54 UP Thakur (<i>n</i> = 48)	TH	Uttar Pradesh	Indo-European	Warrior, agriculture	Original inhabitant	Tandon <i>et al.</i> 2002

Appendix 2. Exact test of HWE at 15 STR loci in predominant Indian populations (*genotyped for nine co-amplified loci only).

Caste	vWa	D8S1179	D3S1358	D5S818	D13S317	D7S820	FGA	D21S11	TH01	D18S51	CSF1PO	D16S539	TPOX	Penta D	Penta E
NE	0.977	0.792	0.365	0.109	0.220	0.128	0.587	0.992	0.559	0.225	0.558	0.463	0.570	0.187	0.236
BH	0.376	0.120	0.349	0.949	0.719	0.967	0.678	0.318	0.259	0.125	0.646	0.189	0.762	0.236	0.236
LE	0.426	0.293	0.394	0.428	0.144	0.372	0.601	0.411	0.382	0.242	0.317	0.284	1.000	0.053	0.053
OB	0.441	0.470	0.192	0.649	0.267	0.186	0.174	0.767	0.248	0.652	0.135	0.601	0.881	0.155	0.830
KH	0.089	0.735	0.092	0.582	0.654	0.414	0.222	0.158	0.306	0.706	0.616	0.123	0.651	0.412	0.571
KA	0.429	0.221	0.326	0.244	0.897	0.250	0.212	0.427	0.243	0.294	0.657	0.095	0.951	0.366	0.821
GP	0.051	0.416	0.335	0.070	0.854	0.092	0.987	0.167	0.819	0.972	0.634	0.754	0.174	0.476	0.741
DB	0.093	0.493	0.944	0.501	0.261	0.122	0.820	0.561	0.330	0.810	0.762	0.827	0.470	0.598	0.210
MA	0.773	0.169	0.897	0.375	0.068	0.671	0.293	0.145	0.179	0.773	0.572	0.854	0.392	0.727	0.010
CB	0.305	0.241	0.660	0.272	0.065	0.219	0.660	0.116	0.400	0.299	0.951	0.329	0.111	0.182	0.591
DH	0.476	0.000	0.816	0.413	0.142	0.406	0.149	0.552	0.801	0.255	0.275	0.423	0.585	0.091	0.158
IB	0.427	0.419	0.114	0.219	0.308	0.579	0.390	0.627	0.888	0.089	0.173	0.965	0.984	0.879	0.885
GW	0.023	0.330	0.219	0.584	0.074	0.909	0.159	0.388	0.372	0.599	0.259	0.679	0.571	0.077	0.002
LI	0.574	0.053	0.557	0.991	0.324	0.149	0.193	0.157	0.132	0.688	0.454	0.681	0.512	0.387	0.740
MU	0.126	0.244	0.341	0.459	0.600	0.321	0.144	0.278	0.093	0.000	0.235	0.059	0.506	0.145	0.038
TK	0.465	0.940	0.166	0.337	0.989	0.494	0.343	0.064	0.151	0.632	0.776	0.211	0.054	0.484	0.345
VA	0.161	0.144	0.732	0.558	0.721	0.038	0.506	0.914	0.715	0.207	0.399	0.914	0.246	0.382	0.467
PA	0.099	0.333	0.974	0.677	0.708	0.539	0.219	0.916	0.360	0.113	0.910	0.401	0.339	0.860	0.647
PR	0.942	0.851	0.512	0.725	0.261	0.536	0.469	0.561	0.137	0.814	0.127	0.172	0.112	0.500	0.501
BR	0.126	0.342	0.069	0.506	0.432	0.728	0.376	0.379	0.201	0.176	0.633	0.418	0.370	0.674	0.659
BU	0.394	0.319	0.690	0.132	0.799	0.641	0.073	0.170	0.488	0.299	0.840	0.169	0.769	0.509	0.759
RA	0.769	0.839	0.814	0.363	0.840	0.608	0.702	0.650	0.288	0.827	0.630	0.104	0.580	0.687	0.499
BK	0.757	0.527	0.627	0.466	0.838	0.607	0.534	0.484	0.670	0.065	0.738	0.163	0.161	0.189	0.148
YA	0.489	0.727	0.586	0.479	0.737	0.427	0.454	0.319	0.152	0.441	0.838	0.753	0.331	0.823	0.471
KU	0.755	0.399	0.125	0.078	0.537	0.279	0.447	0.618	0.820	0.681	0.451	0.442	0.245	0.292	0.643
BY	0.300	0.229	0.224	0.239	0.651	0.388	0.367	0.501	0.190	0.299	0.197	0.052	0.425	0.527	0.107
KT	0.060	0.127	0.594	0.695	0.095	0.217	0.109	0.922	0.470	0.757	0.723	0.128	0.584	0.149	0.314
KM	0.062	0.044	0.250	0.840	0.182	0.148	0.363	0.126	0.597	0.144	0.308	0.065	0.363	0.768	0.088
JA	0.050	0.262	0.878	0.702	0.093	0.910	0.809	0.395	0.161	0.116	0.292	0.895	0.108	0.352	0.603
TH	0.043	0.541	0.071	0.859	0.561	0.432	0.383	0.214	0.503	0.077	0.073	0.618	0.621	0.159	0.203
AG	0.234	0.084	0.863	0.572	0.218	0.312	0.000	0.687	0.346	0.060	0.615	0.788	0.599	0.490	0.148
DG	0.372	0.916	0.436	0.308	0.218	0.335	0.679	0.242	0.346	0.307	0.656	0.618	0.644	0.655	0.338
SA	0.069	0.048	0.530	0.516	0.364	0.298	0.883	0.089	0.127	0.057	0.575	0.887	0.231	0.289	0.291
TE	0.094	0.106	0.303	0.085	0.255	0.274	0.120	0.125	0.394	0.315	0.207	0.378	0.392	0.634	0.170
GU	0.641	0.018	0.898	0.594	0.811	0.771	0.932	0.117	0.970	0.721	0.876	0.128	0.785	0.394	0.413
IR	0.243	0.198	0.565	0.300	0.907	0.239	0.000	0.214	0.213	0.316	0.052	0.396	0.209	0.942	0.049
CK	0.595	0.104	0.611	0.672	0.933	0.418	0.000	0.568	0.971	0.805	0.564	0.582	0.141	0.139	0.411
RC	0.219	0.718	0.203	0.159	0.466	0.917	0.069	0.598	0.516	0.306	0.063	0.162	0.235	0.086	0.921
SP	0.459	0.394	0.067	0.094	0.415	0.376	0.545	0.201	0.247	0.252	0.433	0.778	0.380	0.239	0.077
JU	0.459	0.394	0.067	0.094	0.226	0.195	0.437	0.201	0.247	0.252	0.634	0.514	0.553	0.217	0.077
SO	0.299	0.601	0.030	0.138	0.543	0.390	0.132	0.691	0.605	0.292	0.909	0.796	0.303	0.331	0.863
ME*	0.057	0.271	0.781	0.583	0.411	0.064	0.496	0.710	0.059	0.198	0.730	-	0.110	-	-
MM*	0.111	0.289	0.347	0.383	0.112	0.813	0.552	0.603	0.130	0.337	0.325	-	0.167	-	-
GR*	0.118	0.197	0.835	0.117	0.325	0.214	0.136	0.119	0.227	0.217	0.078	-	0.114	-	-
NG*	0.235	0.717	0.152	0.708	0.050	0.434	0.140	0.166	0.057	0.250	0.114	-	0.554	-	-
KK*	0.042	0.663	0.612	0.856	0.291	0.171	0.054	0.208	0.547	0.250	0.359	-	0.165	-	-
HM*	0.127	0.168	0.386	0.730	0.166	0.129	0.049	0.187	0.100	0.250	0.133	-	0.416	-	-
WB*	0.305	0.281	0.870	0.988	0.704	0.166	0.188	0.484	0.285	0.211	0.578	-	0.104	-	-
KW*	0.500	0.403	0.425	0.912	0.070	0.050	0.100	0.181	0.424	0.443	0.140	-	0.204	-	-
GO*	0.758	0.315	0.410	0.114	0.245	0.242	0.105	0.355	-	0.740	0.000	-	0.000	-	-
BD*	0.489	0.570	0.197	0.692	0.053	0.472	0.302	0.501	-	-	-	-	-	-	-
AR*	0.051	0.320	-	0.129	0.000	0.172	-	-	-	-	-	-	-	-	-
DR*	0.243	-	0.146	1.000	0.012	0.328	0.205	-	-	-	-	-	-	-	-
BA*	0.021	0.127	0.064	0.054	0.054	-	0.066	-	-	-	-	-	-	-	-

