Genomic Diversities and Affinities among Four Endogamous Groups of Punjab (India) Based on Autosomal and Mitochondrial DNA Polymorphisms

INDERJEET KAUR,[1] SANGITA ROY,[2] SUBHABRATA CHAKRABARTI,[3] VIRINDER KAUR SARHADI,[3] PARTHA P. MAJUMDER,[4] A.J.S BHANWER,[1] AND JAI RUP SINGH[1,3]

Abstract Nineteen insertion/deletion and restriction site polymorphisms on autosomal and mitochondrial genomes and mitochondrial DNA hypervariable segment 1 sequences were used to study genetic diversities and affinities among four endogamous groups of Punjab, India. High values of heterozygosity were noted in all four groups, both in the autosomal and mitochondrial genomes. The coefficient of gene differentiation among the groups, however, was found to be low. Genetic distance and phylogenetic analyses based on these data indicated that inferences on affinities among the populations were different when the two sets of loci (autosomal and mitochondrial) were considered separately. We have interpreted these results on the basis of some known historical data on migrations into this region. The results of this study when compared with the findings of some previous studies indicate that there are regional differences in the patterns of correlation between genomic and sociocultural affinities within India.

The study of evolution and genetic relationships among contemporary human populations has shed light on the origin of modern humans and also on the nature, extent, and causes of genetic differentiation. India represents one of the most interesting global regions in human evolution, primarily because some of the most ancient out-of-Africa waves of human migration appear to have passed through India (Cann 2001). Further, northern India has been a virtual melting pot of humans who have entered India in historic times (Thapar 1966). As groups of humans entered India, there must have been differential admixture of genes brought in by these immigrants with those of preexisting local ethnic groups, who were already socially structured in a hierarchical manner. Therefore, population groups of northern India are of great interest from the standpoint of the peopling of India. In the present study, we have included four endogamous groups resident in the

[1]Department of Human Genetics, Guru Nanak Dev University, Amritsar, India.
[2]Human Genetics and Genomics Department, Indian Institute of Chemical Biology, Calcutta, India.
[3]Centre for Genetic Disorders, Guru Nanak Dev University, Amritsar, India.
[4]Anthropology and Human Genetics Unit, Indian Statistical Institute, Calcutta, India.

Human Biology, December 2002, v. 74, no. 6, pp. 819–836.

KEY WORDS: ALU INSERTION/DELETION, MTDNA, GENETIC DIVERSITY, POLYMORPHISM, GENETIC DISTANCE

north Indian State of Punjab, with a view to understanding the extent of genetic diversities and affinities among them. This study has been conducted using two sets of DNA markers, autosomal and mitochondrial. The autosomal set includes eight insertion/deletion (InDel) polymorphisms. The mitochondrial set includes 11 markers (10 restriction site polymorphisms [RSPs] and 1 InDel) and hypervariable segment–1 (HVS1) sequences.

Alu elements are a family of short interspersed elements (SINEs) with about 500,000 members distributed throughout the primate genome (Batzer and Deininger 1991; Deininger and Batzer 1993). The human-specific *Alu* InDels are highly polymorphic in most human populations, and since their ancestral states are known, these InDels are particularly useful in tracing the ancestry through phylogenetic trees more effectively (Perna et al. 1992; Batzer et al. 1994, 1995, 1996). *Alu* insertions do not occur multiple times at the same chromosomal location; therefore, the sharing of *Alu* elements is necessarily because of common ancestry. This fact makes the *Alu* elements useful markers for various population genetic studies (Batzer et al. 1994; Stoneking et al. 1997). Only a single major study using *Alu* InDel polymorphisms has been conducted in India (Majumder et al. 1999). This study on 14 ethnic groups of India, however, did not include populations from Punjab. Lack of data on north Indian populations on these InDel polymorphisms, especially from Punjab, motivated us to undertake this study. Further, although some earlier mitochondrial DNA (mtDNA) studies (Kivisild et al. 1999) on Indian populations have been based on HVS1 sequence data from individuals drawn from Punjab, these samples were included without regard to their endogamous group identity. It was also of interest to cross-validate the findings of Bamshad et al. (2001) on the origin of Indian castes. Bamshad et al. (2001) found that caste groups in the uppermost rung of the social hierarchy were the closest to Europeans, and that this affinity declined as one descended the social ladder of the caste groups.

