Genetic variation at fifteen microsatellite loci in human populations of India

V. K. Kashyap[†,*], Neeta Sarkar[†], Sanghamitra Sahoo[†], B. N. Sarkar[#] and R. Trivedi[†]

[†]DNA Typing Unit, Central Forensic Science Laboratory, 30, Gorachand Road, Kolkata 700 014, India
[#]Anthropological Survey of India, Kolkata

The genetic diversity in the Indian populations has been analysed at fifteen microsatellite loci among fifty-four endogamous groups comprising 3647 individuals. The populations belong to all major linguistic families and ethnic groups from diverse geographical regions of the country, occupying different positions in the socio-cultural hierarchy. Allelic variation was studied at all the fifteen microsatellite loci and was found to be 83.6%, which is higher than the values obtained from classical genetic markers. The coefficient of genetic differentiation (G_{ST}) among the fifty-four populations was also found to be higher (4.1%) than that of the most global populations.

Analysis of molecular variance and the genetic affinity among various linguistic, geographical and ethnic groups were plotted based on principal component analysis, which indicated that geographical proximity was the main factor determining genetic affinity among the populations (0.92%). The maximum genetic variance 'between populations within groups' could be attributed more to the social status (2.37%) than to the geographical (1.66%), ethnic (0.83%) or linguistic groups (0.79%). Our work clearly demonstrates that STR markers may delineate different ethnic, linguistic and geographical populations of India.

THE people of India exhibit a unique range of social, cultural, linguistic and biological diversity. Archaeological evidence suggests that human habitation in the Indian subcontinent began from the Early Stone Age, dating around 250,000 yrs BP (YBP). Stretching back in an unbroken sweep over 5000 YBP, people of different ethnic stocks, cultures and languages started migrating to India from different directions and contributed significantly to the present-day gene pool of the subcontinent[1–5]. The varied ecological regime of the country seems to have nurtured this diversity[6]. The different waves of historical migration in the Indian subcontinent encompass the Indo-Europeans (2000–1400 BC), Greeks (400–200 BC), Sakas (200 BC), Kushanas (AD 100), Huns (AD 200–500) and Arabs (AD 800). These populations entered India and slowly merged with the local populations[7,8]. The main sources of this genetic influx were: (i) the Maracan coast and Baluchistan in the west; and (ii) Southeast Asia and South China through India's northeastern border[9]. Many scholars believe that the Dravidians migrated from Asia Minor region and brought the language to India, but a few believe that they owe their origin to Neolithic people of southern India[3,10].

Indian populations include four ethnic groups: Austroloid, Negrito, Mongoloid, and Caucasoid. Caucasoid and Mongoloid populations are mainly concentrated in the north and northeastern parts of the country. The Australoid groups are mostly confined to the central, western and southern India, while the Negritos are restricted only to the Andaman Islands[11]. Linguistically, Indian populations belong to four major language families: Dravidian, Austro-Asiatic, Tibeto-Burman and Indo-European (Indo-Aryan). Majority of Indians speak Indo-European or Dravidian languages, spread over the northern and southern parts of the subcontinent, respectively. The Tibeto-Burman languages are largely spoken in the north-eastern parts, while the Austro-Asiatic languages are restricted among the tribal groups, viz. Korkus, Mundas, Santals, Khasis, and Nicobarese. It is believed that the speakers of this language family came through the eastern region of the subcontinent much before the Indo-European language came to India[2]. The people of India form thousands of endogamous units under the broad categories of castes, tribes and different religious groups. The *People of India* project recognized 4635 distinct communities. Many of them are clusters of endogamous groups with similar traditional occupations and social status. However, it is believed that the actual number of endogamous groups is much larger and of the order of 50,000 to 60,000 in India[11,12].

Several studies based on the classical genetic markers were carried out to understand the genetic variations among the Indian populations. The blood groups, red cell enzymes and serum proteins were extensively studied among regional, ethnic and linguistic groups to understand the genetic history, origin and affinities of the Indian populations[1,8,13–19]. All these studies gave only an outline trend of genetic variations among geographical groups of the country. In addition, due to the low level of

*For correspondence. (e-mail: vkk2k@hotmail.com)

polymorphism observed with classical genetic markers, there has been a paradigm shift in the tools used in population genetic studies. The application of PCR technology and availability of a large number of polymorphic loci have revolutionized the analysis of genetic variation in humans. Among these loci, the microsatellites are abundant in the genome[20], having a relatively high heterozygosity (usually > 0.5) and large number of alleles due to high mutation rates[21]. These features, along with automated typing procedure[22], make them the most desirable tool for scoring genetic variability in the human population, inferring evolutionary relationships and demographic histories[23–25], and in human identification for determining parentage and relatedness of individuals[26–28]. Although a considerable amount of information on polymorphism at microsatellite loci in the Indian human population is now available[29–41], the studies are confined to limited population groups with different sets of marker for the various populations. These genotype data are inconsistent, heterogeneous and non-comparable. Thus, it is not possible to analyse them statistically for delineation of the populations.

With this rationale, we have analysed fifteen STR markers with two main objectives: (i) to assess the nature and extent of variations in microsatellite loci and (ii) to examine the genetic diversity and affinity amongst various Indian populations.

Materials and methods

Populations

Fifty-four populations analysed in this study were selected on the basis of ethnicity, prevalence of endogamy and anthropological significance, among which nine populations belong to the Mongoloids, twenty-three represent the Caucasoids and fourteen represent the Australoids. Besides these, eight have been categorized as mixed populations.

Linguistically, this study includes fifteen Dravidian-speaking populations, twenty-four Indo-European and eleven Tibeto-Burman speakers. Only four populations represent the Austro-Asiatic speakers. The criteria for the selection of population and other ethical issues are already published[42].

With regard to the geographical distribution of these populations, fourteen have been drawn from the southern region, from the states of Andhra Pradesh, Tamil Nadu and Karnataka, four each from the western and central regions, fifteen from the northern and seventeen from eastern regions of the country. Thus, the populations selected in the present study represent a wide-ranging spectrum of linguistically, geographically and ethnically diverse groups of India. Details of the studied populations are given in Table 1 and the distribution of sampled populations is shown in Figure 1.