Note: The caste abbreviation can be referred from appendix 1.

Intrapopulation genomic diversity in India

Appendix 3. Average allele frequencies (%) at 15 microsatellite loci among predominant Indian populations.

Allele size (bp)	115	119	123	127	129	131	135	139	143	147	151	155	156	159	160	163	164	167	168	169	171	172	173	176	177	179	181	183	185	187	189	193	197	201		
L1*		0.8	0.3		0.2	1	13.8	11.2	21.2	26.4	16.2		7.3		1.4		0.2																			
L2																																				
L3																																				
L4	0.1	0.4	2.7	4.9		12.8	30.1	28.8	15.7	2.4	1.7	0.3		0.1																						
L5																				1.1			20.4		11.2		11.6		23.1		21.7	7.8	2.8	0.3		
L6																																				
L7																																				
L8																																				
L9	0.7	1.5	6.1	32.1	1.9	26.1	20.6	9.4	1.4	0.2																										
L10												0.2	0.4	25.6	15				15	31.5	9.8	2	0.5													
L11																																				
L12																																				0.1
L13																																				
L14																																				
L15																																				

Allele size (bp)	203	207	211	215	217	219	221	223	225	227	229	231	233	235	237	239	241	243	245	247	249	251	255	262	266	270	274	276	278	280	282	284	286	288		
L1																																				
L2	0.1	1.1	1.8	15.9		10.4		11.1		15.2		18.9		15.7		7.1		2.1		0.6																
L3																									0.2	3	31.8	14.5		11		33.8		4.2		
L4																																				
L5																																				
L6				1.5	5.9	18.3	10	20.4	23.9	15.2	3.8	0.9	0.1																							
L7																																				
L8																																				
L9																																				
L10																																				
L11																																				
L12	0.2	0.4	0.4	1.0	0.2	11.3	1.5	21.1	1.4	19	3	4.4	10.4	2.6	15.2	1.3	4.9	0.3	0.6	0.3	0.2	0.1	0.2													
L13																																				
L14																																				
L15																																				