In the present study, eight human-specific InDel loci were examined on 192 individuals, 48 each from four ethnic groups, namely, Brahmins, Khatris, Jat Sikhs, and Scheduled Castes of Punjab, India. Of these eight loci, six were *Alu* insertion loci; the seventh locus (*CD4*) was a deletion locus. Further, we have also included an eighth locus, which is an mtDNA segment insertion in the human nuclear genome (Zischler et al. 1995). Ten mtDNA RSPs and one mtDNA InDel have been screened. In addition, we have sequenced the HVS1 segment (nt [nucleotide] 16024–16380) of the mtDNA in a subset of 57 individuals (about 15 individuals drawn randomly from each of the four populations).

Materials and Methods

Populations Studied. Four endogamous groups of Punjab consisting of Brahmins, Khatris, Jat Sikhs, and Scheduled Castes were studied for eight InDel loci. The Brahmins are geographically distributed over the entire State of Punjab. They are traditionally priests and torch-bearers of Aryan rituals, but now they participate in various, primarily white-collar, occupations. The Khatris occupy the central region of Punjab. Their main occupation is trade. There are some subgroups of this caste. Consanguinity is practiced within the subgroups, but the level of inbreeding is low. Jat Sikhs are predominantly agriculturists and inhabit all regions of the state. They may have entered Punjab after the Brahmins and Khatris (Ibbetson 1916a, b). The Scheduled Castes are primarily menial workers and occupy the lowest rung in the social hierarchy. They also inhabit all regions of the state. According to the Hindu caste system in Punjab, Brahmins are at the top of the social ladder and the Scheduled Castes are at the lowermost rung. The position of Jat Sikhs and Khatris is controversial. According to one view Jat Sikhs are at the top since they are the most predominant caste group (≈60%) among the Sikhs (followers of Sikhism). The social ranking of the Khatris, or the Kshatriyas of the original Manu's classification of castes, is below the Brahmins and just above the Scheduled Castes. The origin of Jat Sikhs is also disputed. While one view is that they are one of the Rajput tribes, another is that the Rajputs belong to the original Aryan stock that entered India, and the Jats belong to a later wave of immigrants of Indo-Scythian stock (Ibbetson 1916a,b). A total of 192 blood samples of unrelated individuals, 48 from each endogamous group belonging to different social strata, were collected randomly from different geographical locations of Punjab, India.

Laboratory Analyses. After obtaining prior informed consent of the study participants by researchers of the Guru Nanak Dev University, Amritsar, blood samples (5–10 mL by venipuncture) were collected in sterile EDTA vials. DNA was isolated in the laboratory of Guru Nanak Dev University, using the method of Gill et al. (1987). Each DNA sample was screened with respect to 8 autosomal *Alu* In/Del polymorphisms. DNA samples were amplified by polymerase chain reaction (PCR) using locus-specific primers. The primer sequences and PCR protocols used in this study are given in Majumder et al. (1999). Amplified PCR products were run on agarose gel, stained with ethidium bromide, and visualized under UV light. To size the alleles, a *Hae*III-digested (X174 DNA marker was also run simultaneously.

Each DNA sample was also screened for 10 mtDNA RSPs and one InDel polymorphism (IDP). The RSPs screened were *Hae*III nt 663, *Hpa*I nt 3592, *Alu*I nt 5176, *Alu*I nt 7025, *Dde*I nt 10394, *Alu*I nt 10397, *Hin*fI nt 12308, *Hin*cII nt 13259, *Alu*I nt 13262, and *Hae*III nt 16517; the IDP screened was the COII/tRNA^{Lys} intergenic 9-base-pair (bp) deletion. These sites were chosen such that individuals could be classified into haplogroups that are most relevant for Indian populations. Mitochondrial DNA RSP analyses were performed using standard primers and protocols (Torroni et al. 1993, 1996).