STR markers used

The fifteen autosomal microsatellite loci used in this study include thirteen tetranucleotide repeats (D3S1358, THO1, D18S51, D21S11, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX and FGA) and two pentanucleotide repeats (PENTA E and PENTA D). These microsatellites are located on different chromosomes. The only exceptions are a subset of two loci on chr 21 and 5 but they do not have any reported linkage on the chromosome. Thus, all the loci used in the present study are substantially unlinked, which make them ideal tools for analysing inter population genetic diversity.

Sample collection, DNA isolation and amplification

Blood was obtained in EDTA-coated vaccutainers from consenting unrelated donors. Community, health status and family disease histories were recorded on blood-donor cards of the DNA Typing Unit, Central Forensic Science Laboratory (CFSL), Kolkata. DNA was extracted by organic method[43]. Quality and quantity assessment of the isolated DNA was performed using 0.8% agarose yield gel and slotblot procedures using probe D17Z1 (ref. 44).

The amplification was performed using a 25 µl final reaction volume containing 1 ng of sample DNA and 0.5 unit of Taq DNA polymerase with STR buffer and primers specific for the multiplex. The primers and buffers were obtained for fifteen co-amplified STR loci system (Trade name: Power Plex 16™ System) from Promega Corporation (Madison, WI), and nine co-amplified STR system (Trade name: AmpF*l*STR Profiler Plus) from Applied Biosystem, USA. PCR amplification was carried out at PE 2400 Model PCR Machine (Perkin Elmer Biosystems, Palo Alto, CA) following the manufacturer's protocols. The amplified products were separated for genotyping on 377 Automated DNA Sequencer.

Statistical analysis

Genetic diversity of the fifty-four populations was analysed by calculating average heterozygosity across the fifteen STR loci[45,46]. Locus-wise gene diversity of population for the different linguistic, ethnic and geographical regions was also calculated independently. Two different statistical measures, coefficient of gene differentiation (G_{ST}) and molecular variance analysis (AMOVA) were conducted at the ethnic, linguistic and geographic levels to understand the structure of genetic variability among them. AMOVA was performed using Arlequin version 2.1 (ref. 47), at two levels of population substructuring (among groups and between populations within each group), based on fifteen microsatellite loci among forty-one diverse Indian populations. Principal component

Table 1. Sample size and some important demographic information on the studied populations of India

Population	Abbreviation used	Location	Sample size	Linguistic affiliation	Traditional occupation	Socio-cultural affiliation	Demographic history
Buddhist	BD	Ladakh, Jammu and Kashmir	156	Tibeto-Burman	Priesthood	Religious category	Migrated from Tibet
Argon	AR	Ladakh, Jammu and Kashmir	51	Tibeto-Burman	Trade and commerce	–	Mixed population
Drokpa	DR	Ladakh, Jammu and Kashmir	38	Tibeto-Burman	Trade and agriculture	–	–
Baltis	BA	Ladakh, Jammu and Kashmir	67	Tibeto-Burman	Trade and agriculture	Tribe	Migrated from Balistan
Nepali	NE	Sikkim	110	Indo-European	Agriculture	Tribe	Migrated from adjacent area
Lepcha	LP	Sikkim	48	Tibeto-Burman	Traditionally hunter-gatherer; at present cultivar	Tribe	Migrated from adjacent areas
Bhutia	BH	Sikkim	75	Tibeto-Burman	Terrace cultivation	Tribe	Migrated from Tibet
Naga	NG	Manipur	106	Tibeto-Burman	Shifting cultivation	Tribe	Original inhabitant
Kuki	KK	Manipur	105	Tibeto-Burman	Shifting cultivation	Tribe	Original inhabitant
Hmar	HM	Manipur	101	Tibeto-Burman	Shifting cultivation	Tribe	Migrated from adjacent areas
Garo	GR	West Bengal	110	Tibeto-Burman	Shifting cultivation	Tribe	Migrated from Garo hill
Manipuri Muslim	MM	Manipur	101	Indo-European	Agriculture	Religious group	Migrated from adjacent areas
Meitei	ME	West Bengal	105	Indo-European	Agriculture	Caste	Original inhabitant
Bengali Brahmin	WB	West Bengal	110	Indo-European	Priesthood	Caste	Migrated from adjacent areas
Bengali Kayastha	KW	West Bengal	103	Indo-European	Service	Caste	Original inhabitant
Agharia	AG	Madhya Pradesh	70	Indo-European	Agriculture	Backward caste	Migrated from Agra
Dheria Gond	DG	Madhya Pradesh	35	Dravidian	Agriculture	Tribal	Original inhabitant
Satnami	SA	Madhya Pradesh	50	Indo-European	Traditional leather worker	Lower caste	Oldest resident of Chhattisgarh
Teli	TL	Madhya Pradesh	50	Indo-European	Oil-pressing community	Backward class	Migrated from adjacent areas
Oriya Brahmin	OB	Orissa	57	Indo-European	Priesthood	Upper caste	Migrated from adjacent areas
Khandayat	KH	Orissa	62	Indo-European	Agriculture	Caste	Original inhabitant
Karan	KA	Orissa	62	Indo-European	Traditional scribers, presently agriculture	Caste	Original inhabitant
Gope	GP	Orissa	60	Indo-European	Milkman	Caste	Original inhabitant
Juang	JU	Orissa	50	Austro-asiastic	Traditionally shifting cultivation	Tribe	Original inhabitant
Saora	SO	Orissa	35	Austro-asiatic	Traditionally shifting cultivation	Tribe	Original inhabitant
Bihar Brahmin	BR	Bihar	59	Indo-European	Priesthood	Caste	Migrated from adjacent areas
Bhumihar	BU	Bihar	65	Indo-European	Agriculture	Caste	Descendant from Brahman
Rajput	RA	Bihar	58	Indo-European	Agriculture	Caste	Migrated from adjacent areas
Bihar Kayastha	BK	Bihar	53	Indo-European	Traditional scribes, presently agriculture	Caste	Migrated from adjacent areas
Yadav	YA	Bihar	44	Indo-European	Animal husbandary	Caste	Original inhabitant
Bihar Kurmi	KU	Bihar	50	Austro-Asiatic	Agriculture	Tribe	Original inhabitant
Bihar Baniya	BN	Bihar	45	Indo-European	Moneylenders and business	Caste	Migrated from adjacent areas
Sakanupakshollu	SP	Andhra Pradesh	30	Dravidian	–	Caste	Migrated from adjacent areas
Reddy	RC	Andhra Pradesh	30	Dravidian	Cultivator	Caste	Original inhabitant
Gounder	GU	Tamil Nadu	56	Dravidian	Agriculture	Caste	Original inhabitant
Irular	IR	Tamil Nadu	54	Dravidian	Hunter-gatherer	Tribe	Original inhabitant

Contd...