Allele size (bp)	290	292	294	296	298	300	304	308	312	314	318	320	322	325	326	329	330	332	333	334	336	337	338	340	341	342	344	345	346	348	349	350	352	353		
L1*																																				
L2												7.3		1.5		0.2																				
L3	1.5																																			
L4																																				
L5																																				
L6																																				
L7		21.8		13.2		2.3	0.5																													
L8														0.5	0.5		2.7	19.8		28.6		38.5		7.4												
L9																																				
L10																																				
L11	0.1		0.8		1.2	0.4	2.8	8.6	12.5	2.8	21.9	0.3	16.8		12.2		8		4.2		3.3		1.6		1.2							0.3				
L12																																				
L13														0.2	0.8		1.7	0.4	5.7	0.8	10.9	1.4	10.3	3.1	13.8	3.2	15.5	2								
L14																																				
L15																																				

Allele size (bp)	354	356	357	358	360	362	364	366	370	376	379	381	384	386	389	394	399	404	409	414	419	424	429	434	439	444	449	451	454	459	464	469	474	479		
L1																																				
L2																																				
L3																																				
L4																																				
L5																																				

contd . . .

Appendix 3. (contd) Average allele frequencies (%) at 15 microsatellite loci among predominant Indian populations.

L6																												
L7																												
L8		0.3																										
L9																												
L10																												
L11	0.4		0.1	0.2	0.2	0.1																						
L12																												
L13	14.2	1.2	9.4	0.5	3.1	0.1	1.0	0.5			0.1	0.1																
L14						0.1	0.1		0.1	0.9	2.3	20	20.7	23.7	13.3	10.3	5.8	2.1	0.2	0.4								
L15							5.4	0.2	5.8	0.8	1.9	4.2	13.3	12.7	7.4	6.2	7.9	11.3	8.5	6.1	3.1	0.1	2.7	1.2	0.4	0.3	0.4	0.1

*Note: L1 = vWA; L2 = D8S1179; L3 = TPOX; L4 = D5S818; L5 = D13S317; L6 = D7S820; L7 = D16S539; L8 = CSF1PO; L9 = D3S1358; L10 = TH01; L11 = D18S51; L12 = D21S11; L13 = FGA; L14 = Penta D; L15 = Penta E .

Appendix 4. Observed heterozygosity at 15 microsatellite loci in predominant Indian populations.