Sequencing of the mtDNA HVS1 was carried out by the cycle sequencing method in an ABI-3100 automated DNA sequencer and the ABI prism dideoxyterminator system. The HVS1 region (nt 16024–nt 16380) was amplified using standard primers (Vigilant et al. 1991). Genotyping and DNA sequencing were carried out at the Indian Statistical Institute.

Statistical Analyses. Allele frequencies were calculated by the gene counting method. The proportion of heterozygous individuals at each locus, standard error (SE), and the average heterozygosity were calculated from the estimated allele frequencies. The coefficient of gene differentiation (G_{ST}), was also calculated (Nei 1973). Genetic distances between pairs of populations based on allele frequencies were estimated using the D_A distance measure (Nei 1973). Neighbor-joining (NJ) trees were constructed on the basis of these estimated distances (Saitou and Nei 1987). The results have been compared with the available data for other Indian populations to obtain an overview of the populations studied. DNA sequences were aligned using ClustalW. Appropriate statistics pertaining to nucleotide sequence variability (nucleotide diversity, average number of nucleotide differences) were computed. A maximum-likelihood tree of the distinct sequences was calculated using PHYLIP (http://evolution.genetics.washington.edu/phylip.html).

Results

Locus-specific allele frequencies for the populations studied are presented in Table 1. Allele frequencies at the seven insertion loci ranged from 35% to 93%. The deletion allele at the *CD4* locus exhibited low frequencies in all the populations. Majumder et al. (1999) had also reported allele frequencies of between 29% and 94% for the insertion loci in their study populations and low frequencies of the deletion allele at the *CD4* locus.

The heterozygosity values at the individual loci and the average heterozygosity values and their standard errors for each population were estimated (Table 2). Except for *CD4* and *APO*, high levels of heterozygosity were observed at all the loci in all the populations. The average heterozygosities were remarkably sim-

Table 1. Allele Frequencies at Eight Polymorphic Autosomal InDel Loci in Four Endogamous Groups of Punjab

	Alu	*Alu*	*Alu*	*Alu*	*Alu*	*Alu*	*Alu*	
	ACE	*PV92*	*FXIIIB*	*D1.1*	*CD4*	*APO*	*PLAT*	*MtNUC*
Population	+[a]	+	+	+	–	+	+	+
Brahmin ($n = 48$)	0.4895	0.3541	0.5418	0.4687	0.0834	0.9062	0.4468	0.5729
Khatri ($n = 48$)	0.6250	0.4062	0.5937	0.5104	0.0625	0.8750	0.5119	0.5833
Jat Sikh ($n = 48$)	0.6979	0.3958	0.5937	0.5208	0.1563	0.9270	0.4559	0.5625
Scheduled Castes ($n = 48$)	0.4791	0.5000	0.6875	0.3645	0.0521	0.9062	0.4896	0.4895

a. Insertion (+) and deletion (–) are relative to primates.

Table 2. Heterozygosity Values at Eight Polymorphic InDel Loci in Four Endogamous Groups of Punjab

Population	*ACE*	*FXIIIB*	*PV92*	*D1.1*	*CD4*	*APO*	*PLAT*	*MtNUC*	*Average Heterozygosity ± S.E.*
Brahmin	0.4996	0.4996	0.4573	0.4980	0.1528	0.1699	0.4943	0.4996	0.4160 ± 0.0550
Khatri	0.4687	0.4823	0.4823	0.4997	0.1171	0.2187	0.4997	0.4921	0.4156 ± 0.0543
Jat Sikh	0.4216	0.4824	0.4782	04991	0.2636	0.1351	0.4961	0.4860	0.4173 ± 0.0490
Scheduled Castes	0.4990	0.4296	0.5	0.4632	0.099	0.1699	0.4997	0.4892	0.4034 ± 0.0592

ilar and high in all the four populations studied. These estimated values are consistent with those reported from other Indian populations (Majumder et al. 1999)

The observed heterozygosity (H_S) within the populations and the total heterozygosity (H_T) at all the loci were calculated (Table 3). There was wide variation (0.0027–0.0348) in the estimated coefficients of gene differentiation (G_{ST}) across loci. G_{ST} values at the *APO*, *PLAT*, and *MtNUC* loci were about an order of magnitude lower than those observed for the other loci. The coefficient of gene differentiation for all the loci considered together was 0.0134, indicating that within-population variation is much greater than between-population genetic variation.