Table 1. (*Contd...*)

Population	Abbreviation used	Location	Sample size	Linguistic affiliation	Traditional occupation	Socio-cultural affiliation	Demographic history
Chakkliyar	CK	Tamil Nadu	49	Dravidian	Washerman	Caste	Migrated from Andhra Pradesh
Golla	GO	Andhra Pradesh	65	Dravidian	Pastoral	Caste	Original inhabitant
Vanniyar	VN	Tamil Nadu	87	Dravidian	Oil-pressing community	Caste	Original inhabitant
Pallar	PA	Tamil Nadu	33	Dravidian	Agriculture	Caste	Original inhabitant
Parriyar	PR	Tamil Nadu	21	Dravidian	Agriculture	Caste	Original inhabitant
Kallar	KL	Tamil Nadu	101	Dravidian	Traditional feudal chieftains	Caste	Original inhabitant
Iyengar Brahmin	IB	Karnataka	65	Dravidian	Priesthood	Caste	Migrated from adjacent areas
Lingayat	LI	Karnataka	98	Dravidian	Cultivation	Caste	Migrated from adjacent areas
Gowda	GW	Karnataka	56	Dravidian	Agriculture	Caste	Migrated from neighbouring areas
Karnataka Muslim	MU	Karnataka	65	Dravidian	Agriculture	Religious group	Migrant
Desasth Brahmin	DB	Maharashtra	70	Indo-European	Priesthood	Caste	Migrated from adjacent areas
Marathas	MA	Maharashtra	65	Indo-European	Agriculture	Caste	–
Dhangar	DH	Maharashtra	150	Indo-European	Shepherds	Backward caste	Migrated from adjacent areas
Chitpavan Brahmin	CB	Maharashtra	78	Indo-European	Priesthood	Caste	Migrated from adjacent areas
Khatri	KT	Uttar Pradesh	47	Indo-European	Businessmen	Caste	Migrated from western India
UP Kurmi	KM	Uttar Pradesh	45	Austro-Asiatic	Agriculture	Backward community	Migrated from Bihar
UP Jat	JA	Uttar Pradesh	48	Indo-European	Agriculture	Caste	Migrated from Punjab and Rajasthan
UP Thakur	TH	Uttar Pradesh	48	Indo-European	Agriculture	Caste	Original inhabitant

analysis was performed to determine the genetic distance between each group of the geographic, ethnic and linguistic categories using the SPSS software.

Results

Number of alleles and heterozygosity at microsatellite loci

The distribution of allele frequency at fifteen microsatellite loci shows a high degree of polymorphism in all fifty-four populations. The numbers of alleles range from 16 to 28 at D18S51, D21S11, FGA and PENTA E loci and 5 to 10 at other loci, suggesting modest polymorphism in the latter. Locus-specific allele frequencies and other statistical parameters of population genetics (HWE and exact test) and forensic analysis for all the studied populations are reported in other studies[48–62]. The low-frequency alleles in any particular group do not necessarily preclude their presence in the other groups because of the small number of individual samples analysed.

The average heterozygosity estimates are presented in Table 2. The heterozygosity ranges from 0.761 ± 0.018 (Manipuri Muslim) to 0.866 ± 0.012 (Thakur of Uttar Pradesh), with the overall mean for fifty-four populations of India being 0.813. Although the range of average heterozygosity overlaps between different geographical regions, it is observed that Indo-European, Tibeto-Burman and Dravidian speakers show an extended range of heterozygosity (0.761–0.866), while it is limited in the Austro-Asiatic speakers (0.803–0.828).

Gene diversity

To examine the intra- and inter-population variations in 15 microsatellite loci of the studied Indian populations, the genetic diversity (H_s) and G_{ST} estimates were analysed locus-wise and collectively under the categories of ethnic, language and geographical groups. The gene diversity for the populations classified according to linguistic affiliation is presented in Table 3. Variations within the populations as described by the H_s values show that Indo-Europeans have a higher value (80.1%) compared to the Dravidians, Tibeto-Burmans and Austro-Asiatics (78.8%). The overall value of the coefficient of gene differentiation for the Dravidians is 3.4%, which is higher

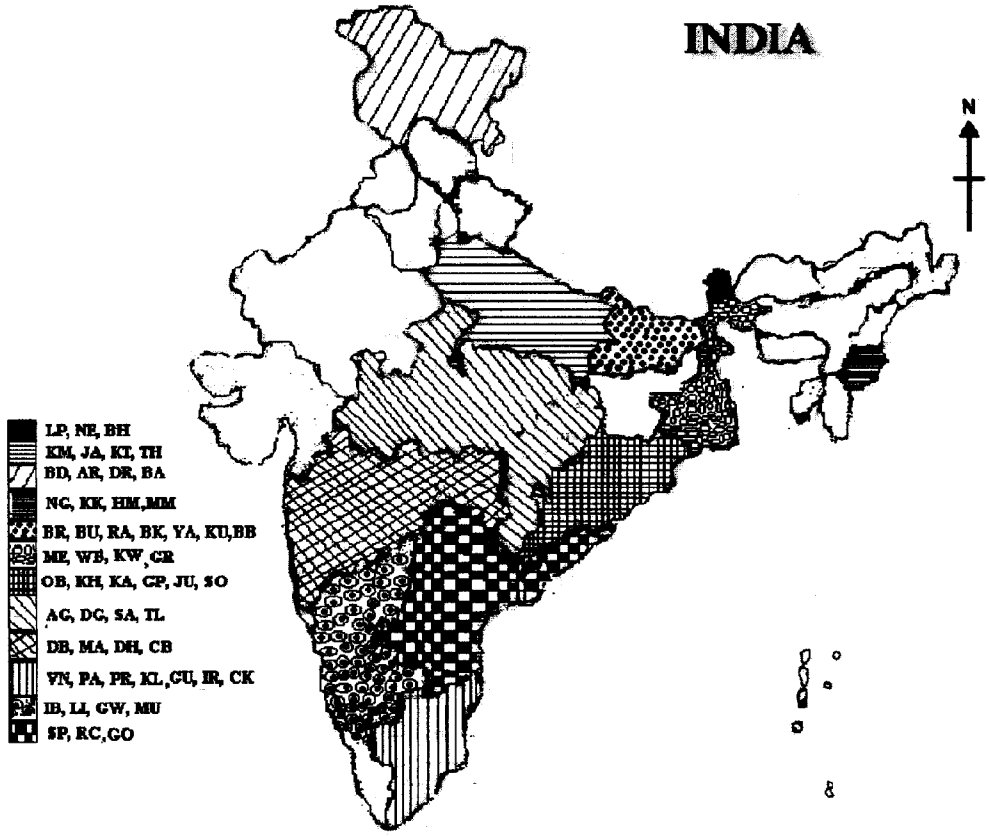

Figure 1. Distribution of samples studied for genetic variation at fifteen microsatellite loci in Indian population. (Abbreviations correspond to populations presented in Table 1.)