Sampling location	Group	vWA	D18S51	D8S1179	FGA	D5S818	D13S317	D7S820	D3S1358	D21S11	D16S53	TPOX	TH01	Penta-D	CSF1PO	Penta-E
Bihar	Brahmin	0.714	0.789	0.785	0.767	0.727	0.672	0.679	0.74	0.824	0.793	0.636	0.771	0.706	0.719	0.762
	Bhumihar	0.672	0.746	0.838	0.793	0.64	0.846	0.734	0.711	0.903	0.738	0.725	0.709	0.828	0.677	0.793
	Rajput	0.803	0.719	0.826	0.886	0.537	0.844	0.719	0.55	0.736	0.785	0.618	0.716	0.725	0.793	0.872
	Kayasth	0.803	0.846	0.82	0.857	0.641	0.867	0.811	0.653	0.826	0.754	0.306	0.846	0.826	0.653	0.826
	Yadav	0.763	0.842	0.811	0.8	0.725	0.846	0.794	0.736	0.789	0.868	0.829	0.75	0.816	0.757	0.947
	Kurmi	0.787	0.857	0.894	0.891	0.708	0.729	0.787	0.694	0.878	0.674	0.638	0.837	0.857	0.694	0.898
	Baniya	0.778	0.822	0.911	0.578	0.644	0.822	0.711	0.556	0.733	0.756	0.689	0.733	0.644	0.667	0.8
Maharashtra	Maratha	0.75	0.792	0.854	0.802	0.745	0.774	0.752	0.733	0.882	0.861	0.624	0.804	0.778	0.73	0.618
	Desasth. B	0.82	0.865	0.888	0.826	0.757	0.783	0.813	0.745	0.841	0.766	0.707	0.785	0.813	0.738	0.829
	Chitpavan. B	0.75	0.881	0.833	0.852	0.853	0.779	0.779	0.721	0.838	0.765	0.613	0.676	0.881	0.612	0.954
	Dhangar	0.714	0.857	0.8	0.778	0.81	0.848	0.808	0.766	0.823	0.823	0.661	0.494	0.873	0.785	0.948
Madhya	Gond	0.692	0.762	0.875	0.6	0.727	0.6	0.478	0.696	0.75	0.75	0.412	0.65	0.696	0.454	0.619
Pradesh	Agharia	0.86	0.82	0.771	0.86	0.736	0.706	0.699	0.698	0.863	0.74	0.553	0.696	0.804	0.745	0.84
	Teli	0.698	0.756	0.838	0.898	0.674	0.829	0.668	0.589	0.791	0.658	0.742	0.769	0.833	0.78	0.878
	Satnami	0.846	0.771	0.697	0.773	0.711	0.889	0.698	0.778	0.842	0.814	0.444	0.75	0.605	0.625	0.727
Orissa	O.Brahmin	0.859	0.824	0.818	0.811	0.684	0.824	0.771	0.842	0.842	0.789	0.745	0.912	0.75	0.785	0.947
	Khandyat	0.741	0.725	0.854	0.819	0.806	0.887	0.774	0.854	0.758	0.741	0.786	0.7	0.758	0.661	0.967
	Karan	0.868	0.819	0.754	0.879	0.709	0.774	0.774	0.803	0.854	0.838	0.733	0.758	0.836	0.661	0.9
	Gopa	0.916	0.883	0.733	0.866	0.783	0.75	0.75	0.733	0.75	0.816	0.716	0.75	0.833	0.7	0.916
	Juang	0.92	0.88	0.82	0.84	0.74	0.82	0.68	0.68	0.88	0.78	0.74	0.74	0.86	0.76	0.86
Saora	Saora	0.8	0.828	0.857	0.942	0.828	0.8	0.828	0.571	0.829	0.774	0.742	0.8	0.857	0.657	0.942
Karnataka	Iyenger	0.784	0.907	0.861	0.861	0.692	0.753	0.723	0.661	0.876	0.861	0.707	0.815	0.815	0.723	0.921
	Lingayat	0.734	0.938	0.822	0.894	0.724	0.714	0.734	0.793	0.857	0.846	0.581	0.785	0.755	0.734	0.876
	Gowda	0.779	0.779	0.745	0.803	0.525	0.745	0.754	0.559	0.779	0.83	0.542	0.678	0.741	0.745	0.864
	Muslim	0.688	0.888	0.755	0.911	0.733	0.733	0.8	0.488	0.733	0.777	0.555	0.688	0.933	0.733	0.8
Sikkim	Nepali	0.81	0.843	0.858	0.8	0.713	0.778	0.855	0.667	0.921	0.742	0.634	0.715	0.804	0.742	0.902
	Bhutia	0.75	0.844	0.813	0.563	0.717	0.688	0.719	0.75	0.652	0.875	0.657	0.657	0.875	0.75	0.902
	Lepcha	0.796	0.812	0.682	0.523	0.773	0.75	0.637	0.728	0.819	0.614	0.546	0.432	0.66	0.66	0.841
Tamil Naidu	T.Kallar	0.842	0.851	0.851	0.891	0.762	0.733	0.752	0.663	0.832	0.792	0.683	0.713	0.733	0.628	0.921
	Vanniyar	0.747	0.839	0.793	0.897	0.747	0.793	0.76	0.805	0.851	0.782	0.678	0.816	0.805	0.678	0.908
	Pallar	0.818	0.901	0.818	0.788	0.697	0.758	0.848	0.848	0.879	0.788	0.818	0.788	0.848	0.606	0.939
	Paraiyar	0.809	0.952	0.857	0.905	0.762	0.857	0.952	0.619	0.809	0.857	0.571	0.714	0.857	0.619	0.857
	Goundar	0.607	0.768	0.839	0.839	0.821	0.768	0.75	0.714	0.857	0.75	0.732	0.768	0.696	0.625	0.821
	Irular	0.759	0.889	0.926	0.759	0.611	0.796	0.593	0.704	0.833	0.704	0.648	0.63	0.796	0.685	0.944
Chakkiliyar	Chakkiliyar	0.694	0.776	0.857	0.816	0.633	0.837	0.857	0.592	0.878	0.694	0.898	0.714	0.694	0.633	0.857
Uttar Pradesh	Thakur	0.76	0.714	0.822	0.8	0.7	0.63	0.771	0.81	0.791	0.814	0.744	0.791	0.73	0.8	0.92
	Jat	0.83	0.74	0.733	0.833	0.73	0.75	0.714	0.85	0.76	0.81	0.691	0.76	0.913	0.79	0.913
	Kurmi	0.882	0.764	0.78	0.966	0.884	0.711	0.723	0.799	0.903	0.823	0.52	0.681	0.76	0.8	0.941
	Khatri	0.88	0.87	0.842	0.69	0.83	0.745	0.79	0.67	0.935	0.77	0.725	0.745	0.704	0.78	0.96

contd . . .

Intrapopulation genomic diversity in India

Appendix 4. (contd) Observed heterozygosity at 15 microsatellite loci in predominant Indian populations.

Sampling location	Group	vWA	D18S51	D8S1179	FGA	D5S818	D13S317	D7S820	D3S1358	D21S11	D16S53	TPOX	THO1	Penta-D	CSFIPO	Penta-E
Andhra Pradesh	Reddy	0.404	0.687	0.823	0.47	0.571	0.711	0.7	0.444	0.65	0.684	0.505	0.679	0.7	0.148	0.889
	SP	0.501	0.8	0.778	0.301	0.48	0.776	0.625	0.381	0.75	0.909	0.884	0.114	0.77	0.664	0.786
	Golla	0.75	0.833	0.811	0.772	0.785	0.647	0.732	0.741	0.826	N.A	N.A	N.A	N.A	N.A	N.A
West Bengal	Brahmin	0.6	0.68	0.6	0.85	0.8	0.688	0.7	0.7	0.65	N.A	0.754	0.739	N.A	0.829	N.A
	Kayasth	0.65	0.7	0.65	0.841	0.795	0.7	0.725	0.695	0.6	N.A	0.755	0.754	N.A	0.828	N.A
	Garó	0.798	0.674	0.674	0.826	0.783	0.797	0.696	0.631	0.888	N.A	0.591	0.736	N.A	0.782	N.A
	Meitei	0.875	0.882	0.688	0.896	0.813	0.759	0.75	0.771	0.788	N.A	0.614	0.659	N.A	0.644	N.A
Ladhak	Buddhist	0.9	0.826	0.814	0.8	0.786	0.842	0.671	0.828	0.828	N.A	N.A	N.A	N.A	N.A	N.A
	Argon	0.778	0.667	0.879	0.939	0.788	0.818	0.758	0.878	0.697	N.A	N.A	N.A	N.A	N.A	N.A
	Drokpa	0.92	0.96	0.96	0.84	0.8	0.68	0.6	0.84	0.84	N.A	N.A	N.A	N.A	N.A	N.A
	Balti	0.707	0.879	0.879	0.948	0.783	0.883	0.776	0.817	0.948	N.A	N.A	N.A	N.A	N.A	N.A
Manipur	Naga	0.667	0.869	0.792	0.792	0.792	0.7	0.667	0.584	0.885	N.A	0.669	0.715	N.A	0.765	N.A
	Kuki	0.834	0.895	0.792	0.542	0.834	0.821	0.709	0.874	0.775	N.A	0.524	0.773	N.A	0.781	N.A
	Hmar	0.75	0.7	0.792	0.667	0.813	0.781	0.688	0.73	0.58	N.A	0.579	0.755	N.A	0.785	N.A
	Muslim	0.889	0.712	0.667	0.834	0.889	0.701	0.778	0.5	0.811	N.A	0.763	0.806	N.A	0.63	N.A