The neighbor-joining tree based on the estimated D_A distances (Table 4), which were calculated on the basis of allele frequencies of all eight loci, is depicted in Figure 1. As can be seen from Table 4 and Figure 1 (an unrooted tree), the Brahmins and the Khatris are genetically the closest to each other. The Jat Sikhs are genetically close to the Khatris, but are distant from the Brahmins. All three groups are genetically very distant from the Scheduled Castes.

Table 3. Results of Genetic Diversity Analysis Based on Autosomal Loci

	Heterozygosity		
Locus	*Within-Population (H_S)*	*Total (H_T)*	*Coefficient of Gene Differentiation (G_{ST})*
ACE	0.4723	0.4894	0.0348
FXIIIB	0.4728	0.4783	0.0115
PV92	0.4795	0.4852	0.0117
D1.1	0.4901	0.4977	0.0154
CD4	0.1581	0.1614	0.0205
APO	0.1735	0.1742	0.0040
PLAT	0.4975	0.4988	0.0027
MtNUC	0.4919	0.4946	0.0055
Combined	0.4045	0.4100	0.0134

Table 4. Pairwise Genetic Distances based on Frequencies at Autosomal InDel Loci (above the Diagonal) and Frequencies of mtDNA Haplotypes (below the Diagonal) among Four Endogamous Groups of Punjab

	Brahmin	*Khatri*	*Jat Sikh*	*Scheduled Castes*
Brahmin	–	0.0022	0.0042	0.0043
Khatri	0.0937	–	0.0026	0.0046
Jat Sikh	0.0461	0.0579	–	0.0084
Scheduled Castes	0.1305	0.0650	0.0961	–

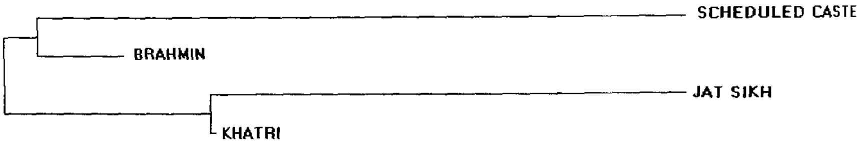

Figure 1. Neighbor-joining tree depicting genetic relationships among four endogamous groups of Punjab based on eight autosomal insertion/deletion polymorphisms.

The *Alu* ACE polymorphism is known to be associated with cardiovascular disease and may be influenced by selective pressures. We therefore examined whether the pattern of affinities changes by exclusion of this locus. The topology of the neighbor-joining tree based on D_A distances calculated after removing data on the *Alu ACE* locus remained unchanged.

With respect to the 11 biallelic mtDNA polymorphisms, 13 distinct haplotypes were observed among the 192 individuals screened from the four populations (Table 5). None of the individuals possessed the COII/tRNA^{Lys} intergenic 9-bp deletion. Three of the 10 restriction sites (*Hae*III nt 663, *Hpa*I nt 3592, and *Alu*I nt 5176) were also monomorphic. Only three of the haplotypes are common in all populations; the remaining 10 haplotypes are present sporadically. While the haplotype 00111101010 is the most frequent haplotype among the Khatris, Jat Sikhs, and Scheduled Castes, this is not the most common haplotype among the Brahmins. The most common haplotype (00110001010) among the Brahmins is the second most frequent haplotype among the Khatris and the Jat Sikhs, but is uncommon among the Scheduled Castes. Based on the haplotype frequencies, we estimated the haplotype diversities within each of the four populations, which are: Brahmins = 0.836, Jat Sikhs = 0.829, Khatris = 0.790, and Scheduled Castes = 0.731. The coefficient of differentiation with respect to mtDNA haplotype frequencies, G_{ST}, was estimated to be 0.0264, which is much higher than the value estimated with respect to autosomal DNA markers. Thus, compared to autosomal markers, there is substantially greater within-population variation in mtDNA hap-

Table 5. Mitochondrial DNA Haplotype Frequencies (Percentages) in Four Endogamous Groups of Punjab