Table 2. Average heterozygosity based on fifteen microsatellite loci in 54 populations of India

Population	Average heterozygosity	SE	Population	Average heterozygosity	SE
Buddhist	0.795	± 0.016	Dhangar	0.809	± 0.019
Argon	0.782	± 0.021	Golla	0.811	± 0.018
Drokpa	0.791	± 0.017	Vanniyar	0.811	± 0.016
Baltis	0.803	± 0.020	Pallar	0.841	± 0.018
Agharia	0.824	± 0.019	Parriyar	0.833	± 0.027
Dheria Gond	0.813	± 0.017	Manipuri Muslim	0.761	± 0.018
Satnami	0.809	± 0.019	Bengali Brahmin	0.823	± 0.010
Teli	0.811	± 0.019	Bengali Kayastha	0.828	± 0.011
Nepali	0.822	± 0.017	Meitei	0.813	± 0.017
Lepcha	0.831	± 0.012	Yadav	0.815	± 0.019
Bhutia	0.786	± 0.015	Bihar Kurmi	0.803	± 0.019
Oriya Brahmin	0.816	± 0.017	Bihar Baniya	0.817	± 0.016
Khandayat	0.811	± 0.015	Naga	0.822	± 0.016
Karan	0.817	± 0.018	Kuki	0.820	± 0.010
Gope	0.815	± 0.014	Hmar	0.824	± 0.012
Sakanupakshollu	0.809	± 0.027	Garo	0.806	± 0.022
Reddy	0.784	± 0.026	Iyengar Brahmin	0.824	± 0.025
Gounder	0.809	± 0.020	Lingayat	0.822	± 0.019
Irular	0.787	± 0.022	Gowda	0.834	± 0.013
Chakkliyar	0.809	± 0.023	Karnataka Muslim	0.842	± 0.019
Bihar Brahmin	0.821	± 0.019	Khatri	0.852	± 0.011
Bhumihar	0.821	± 0.018	UP Kurmi	0.828	± 0.022
Rajput	0.835	± 0.015	UP Jat	0.816	± 0.016
Bihar Kayastha	0.819	± 0.019	Thakur	0.866	± 0.012
Desastha Brahmin	0.832	± 0.016	Kallar	0.821	± 0.015
Marathas	0.814	± 0.018	Juang	0.804	± 0.015
Chitpavan Brahmin	0.809	± 0.014	Saora	0.810	± 0.016

Table 3. Gene diversity at fifteen STR loci amongst the four linguistic groups in India

	Dravidian (n = 14)		Indo-European (n = 20)		Tibeto-Burman (n = 3)		Austro-Asiatic (n = 4)	
Locus	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}
D3S1358	0.717	0.064	0.745	0.023	0.735	0.005	0.746	0.025
THO1	0.757	0.018	0.841	0.025	0.699	0.015	0.765	0.023
D18S51	0.849	0.027	0.842	0.025	0.867	0.012	0.840	0.019
D21S11	0.846	0.047	0.902	0.015	0.844	0.010	0.826	0.027
PENTAE	0.887	0.026	0.740	0.032	0.898	0.019	0.884	0.018
D5S818	0.717	0.028	0.807	0.019	0.787	0.025	0.705	0.026
D13S317	0.795	0.029	0.795	0.014	0.781	0.018	0.765	0.014
D7S820	0.778	0.037	0.795	0.011	0.794	0.022	0.787	0.019
D16S539	0.778	0.039	0.727	0.016	0.801	0.004	0.773	0.021
CSF1PO	0.675	0.033	0.819	0.014	0.719	0.023	0.715	0.012
PENTAD	0.790	0.022	0.797	0.026	0.831	0.021	0.826	0.020
vWA	0.798	0.021	0.841	0.017	0.775	0.026	0.803	0.023
D8S1179	0.839	0.020	0.723	0.042	0.831	0.036	0.818	0.029
TPOX	0.710	0.054	0.872	0.041	0.672	0.024	0.701	0.024
FGA	0.847	0.041	0.768	0.019	0.785	0.089	0.869	0.019
Mean	0.786	0.034	0.801	0.023	0.788	0.024	0.788	0.021

Table 4. Gene diversity at fifteen STR loci in five geographical regions of India

	Southern (n = 13)		Northern (n = 11)		Western (n = 4)		Eastern (n = 9)		Central (n = 4)	
Locus	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}
D3S1358	0.715	0.065	0.752	0.025	0.748	0.024	0.839	0.019	0.722	0.019
THO1	0.758	0.014	0.762	0.015	0.705	0.041	0.892	0.020	0.762	0.022
D18S51	0.851	0.027	0.846	0.017	0.847	0.011	0.815	0.028	0.832	0.014
D21S11	0.843	0.050	0.857	0.024	0.824	0.034	0.747	0.028	0.853	0.014
PENTA E	0.887	0.024	0.907	0.012	0.898	0.018	0.788	0.025	0.893	0.014
D5S818	0.716	0.028	0.742	0.047	0.737	0.012	0.796	0.015	0.731	0.014
D13S317	0.797	0.029	0.801	0.027	0.794	0.037	0.793	0.020	0.799	0.008
D7S820	0.779	0.037	0.784	0.025	0.796	0.008	0.784	0.013	0.788	0.021
D16S539	0.778	0.039	0.795	0.014	0.805	0.015	0.710	0.020	0.781	0.012
CSF1PO	0.675	0.034	0.738	0.016	0.731	0.015	0.821	0.021	0.706	0.009
PENTAD	0.789	0.019	0.821	0.009	0.823	0.019	0.787	0.027	0.812	0.033
vWA	0.799	0.021	0.799	0.020	0.772	0.029	0.826	0.031	0.791	0.008
D8S1179	0.844	0.016	0.845	0.017	0.845	0.011	0.694	0.025	0.822	0.018
TPOX	0.704	0.056	0.726	0.061	0.717	0.021	0.837	0.051	0.724	0.095
FGA	0.848	0.040	0.869	0.054	0.867	0.041	0.742	0.027	0.871	0.021
Mean	0.786	0.033	0.803	0.026	0.794	0.023	0.741	0.023	0.793	0.021