Appendix 5. Power of discrimination / matching probability at 15 microsatellite loci among predominant Indian populations.

Sampling location	Group	vWA	D18S51	D8S1179	FGA	D5S818	D13S317	D7S820	D3S1358	D21S11	D16S53	TPOX	THO1	Penta-D	CSFIPO	Penta-E
Bihar	Brahmin	0.81	0.843	0.84	0.878	0.702	0.284	0.809	0.73	0.835	0.758	0.676	0.751	0.787	0.742	0.905
	Bhumihar	0.819	0.831	0.835	0.871	0.713	0.786	0.785	0.756	0.846	0.806	0.685	0.776	0.82	0.715	0.907
	Rajput	0.798	0.879	0.843	0.883	0.743	0.796	0.803	0.806	0.824	0.825	0.676	0.743	0.832	0.759	0.926
	Kayasth	0.807	0.822	0.84	0.885	0.728	0.824	0.764	0.718	0.842	0.765	0.728	0.769	0.834	0.698	0.901
	Yadav	0.812	0.85	0.848	0.878	0.718	0.797	0.815	0.697	0.819	0.747	0.729	0.779	0.837	0.686	0.912
	Kurmi	0.808	0.848	0.832	0.858	0.671	0.784	0.76	0.755	0.83	0.803	0.729	0.779	0.836	0.726	0.904
	Baniya	0.84	0.835	0.838	0.86	0.791	0.815	0.773	0.69	0.835	0.798	0.76	0.739	0.81	0.793	0.894
Maharashtra	Maratha	0.785	0.858	0.841	0.834	0.708	0.81	0.8	0.725	0.85	0.823	0.704	0.75	0.812	0.72	0.888
	Desasth. B	0.806	0.866	0.841	0.893	0.746	0.842	0.816	0.754	0.839	0.802	0.719	0.788	0.813	0.72	0.908
	Chitpavan.	0.742	0.84	0.856	0.835	0.761	0.798	0.803	0.742	0.819	0.774	0.694	0.748	0.85	0.722	0.918
	Dhangar	0.718	0.86	0.868	0.841	0.754	0.754	0.804	0.78	0.862	0.836	0.808	0.767	0.827	0.739	0.91
Madhya Pradesh	Gond	0.781	0.822	0.785	0.835	0.735	0.766	0.773	0.743	0.849	0.775	0.789	0.741	0.806	0.683	0.891
	Agharia	0.814	0.87	0.835	0.905	0.73	0.807	0.795	0.737	0.852	0.772	0.69	0.767	0.835	0.696	0.909
	Teli	0.848	0.796	0.842	0.872	0.746	0.822	0.8	0.72	0.843	0.771	0.718	0.763	0.821	0.743	0.883
	Satnami	0.781	0.841	0.828	0.873	0.733	0.804	0.785	0.688	0.869	0.808	0.697	0.779	0.802	0.702	0.892
Orissa	O.Brahmin	0.794	0.846	0.841	0.846	0.694	0.821	0.787	0.77	0.842	0.792	0.719	0.838	0.791	0.745	0.895
	Khandyat	0.809	0.792	0.833	0.88	0.748	0.804	0.793	0.737	0.821	0.794	0.748	0.759	0.811	0.736	0.89
	Karan	0.78	0.828	0.85	0.873	0.718	0.833	0.783	0.741	0.85	0.807	0.692	0.761	0.807	0.651	0.904
	Gopa	0.798	0.84	0.833	0.86	0.775	0.795	0.787	0.72	0.827	0.784	0.703	0.775	0.826	0.706	0.854
	Juang	0.829	0.81	0.792	0.861	0.731	0.74	0.802	0.737	0.791	0.718	0.72	0.771	0.829	0.74	0.867
	Saora	0.75	0.837	0.79	0.866	0.71	0.765	0.795	0.77	0.802	0.768	0.656	0.715	0.836	0.681	0.873
Karnataka	Iyenger	0.811	0.809	0.853	0.849	0.701	0.8	0.8	0.698	0.849	0.774	0.687	0.769	0.796	0.704	0.882
	Lingayat	0.789	0.837	0.842	0.865	0.736	0.806	0.793	0.74	0.812	0.808	0.693	0.757	0.807	0.725	0.915
	Gowda	0.806	0.839	0.881	0.858	0.754	0.824	0.817	0.764	0.863	0.818	0.716	0.742	0.792	0.658	0.887
	Muslim	0.783	0.826	0.757	0.815	0.735	0.808	0.82	0.733	0.866	0.804	0.719	0.743	0.82	0.694	0.901
Sikkim	Nepali	0.811	0.87	0.841	0.873	0.753	0.818	0.815	0.742	0.844	0.802	0.711	0.686	0.849	0.728	0.904
	Bhutia	0.772	0.851	0.817	0.693	0.816	0.778	0.744	0.747	0.808	0.796	0.766	0.755	0.815	0.616	0.894
	Lepcha	0.753	0.883	0.844	0.797	0.813	0.814	0.825	0.738	0.889	0.806	0.54	0.661	0.834	0.817	0.899
Tamil Naidu	T. Kallar	0.832	0.868	0.837	0.866	0.777	0.787	0.772	0.741	0.835	0.785	0.684	0.79	0.78	0.669	0.9
	Vanniyar	0.795	0.838	0.842	0.867	0.707	0.782	0.819	0.775	0.85	0.81	0.692	0.775	0.816	0.686	0.9
	Pallar	0.856	0.861	0.835	0.871	0.729	0.828	0.812	0.734	0.819	0.731	0.755	0.781	0.765	0.589	0.882
	Paraiyar	0.814	0.841	0.849	0.853	0.66	0.826	0.796	0.659	0.861	0.829	0.713	0.77	0.82	0.68	0.904
	Goundar	0.722	0.843	0.839	0.859	0.715	0.774	0.82	0.73	0.832	0.76	0.682	0.795	0.762	0.64	0.885
	Irular	0.806	0.817	0.828	0.805	0.7	0.776	0.636	0.765	0.811	0.666	0.674	0.713	0.708	0.701	0.868
	Chakkiliyar	0.811	0.834	0.849	0.848	0.718	0.806	0.782	0.647	0.834	0.787	0.735	0.698	0.74	0.713	0.867