				Scheduled	
	Brahmin	*Khatri*	*Jat Sikh*	*Castes*	*Total*
Haplotype[a]	*(n = 48)*	*(n = 48)*	*(n = 48)*	*(n = 48)*	*(n = 192)*
001**00001000**	1 (2.1)	3 (6.3)	1 (2.1)	3 (6.3)	8 (4.2)
001**00001010**	1 (2.1)	2 (4.2)	1 (2.1)	4 (2.1)	
001**00011000**			1 (2.1)		1 (0.5)
001**10000110**	1 (2.1)	1 (2.1)	2 (4.2)	1 (2.1)	5 (2.6)
001**10001000**	5 (10.4)	4 (8.3)	4 (8.3)	2 (4.2)	15 (7.8)
001**10001010**	14 (29.2)	6 (12.5)	9 (18.8)	2 (4.2)	31 (16.1)
001**10011000**	2 (4.2)	5 (10.4)	3 (6.3)	5 (10.4)	15 (7.8)
001**10011010**	8 (16.7)	5 (10.4)	5 (10.4)	9 (18.8)	27 (14.1)
001**11001000**	1 (2.1)				1 (0.5)
001**11001010**	1 (2.1)		2 (4.2)		3 (1.6)
001**11011010**	2 (4.2)		1 (2.1)	2 (4.2)	5 (2.6)
001**11101000**	1 (2.1)	2 (4.2)	2 (4.2)	1 (2.1)	6 (3.1)
001**11101010**	11 (22.9)	20 (41.7)	17 (35.4)	23 (47.9)	71 (37.0)

a. 0 denotes absence of restriction site or lack of deletion; 1 denotes presence of restriction site or deletion. Order of loci: *Hae*III nt 663, *Hpa*I nt 3592, *Alu*I nt 5176, *Alu*I nt 7025, *Dde*I nt 10394, *Alu*I nt 10397, *Hin*fI nt 12308, *Hinc*II nt 13259, *Alu*I nt 13262, *Hae*III nt 16517, 9-bp deletion. Loci found to be polymorphic in these populations are designated in boldface.

lotype frequencies relative to the within-population variation. We constructed a neighbor-joining tree of relationships among the four population groups on the basis of the pairwise D_A distances (Table 4) estimated from haplotype frequencies. This tree is presented in Figure 2. It can be seen that the relationship among the population groups reconstructed from frequencies of mtDNA markers and the autosomal DNA markers are not congruent. The most striking dissimilarity is that in regard to autosomal DNA markers, the Scheduled Castes were clearly very distant from the other castes, but in regard to mtDNA markers they are not very distant from the Khatris and the Jat Sikhs. The Scheduled Caste group is genetically rather distant from the Brahmins. The Brahmins are fairly close to the Jat Sikhs, but less so with the Khatris. The Jat Sikhs and the Khatris are fairly close.

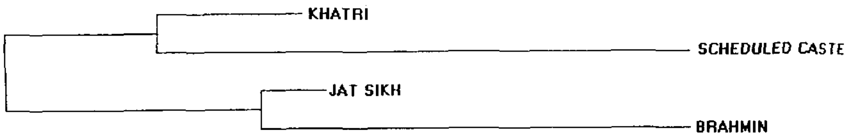

Figure 2. Neighbor-joining tree depicting genetic relationships among four endogamous groups of Punjab based on haplotype frequencies of mtDNA polymorphisms.

Based on the mitochondrial RSP data, we were able to classify individuals into eight haplogroups: Asian and Amerindian haplogroups *A, B, C,* and *D*; east Asian haplogroup *M*; Caucasian haplogroups *U* and *H*; and African haplogroup *L*. The frequencies of the various haplogroups are presented in Figure 3. Haplogroups *A, B, C, D,* and *L* were not observed. Fifty-nine (30.7%) individuals could not be classified into any of the eight haplogroups based on the RSP sites examined. The proportion of individuals with haplogroups other than the eight listed above varied from 50% (Brahmins) to 14.6% (Scheduled Castes). Haplogroup *M* was found to be the most frequent—40.1% of the individuals in the pooled sample belonged to this haplogroup. Although the social ranks of the Jat Sikhs and the Khatris are somewhat controversial, if it is assumed that they occupy a social rank between the Brahmins and the Scheduled Castes, then there is an inverse trend of the frequency of haplogroup *M* with social rank: Brahmins (25%), Jat Sikhs and Khatris (42.7%), and Scheduled Castes (50%). These differences are, however, not statistically significant (χ^2 = 7.17, p = 0.07). No such