than that of the Tibeto-Burman and the Indo-Europeans, 2.3% each, while the value is 2.1% among the Austro-Asiatic speakers. Locus-specific G_{ST} values showed that the populations of Tibeto-Burman are well differentiated at locus FGA and the Dravidians are distinguishable at D3S1358 locus. On the other hand, both Austro-Asiatic and Indo-Europeans can be distinctly differentiated at D8S1179 locus.

Table 4 depicts the gene diversity in the populations of five geographical regions comprising thirteen populations from the southern, eleven from the northern, nine from the eastern and four each from the western and central regions, respectively. The highest genomic diversity between individuals within population (H_S) is observed in the northern region (80.3%) followed by the western (79.4%), central (79.3%), southern (78.6%) and eastern (74.1%) regions of the country. The coefficient of gene diversity (G_{ST}) varies between 2.1% for central India and 3.3% for southern India. G_{ST} values of the western and eastern regions show a similar trend at 2.3%, while the northern region shows a relatively higher value (2.6%).

The gene diversity among ethnic groups is presented in Table 5. The genomic diversity within populations of diverse ethnic origins suggests that the Caucasoid groups have the highest value (80%) followed by the Mongoloids (78.8%), with the least in the Australoids (78.1%). Overall value of the coefficient of gene differentiation for the Australoids is 3.9%, higher than that of the Cauca-

Table 5. Gene diversity at fifteen STR loci across four ethnic groups of India

	Australoid (n = 13)		Mongoloid (n = 3)		Caucasoid (n = 22)		Mixed (n = 3)	
Locus	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}	H_S	G_{ST}
D3S1358	0.719	0.068	0.735	0.005	0.743	0.022	0.748	0.030
THO1	0.756	0.024	0.699	0.015	0.769	0.018	0.752	0.012
D18S51	0.839	0.029	0.867	0.012	0.842	0.024	0.865	0.023
D21S11	0.829	0.050	0.844	0.010	0.843	0.025	0.869	0.024
PENTA E	0.881	0.029	0.898	0.019	0.903	0.015	0.904	0.012
D5S818	0.715	0.032	0.787	0.025	0.741	0.032	0.732	0.003
D13S317	0.786	0.035	0.781	0.018	0.804	0.023	0.814	0.007
D7S820	0.776	0.043	0.794	0.022	0.790	0.017	0.797	0.007
D16S539	0.769	0.043	0.801	0.004	0.790	0.012	0.804	0.006
CSF1PO	0.673	0.037	0.719	0.023	0.726	0.015	0.689	0.015
PENTA D	0.795	0.027	0.831	0.021	0.818	0.014	0.802	0.005
vWA	0.797	0.029	0.775	0.026	0.799	0.024	0.786	0.011
D8S1179	0.828	0.030	0.831	0.036	0.841	0.016	0.858	0.004
TPOX	0.706	0.063	0.672	0.024	0.722	0.040	0.713	0.005
FGA	0.846	0.044	0.785	0.089	0.870	0.040	0.868	0.011
Mean	0.781	0.039	0.788	0.024	0.800	0.023	0.800	0.012

Table 6. Percentage of variation at three levels of population hierarchy by AMOVA in four distinct clusters of populations in India

		Percentage total variance		
Basis	No. of groups	Between groups	Within population	Between population within group
Language	4	0.39	98.86	0.79
Geographical	5	0.92	97.42	1.66
Ethnicity	4	0.34	98.84	0.83
Social	3	0.04	97.59	2.37

soids (2.3%) and the Mongoloids (2.4%), while the G_{ST} value is only 1.2% among the mixed populations.

Genetic diversity was explored by AMOVA among geographic, ethnic, linguistic and social categories (Table 6). The gene differences within populations were found to be large in all groups (>97%). The genomic diversity among the populations within the social groups (high, low-ranking caste groups and tribes) is higher (2.37%) than the geographical (1.66%), ethnic (0.83%) and linguistic groups (0.79%). On the other hand, diversity between populations of different geographical regions is larger (0.92%) than the linguistic groups (0.39%); however, the difference among the ethnic groups (0.34%) and the social categories (0.04%) is low in Indian populations. This is clearly reflected in the plots of the principal component analysis drawn on the basis of data on allele frequencies at fifteen microsatellite loci.

Genetic affinities among Indian populations

To assess the evolutionary relationships of the Indian populations, the allele frequencies of the fifteen STR loci were used for estimating the genetic distances between them and were drawn as PC plots. On the basis of microsatellite diversity, the Dravidians and Tibeto-Burmans appear well differentiated from Austro-Asiatic and Indo-Europeans in the principal component (PC) plots (Figure 2 *a*) in which the total genomic variance accounts for 77.6%. Similarly, the two-dimensional genetic picture (Figure 2 *b*) of geographic groups shows that populations of east, west and north are in different quadrants, but the central and south Indian populations appear in the same quadrant, where only 54.1% depicts the total genetic variance. Thus, the populations of south and central India are genetically closer than populations from other parts of the country. The PC plot (Figure 2 *c*) of ethnicity shows that Caucasoids, Mongoloids and Austroloids are totally distinct from each other, describing 61.1% of the total variance.

Discussion

The present study reveals genetic variation at fifteen microsatellite loci in diversified populations of India. With respect to the distribution of alleles, each STR locus was found to be substantially polymorphic in all the populations, irrespective of their size. This is clearly evident from various measures of genetic variations such as number of alleles and heterozygosity. The range of allele frequencies among the populations of three ethnic groups, viz. Australoids, Mongoloids and Caucasoids was found to be similar and could not be differentiated at CSF1PO, D13S317, D3S1358 and D7S820 loci.