contd . . .

Appendix 5. (contd) Power of discrimination / matching probability at 15 microsatellite loci among predominant Indian populations.

Uttar Pradesh	Thakur	0.856	0.871	0.874	0.915	0.809	0.847	0.798	0.815	0.893	0.819	0.78	0.815	0.833	0.764	0.923
	Jat	0.814	0.789	0.82	0.876	0.696	0.81	0.784	0.764	0.82	0.807	0.742	0.774	0.82	0.745	0.892
	Kurmi	0.727	0.88	0.859	0.884	0.699	0.77	0.79	0.722	0.869	0.801	0.7	0.657	0.79	0.712	0.893
	Khatri	0.855	0.862	0.845	0.902	0.793	0.806	0.792	0.803	0.887	0.818	0.803	0.802	0.812	0.77	0.913
Andhra Pradesh	Reddy	0.77	0.839	0.824	0.855	0.708	0.708	0.645	0.601	0.78	0.806	0.716	0.756	0.813	0.638	0.847
	SP	0.787	0.865	0.821	0.853	0.664	0.84	0.693	0.687	0.832	0.751	0.629	0.778	0.673	0.685	0.874
	Golla	0.824	0.862	0	0.852	0.697	0.813	0.782	0.733	0.893	N.A	N.A	N.A	N.A	N.A	N.A
West Bengal	Brahmin	0.903	0.94	0.855	0.821	0.915	0.899	0.921	0.901	0.888	N.A	0.751	0.826	N.A	0.736	N.A
	Kayasth	0.861	0.861	0.89	0.911	0.91	0.9	0.86	0.89	0.881	N.A	0.752	0.796	N.A	0.75	N.A
	Garo	0.876	0.854	0.739	0.906	0.859	0.874	0.764	0.812	0.954	N.A	0.588	0.777	N.A	0.732	N.A
	Meitei	0.871	0.88	0.871	0.887	0.921	0.9	0.889	0.901	0.911	N.A	0.609	0.758	N.A	0.655	N.A
Ladhak	Buddhist	0.799	0.841	0.827	0.873	0.746	0.817	0.805	0.73	0.834	N.A	N.A	N.A	N.A	N.A	N.A
	Argon	0.821	0.738	0.814	0.844	0.706	0.786	0.747	0.684	0.792	N.A	N.A	N.A	N.A	N.A	N.A
	Drokpa	0.822	0.774	0.816	0.791	0.802	0.744	0.609	0.815	0.824	N.A	N.A	N.A	N.A	N.A	N.A
	Balti	0.772	0.872	0.819	0.81	0.703	0.836	0.82	0.775	0.853	N.A	N.A	N.A	N.A	N.A	N.A
Manipur	Naga	0.732	0.953	0.829	0.869	0.869	0.768	0.732	0.641	0.971	N.A	0.666	0.762	N.A	0.712	N.A
	Kuki	0.915	0.976	0.869	0.694	0.915	0.901	0.878	0.96	0.85	N.A	0.52	0.736	N.A	0.767	N.A
	Hmar	0.827	0.768	0.869	0.732	0.892	0.857	0.765	0.823	0.636	N.A	0.573	0.51	N.A	0.747	N.A
	Muslim	0.871	0.81	0.846	0.877	0.901	0.799	0.814	0.834	0.855	N.A	0.756	0.847	N.A	0.655	N.A