High heterozygosity in Indo-Caucasoids can be correlated to their large effective population size and greater degree of genetic isolation from other ethnic groups. Based on traditional 82 gene markers for forty-two populations, Cavalli-Sforza *et al.*[11] compared the populations

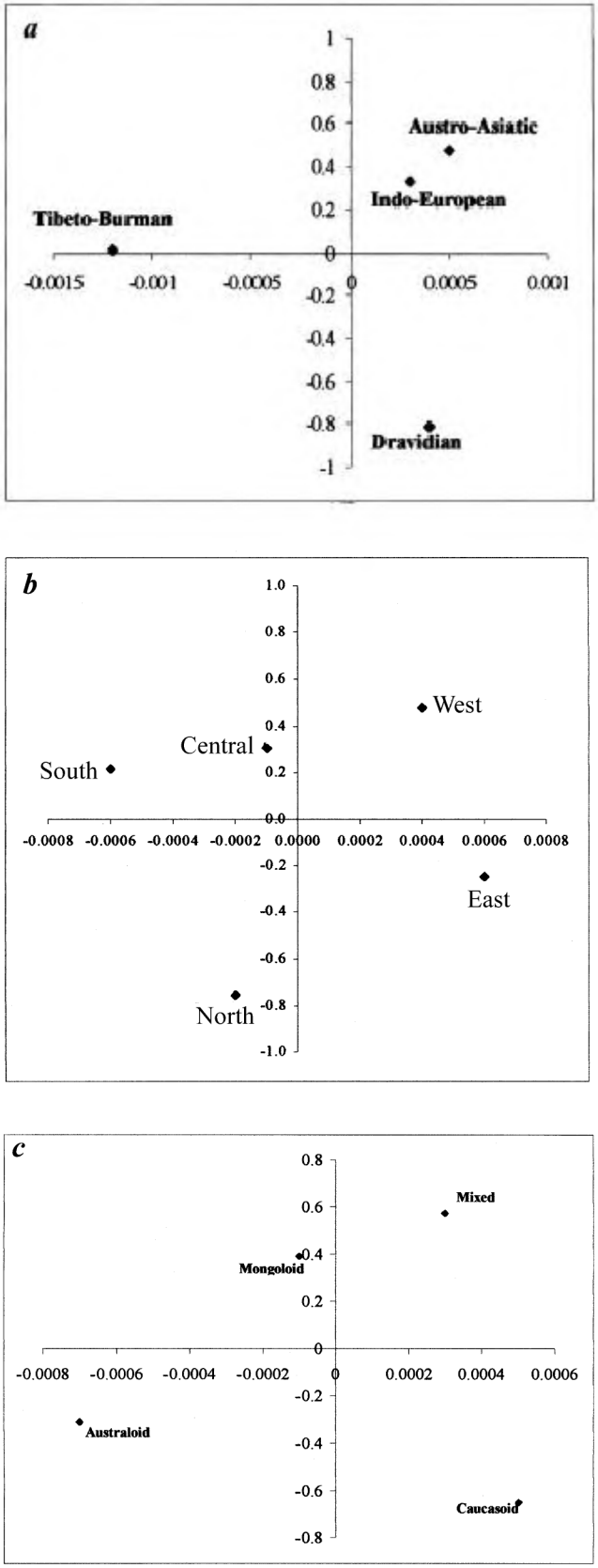

Figure 2. Principal component plot based on fifteen microsatellite loci showing the genetic distance among the four linguistic families of India (*a*), across geographical clusters of India (*b*), and among ethnic groups of India (*c*).

at the global level, which revealed the highest level of heterozygosity (0.35–0.37) for northwestern India, west Asia and continental Europe. Parts of southern and eastern India showed a slightly lower level of heterozygosity (0.33–0.35) compared to western China, Central Asia, Scandinavia and northern Africa. In another global study, Deka *et al.*[29] observed a total diversity of 76%, of which 1.4% was ascribed to the inter-population variations within various ethnic groups. A similar trend was observed based on VNTR loci study from different population groups of India[8]. Our finding based on microsatellite loci shows a high level of heterogeneity in the population groups of southern, northern and eastern India. Likewise, observed heterogeneity was found to be high among the Dravidian and Tibeto-Burman speakers. Among the fifty-four populations, the Mongoloids: Manipuri Muslims, Buddhists, Drokpa, Argon and Bhutia and two Austroloid populations: Irular and Reddy show relatively moderate (< 80%) heterozygosity. The low level of diversity at the STR loci may be due to microevolutionary processes acting in these populations. The high level of genetic heterogeneity observed in the fifteen STR markers across the populations indicates their utility in analysing human identification for forensic purposes as well as in population genetics.

Coefficient of coancestary, AMOVA and phylogenetic analysis based on allele frequencies at fifteen microsatellite loci suggest that in general, populations of the same geographical region tend to have greater degree of similarity. The high level of variation in genetic heterogeneity among the Tibeto-Burman-speaking populations might be due to different degree of admixture in the people from Tibetan origin, while the genetic heterogeneity of the populations living in South India are influenced by genetic drift and close geographical proximity of the populations with each other. The observed genetic similarity between the people inhabiting the south and central regions is not a striking feature. Some Dravidian-speaking tribes in central India have adopted the Indo-European languages in recent times[63]. Similarly, the genetic affinities between the Austro-Asiatic and Indo-European-speaking populations might be due to their prolonged coexistence in close geographical proximity. However, the study based on STR markers among eight Indian populations is not in agreement with our findings, where geographical proximity has a greater influence on genetic affinity among populations than socio-cultural proximity[33].

Our study clearly reveals that Indian populations harbour greater genetic diversity than most of the geographical territories of the world, which is reflected by high intra-population variation (> 97%). However, variation between groups is relatively low as evident from the G_{ST} estimates and AMOVA, implying genetic closeness of populations. The physio-geography of India seems to have played a major role in evolving the observed genetic

differentiation among populations. Another distinct inference of our study is that the social categories depict high inter-population differentiation (2.37%) within groups and an exceedingly low value (0.04%) between groups. This low variance between groups may be due to non-discrete classification of the populations into upper caste, lower caste and the tribal groups. Our study explicitly proves the utility of STR markers in deciphering genetic diversity of Indian populations; these are globally used to understand population dynamics and diseases specific to population in the present scenario.