References

- Agrawal S., Muller B., Bhardwaj U., Bhattnagar S. and Khan F. 2002a Allele frequencies of microsatellite repeat loci in Bhargavas, Chaturvedis and Brahmins of North India. *J. Forensic Sci. USA* **47**, 678–681.
- Agrawal S., Muller B., Bhardwaj U., Bhattnagar S. and Khan F. 2002b Distribution of allele frequencies of six STR markers in North Indians. *J. Forensic Sci. USA* **47**, 682–685.
- Agrawal S., Muller B., Bhardwaj U., Bhattnagar S. and Khan F. 2002c Evaluation of six short tandem repeat in forensics: North Indian populations. *J. Forensic Sci. USA* **47**, 686–689.
- Allchin B. and Allchin R. 1982 The rise of civilization in India and Pakistan. Cambridge University Press, Cambridge.
- Ashma R. and Kashyap V. K. 2002a Genetic study of 15 important STR loci among four major ethnic groups of Bihar, India. *J. Forensic Sci. USA* **47**, 1139–1142.
- Ashma R. and Kashyap V. K. 2002b Genetic polymorphism at 15 STR loci among three important sub-populations of Bihar, India. *Forensic Sci. Int. UK* **130**, 58–62.
- Bhasin M. K., Walter H. and Danker-Hopfe 1994 People of India: An investigation of biological variability. In Ecological, ethnological and linguistic groups. Kamla-Raj Enterprises, Delhi.
- Calafell F., Shuster A., Speed W. C., Kidd J. R. and Kidd K. K. 1998 Short tandem repeat polymorphism evolution in humans. *Eur. J. Hum. Genet.* **6**, 38–49.
- CFSL 2002 A national resource: Indian human population DNA repository. Central Forensic Science Laboratory, Kolkata.
- Chakraborty R., Walter H., Sauber P., Mukherjee B. N., Malhotra K. C., Banerjee S. and Roy M. 1998 Immunoglobulin (Gm and Km) allotypes in nine endogamous groups of West Bengal, India. *Ann. Hum. Biol.* **14**, 155–157.
- Chattopadhyay P. and Kashyap V. K. 2001 Distribution of alleles of three tetrameric STR loci HUMHPRTB, HUMF13B, HUMLPL and other six PCR based loci HLADQA1, LDLR, GYPA, HBGG, D7S8, Gc in eight predominant populations of India. *Forensic Sci. Int. UK* **19**, 134–137.
- Chattopadhyay P., Dutta R. and Kashyap V. K. 2001 Allele Frequency data at nine fluorescent labeled STR loci in four tribal population groups in India. *J. Forensic Sci. USA* **45**, 201–205.
- Chattopadhyay P., Dutta R. and Kashyap V. K. 2001 Population data for nine fluorescent based STR loci among four important tribal populations of India. *J. Forensic Sci. USA* **46**, 184–188.
- Das B., Ghosh A., Chauhan P. S. and Seshadri M. 2002 Genetic polymorphism study at four minisatellite loci (D1S80, D17S5, D19S20 and APOB) among five Indian population groups. *Hum. Biol.* **74**, 345–361.
- Deka R., Shriver M. D., Yu L. M., Ferrell R. E. and Chakraborty R. 1995 Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* **16**, 1659–1664.
- Deka R., Shriver M. D., Yu L. M., Heidrich E. M., Jin L., Zhong Y. et al. 1999 Genetic variation at twenty-three microsatellite loci in sixteen human populations. *J. Genet.* **78**, 99–121.
- Destro-Bisol G., Boschi I., Caglia A., Tofanelli S., Pascali V., Paoli G and Spedini G. 2000 Microsatellite variation in central Africa: An analysis of intrapopulation and interpopulation genetic diversity. *Am. J. Phys. Anthropol.* **112**, 319–337.
- Dutta R. and Kashyap V. K. 2001 Allele frequencies of six STR loci among three predominant population groups of eastern India. *Forensic Sci. Int. UK* **119**, 129–130.
- Dutta R., Chattopadhyay P. and Kashyap V. K. 2000 STR data for AmpF1STR Profiler Plus loci among four predominant population of eastern India. *J. Forensic Sci. USA* **45**, 353–357.
- Eller E. 1999 Population substructure and isolation by distance in three continental regions. *Am. J. Phys. Anthropol.* **108**, 147–159.
- Fairservis W. A. 1971 The roots of ancient India. Macmillan, New York.
- Gadgil M., Joshi N. V., Prasad U. V., Manoharan S. and Patil S. 1998 Peopling of India. In *The Indian human heritage* (ed. D. Balasubramanian and N. Appaji Rao), pp. 100–129. Universities Press, Hyderabad.
- Gaikwad S. and Kashyap V. K. 2002 Polymorphism at fifteen hypervariable microsatellite loci in four populations of Maharashtra, India. *Forensic Sci. Int. UK* **126**, 267–271.
- Graber R. A. and Morris J. W. 1983 General equations for the average power of exclusion for genetic system of n co-dominant alleles in one-parent cases of disputed parentage. In *Inclusion probabilities in parentage testing* (ed. R. H. Walker), pp. 277–280. American Association of Blood Banks, Arlington.