1. Bhasin, M. K., Walter, H. and Danker-Hopfe, *People of India: An Investigation of Biological Variability in Ecological Ethno-economic and Linguistic Groups*, Kamla-Raj Enterprises, Delhi, 1994.
2. Gadgil, M., Joshi, N. V., Prasad, U. V., Manoharan, S. and Patil, S., in *The Indian Human Heritage* (eds Balasubramanian, D. and Appaji Rao, N.), Universities Press, Hyderabad, 1998, pp. 100–129.
3. Allchin, B. and Allchin, R., *The Rise of Civilization in India and Pakistan*, Cambridge University Press, Cambridge, 1982.
4. Sankalia, H. D., Prehistory and Protohistory of India and Pakistan, Deccan College, Poona, 1974.
5. Mishra, S., The age of the Acheulian in India: New evidence. *Curr. Anthropol.*, 1992, **33**, 325.
6. Gadgil, M. and Guha, R., *The Fissured Land: An Ecological History of India*, Oxford University Press, New Delhi, 1992.
7. Malhotra, K. C., Morphological composition of the people of India. *J. Hum. Evol.*, 1978, **7**, 45–53.
8. Papiha, S. S., Mastana, S. S., Purandare, C. A., Jayasekera, R. and Chakraborty, R., Population genetic study of three VNTR loci (D2S44, D7S22 and D12S11) in five ethnically defined populations of the Indian subcontinent. *Hum. Biol.*, 1996, **68**, 819–835.
9. Fairservis, W. A., *The Root of Ancient India*, Macmillan, New York, 1971.
10. Rapson, E. J., in *Cambridge History of India*: *Ancient India* (ed. Rapson, E. J.), S. Chand & Co, Delhi, 1955, vol. 1, pp. 33–57.
11. Cavalli–Sforza, L. L., Menozzi, P. and Piazza, A., *The History and Geography of Human Genes*, Princeton University Press, Princeton , NJ, 1994, pp. 208–213.
12. Singh, K. S., in *People of India, National Series Volume (IV–VI)*, Anthropological Survey of India and Oxford University Press, Delhi, 1998.
13. Joshi, N. V., Gadgil, M., and Patil, S., Exploring cultural diversity of the people of India. *Curr. Sci.*, 1993, **64**, 10–17.
14. Gadgil, M. and Malhotra, K. C., Adaptive significance of the Indian caste system: an ecological perspective. *Ann. Hum. Biol.*, 1983, **10**, 465–478.
15. Roy Chaudhury, A. K., in *Peoples of India: Some Genetical Aspects* (ed. Satyavati, G. V.), Indian Council of Medical Research, New Delhi, 1983, pp. 1–29.
16. Majumder, P. P. and Mukherjee, B. N., in *Human Population Genetics* (ed. Majumder, P. P.), Plenum Press, New York, 1993, pp. 255–275.
17. Malhotra, K. C. and Vasulu, T. S., in *Proceedings of Haldane Centenary Conference of Human Genetics* (ed. Majumder, P. P.), Plenum Press, New York, 1993, pp. 207–223.
18. Das, K., Malhotra, K. C., Mukherjee, B. N., Walter, H., Majumder, P. P. and Papiha, S. S., Population structure and genetic differentiation among 16 tribal populations of Central India. *Hum. Biol.*, 1996, **68**, 679–705.
19. Saha, N. and Tay, J. S., Genetic studies among the Nagas and Hmars of eastern India. *Am. J. Phys. Anthropol.*, 1990, **82**, 101–102.
20. Dib, C., Fizames, F., Samson, D., Drouot, N., Vignal, A., Millasseau, P. and Mara, P., A comprehensive genetic map of the human genome based on 5624 microsatellites. *Nature*, 1996, **380**, 152–154.
21. Weber, J. L. and Wong, C., Mutation of human short tandem repeats. *Hum. Mol. Genet.*, 1993, **2**, 1123–1128.
22. Lin, Z., Cui, X. and Li, H., Multiplex genotype determination at a large number of gene loci. *Proc. Natl. Acad. Sci (USA)*, 1996, **93**, 2582–2587.
23. Bowcock, A. M., Ruiz-Linares, A., Tomforde, J., Minch, E., Kidd, J. R. and Cavalli-Sforza, L. L., High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 1994, **368**, 455–457.
24. Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S. and Jorde, L. B., Signatures of population expansion in microsatellite repeat data. *Genetics*, 1998, **148**, 1921–1930.
25. Calafell, F., Shuster, A., Speed, W. C., Kidd, J. R. and Kidd, K. K., Short tandem repeat polymorphism evolution in human. *Eur. J. Hum. Genet.*, 1998, **6**, 38–49.
26. Budowle, B., Moretti, T. R., Niezgoda, S. J. and Brown, B. L., in *Second European Symposium on Human Identification*, Promega Corporation, Madison, WI, 1998, pp. 73–88.
27. Chakraborty, R. and Jin, L., Determination of relatedness between individuals using DNA fingerprinting. *Hum. Biol.*, 1993, **65**, 875–895.
28. Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y. and Budowle, B., The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis*, 1990, **20**, 1682–1696.
29. Deka, R., Shriver, M. D., Yu, L. M., Ferrell, R. E. and Chakraborty, R., The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis*, 1995, **16**, 1659–1664.
30. Deka, R. *et al.*, Genetic variation at twenty-three microsatellite loci in sixteen human populations. *J. Genet.*, 1999, **78**, 99–121.
31. Mountain, J. L., Hebert, J. M., Bhattacharyya, S., Underhill, P. A., Ottolenghi, C., Gadgil, M. and Cavalli-Sforza, L. L., Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.*, 1995, **56**, 979–992.
32. Pandian, S. K. *et al.*, Allele frequency distribution for the variable number of tandem repeat locus D10S28 in Tamil Nadu (South India) population. *Electrophoresis*, 1995, **16**, 1689–1692.
33. Mukherjee, N., Majumder, P. P., Roy, B., Roy, M., Dey, B., Chakraborty, M. and Banerjee, S., Variation at 4 short tandem repeat loci in 8 population groups of India. *Hum. Biol.*, 1999, **71**, 439–446.
34. Dutta, R. and Kashyap, V. K., Allele frequency of six STR loci among three predominant population groups of eastern India. *Forensic Sci. Int.*, 2001, **119**, 129–130.
35. Chattopadhyay, P., Dutta, R. and Kashyap, V. K., Population data for nine fluorescent based STR loci among four important tribal populations of India. *J. Forensic Sci.*, 2001, **46**, 184–188.
36. Reddy, B. M., Sun, G., Ranjan, D. and Deka, R., STR data for the Amp F1STR Profiler Plus loci among Golla population of southern Andhra Pradesh, India. *J. Forensic Sci.*, 2001, **46**, 734–735.
37. Pannerchelvam, S. *et al.*, Allele frequency for D1S80 locus in Tamil population (South India). *J. Forensic Sci.*, 2001, **46**, 177.
38. Tahir, M. A. *et al.*, Distribution of HLA-DQA1, polymarker, CSF1PO, vWA, TH01, TPOX, D16S539, D7S820, D13S317 and D5S818 alleles in East Bengali and West Punjabi populations from Indo-Pak Subcontinent. *J. Forensic Sci.*, 2000, **45**, 1320–1323.

39. Agrawal, S. *et al.*, Allele frequencies of microsatellite repeat loci in Bhargavas, Chaturvedis and Brahmins of north India. *J. Forensic Sci.*, 2002, **47**, 682–685.
40. Das, B., Chauhan, P. S. and Sheshadri, M. S., Genetic polymorphism study at four minisatellite loci (D1S80, D17S5, D19S20 and APOB) among five Indian population groups. *Hum. Biol.*, 2002, **74**, 345–361.
41. Ghosh, A. and Sheshadri, M., Distribution of allele frequencies for two short tandem repeats (HUMTHO1 and F13A01) among five Indian population groups. *J. Forensic Sci.*, 2002, **47**, 415–416.
42. Report, Central Forensic Science Laboratory, Kolkata.
43. Sambrook, J., Fritsch, E. F. and Manniatis, T. in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, 2nd edn.
44. QuantiBlot™ Human DNA Quantitation Kit, Perkin Elmer, USA.
45. Nei, M., Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*, 1973, **70**, 3321–3323.
46. Chakraborty, R., Genetic profile of cosmopolitan populations: effects of hidden subdivision. *Anthropol. Anz.*, 1990, **48**, 313–331.
47. Excoffier, L., Smouse, P. E. and Quattro, J. M., Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 1992, **131**, 479–491.
48. Trivedi, R., Chattopadhyay, P., Maity, B. and Kashyap, V. K., Genetic polymorphism at nine microsatellite loci in four high altitude Himalayan desert human population. *Forensic Sci. Int.*, 2002, **127**, 150–155.
49. Kashyap, V. K., Guha, S. and Trivedi, R., Concordance study on fifteen STR loci in three major populations of Himalayan state, Sikkim. *J. Forensic Sci.*, 2002, **47**, 1163–1167.
50. Chattopadhyay, P., Dutta, R. and Kashyap, V. K, Allele frequency data at nine fluorescent labeled STR loci in four tribal population groups in India. *J. Forensic Sci.*, 2001, **45**, 201–205.
51. Dutta, R., Chattopadhyay, P. and Kashyap, V. K., STR data for AmpF/STR profiler plus loci among four predominant populations of eastern India. *J. Forensic Sci.*, 2000, **45**, 1353–1357.
52. Sarkar, N. and Kashyap, V. K., Genetic diversity at two pentanucleotide and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of Central India. *Forensic Sci. Int.*, 2002, **128**, 196–201.
53. Sahoo, S. and Kashyap, V. K., Allele frequency data for Powerplex 16 loci in four major populations of Orissa, India. *J. Forensic Sci.*, 2002, **47**, 912–915.
54. Sahoo, S. and Kashyap, V. K., Genetic variation at fifteen autosomal microsatellite loci in three highly endogamous tribal populations of Orissa, India. *Forensic Sci. Int.*, 2002, **130**, 189–193.
55. Ashma, R. and Kashyap V. K., Genetic study of fifteen important STR loci among four major ethnic groups of Bihar, India. *J. Forensic Sci.*, 2002, **47**, 1139–1142.
56. Ashma, R. and Kashyap, V. K., Genetic polymorphism at 15 STR loci among three important sub-populations of Bihar, India. *Forensic Sci. Int.*, 2002, **130**, 58–61.
57. Kashyap, V. K., Sarkar, N. and Trivedi, R., Allele frequencies for the fifteen STR loci of the Powerplex 16™ in five endogamous populations of India. *Forensic Sci. Int.*, 2001, **126**, 180–188.
58. Sitalaxmi, T., Trivedi, R. and Kashyap, V. K., Genotype profile for thirteen tetranucleotide repeat loci and two pentanucleotide repeat loci in four endogamous population groups of India. *J. Forensic Sci.*, 2002, **47**, 1168–1173.
59. Sitalaximi, T., Trivedi, R. and Kashyap, V. K., Autosomal microsatellite profile of three socially diverse ethnic Tamil populations of India. *J. Forensic Sci.*, 2002, **48**, 211–214.
60. Rajkumar, R. and Kashyap, V. K., Distribution of alleles of fifteen STR loci of the Powerplex 16 Multiplex in four predominant populations of South India. *Forensic Sci. Int.*, 2002, **126**, 175–179.
61. Gaikwad, S. and Kashyap, V. K., Polymorphism at fifteen hypervariable microsatellite loci in four populations of Maharashtra, India. *Forensic Sci. Int.*, 2002, **126**, 267–271.
62. Tandon, M., Trivedi, R. and Kashyap, V. K., Genomic diversity at fifteen fluorescent labeled short tandem repeat loci in few important populations of state at Uttar Pradesh, India. *Forensic Sci. Int.*, 2002, **123**, 190–195.
63. Rakshit, H. K., Ethnohistory of the tribal populations of middle India. *J. Indian Anthropol. Soc.*, 1980, **15**, 97–112.

ACKNOWLEDGEMENTS. The work was supported by financial grant from IX Plan Funds of Govt. of India to the CFSL, Kolkata. This study would not have been possible without the active volunteer participation of the subjects. The technical assistance of Ms R. Ashma, Ms R. Rajkumar, Mr S. Guha, Mr B. Maity, Ms S. Gaikwad, Ms T. Sitalaximi, Mr M. Tandon, Dr P. Chattopadhyay and Dr R. Dutta, in genotyping of the samples is duly acknowledged. We also thank Dr S. E. Hasnain, Director, CDFD, Hyderabad and Dr Deepak Tyagi, Director, Anthropological Survey of India, Kolkata for the constant encouragement and suggestions. Remarks and comments of the reviewers have helped us to improve the manuscript considerably. S.S. and N.S. thank BPR&D, New Delhi for the fellowship.

Received 26 May 2003; revised accepted 25 June 2003