- 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., *et al.* 1994 Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* **55**, 175–189.
- Hochberg Y. 1988 A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–802.
- Jin L., Baskett M. L., Cavalli-Sforza L. L., Zhivotovsky L. A., Feldman M. W. and Rosenberg N. A. 2000 Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. *Ann. Hum. Genet.* **64**, 117–134.
- Kashyap V. K., Sarkar N. and Trivedi R. 2001 Allele frequencies for the 15 STR loci of the Powerplex 16™ multiplex system in five endogamous populations of India. *Forensic Sci. Int. UK* **126**, 180–188.
- Kashyap V. K., Guha S. and Trivedi R. 2002 Concordance study on 15 STR loci in three major populations of Himalayan state Sikkim. *J. Forensic Sci. USA* **47**, 1163–1167.
- Lin Z., Cui X. and Li H. 1996 Multiplex genotype determination at a large number of gene loci. *Proc. Natl. Acad. Sci. USA* **93**, 2582–2587.
- Malhotra K. C. 1978 Morphological composition of the people of India. *J. Hum. Evol.* **7**, 45–53.
- Mountain J. L., Hebert J. M., Bhattacharya S., Underhill P. A., Ottolenghi C., Gadgil M. and Cavalli-Sforza L. L. 1995 Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56**, 979–992.
- Mukherjee N., Majumder P. P., Roy B., Dey M., Chakraborty M. and Banerjee S. 1999 Variation at short tandem repeat loci in eight population groups of India. *Hum. Biol.* **71**, 439–446.
- Nei M. 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**, 3321–3323.
- Nei M. 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.
- Nei M. and Roychoudhary A. K. 1974 Sampling variance in heterozygosity and genetic distance. *Genetics* **76**, 379–390.
- Panneerchelvam S., Vanaja N., Baskar D., Sivapriya V. and Damodaran C. 2001 Distribution of alleles of 12 STR loci in Tamil population (South India). *Forensic Sci. Int. UK* **119**, 126–128.
- Papiha S. S., Mastana S. S., Purandare C. A., Jayasekera 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. *Hum. Biol.* **68**, 819–835.
- Rajkumar R. and Kashyap V. K. 2002 Distribution of alleles of fifteen STR loci of the Powerplex 16™ multiplex in four predominant population groups of South India. *Forensic Sci. Int. UK* **126**, 175–179.
- Reddy B. M., Sun G., Luis J. R., Crawford M. 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. *Hum. Biol.* **73**, 175–190.
- Relethford J. H. and Jorde L. B. 1999 Genetic evidence for large African population size during recent human evolution. *Am. J. Phys. Anthropol.* **108**, 251–260.
- Roychoudhury A. K. 1992 Genetic relationship of the populations in eastern India. *Ann. Hum. Biol.* **19**, 489–501.
- Sahoo S. and Kashyap V. K. 2002a Allele frequency of data for Powerplex 16™ loci in four major populations of Orissa. *J. Forensic Sci. USA* **47**, 912–915.
- Sahoo S. and Kashyap V. K. 2002b Genetic variation at fifteen autosomal microsatellite loci in three highly endogamous tribal population of Orissa, India. *Forensic Sci. Int. UK* **130**, 189–193.
- Sambrook J., Fritsch E. F. and Maniatis T. 1989 Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Sarkar N. and Kashyap V. K. 2002 Genetic diversity at two pentanucleotide and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of Central India. *Forensic Sci. Int. UK* **128**, 196–201.
- Singh K. S. 1998 India's communities. In *People of India, National Series Volume (IV–VI)*. Anthropological Survey of India and Oxford University Press, Delhi.
- Sitalaximi T., Trivedi R. and Kashyap V. K. 2002 Genotype profile for thirteen tetranucleotide repeat loci and two pentanucleotide repeat loci in four endogamous Tamil population groups of India. *J. Forensic Sci. USA* **47**, 1168–1173.
- Sitalaximi T., Trivedi R. and Kashyap V. K. 2003 Autosomal microsatellite profile of three socially diverse ethnic Tamil populations of India. *J. Forensic Sci. USA* **48**, 211–214.
- Tahir M. A., Herrera R. J., Khan A. A., Kashyap V. K., Duncan G., Barna C., Budowle B., Rowold D. J., Sinha S. and Amjad M. 2002 Distribution of HLA-DQA1, Polymarker, CSF1PO vWA, TH01, TPOX, D16S539, D7S820, D13S317 and D5S818 allele in east Bengali and west Punjabi populations from Indo-Pak subcontinent. *J. Forensic Sci. USA* **45**, 1320–1323.
- Tandon M., Trivedi R. and Kashyap V. K. 2002 Genomic diversity at fifteen fluorescent labeled short tandem repeat loci in few important populations of states of Uttar Pradesh, India. *Forensic Sci. Int. UK* **123**, 190–195.
- Trivedi R., Chattopadhyay P., Maity B. and Kashyap V. K. 2002 Genetic polymorphism at nine microsatellite loci in four and altitude Himalayan desert human populations. *Forensic Sci. Int. UK* **127**, 150–155.
- Watterson G. A. and Guess H. A. 1977 Is the most frequent allele the oldest? *Theor. Popul. Biol.* **11**, 141–160.
- Waye J. S., Presley L., Budowle B., Shulter G. G. and Fournery R. M. 1989 A simple method of quantifying human genomic DNA in forensic specimen extracts. *Biotechniques* **7**, 852–855.
- Weber J. L. and Wong C. 1993 Mutation of human short tandem repeats. *Hum. Mol. Genet.* **8**, 1123–1128.
- Weir B. S. 1992 Independence of VNTR alleles defined by fixed bins. *Genetics* **130**, 873–887.

Received 22 October 2002; in revised form 2 July and 19 November 2003