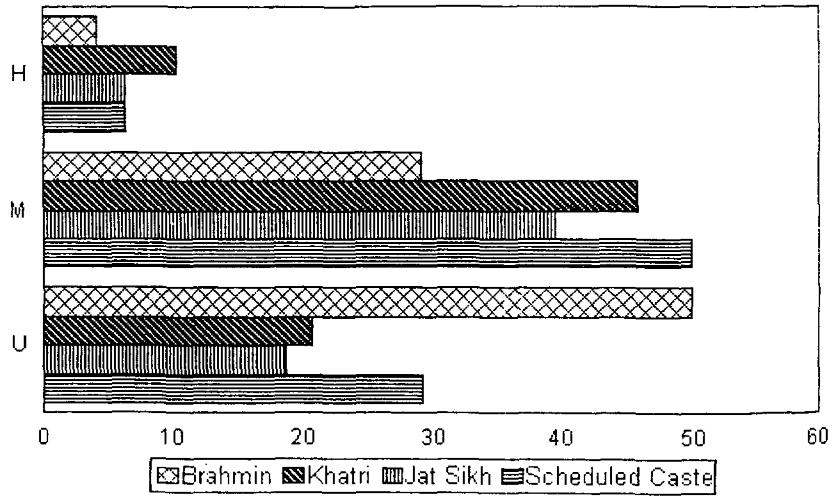

Figure 3. Frequency distributions of various observed mtDNA haplogroups in four endogamous groups of Punjab.

trend is observed in the frequency of haplogroup *U*, which is about 20% among Brahmins, Jat Sikhs, and Khatris, and a little higher (about 30%) among the Scheduled Castes. These differences among ethnic groups are also not statistically significant ($\chi^2 = 1.76$, $p = 0.62$). The remaining observed haplogroup, *H*, occurs sporadically (about 4% to 6%) in all ethnic groups except the Khatris, among whom the frequency is 10.4%.

HVS1 of the mtDNA, comprising 357 nucleotides from positions 16024 to 16380, was sequenced in 57 individuals (12 Brahmins, 15 Khatris, 15 Jat Sikhs, and 15 Scheduled Castes). A summary of the raw sequence data is presented in Table 6. It was observed that 13 sequences (22.8%) were common to individuals within and across the four endogamous groups. The shared sequences were: (KH-64, JT-67), (SC-07, JT-61), (SC-11, JT-66), (JT-52, KH-55), (SC-10, SC-12), (KH-65, BR-59), (SC-09, KH-70), (SC-13, KH-58), (BR-58, BR-62), (BR-66, BR-67), and (BR-64, SC-05, KH-51, JT-53). No striking preponderance of sharing of sequences within or across any two endogamous groups was found. The total observed number of polymorphic sites was 49, of which 19 were singleton sites. The nucleotide diversities and the average number of pairwise nucleotide differences are given in Table 7. It is seen that the Brahmins and the Khatris show the highest nucleotide diversities and average number of pairwise nucleotide differences, followed by the Scheduled Castes and then the Jat Sikhs.

The maximum-likelihood tree depicting relationships among the sequences is given in Figure 4. There is no clear clustering of mtDNA sequences by endogamous group.

Discussion

The *Alu* insertions/deletions are relatively stable polymorphisms, and their ancestral states are known. Allele sharing is due to identity by descent. These properties make them a novel set of highly informative and useful DNA markers for studying human evolution and genetic relationships among various human populations (Batzer et al. 1994). Allele frequencies at all the *Alu* insertion/deletion loci studied, except *CD4*, were found to be highly variable within populations. The *MtNUC* InDel locus was also highly polymorphic in the study populations.

The populations studied exhibited high heterozygosity values for most of these InDel loci. For most loci, the average heterozygosity values in populations were high. Wide variation in the coefficient of gene differentiation (G_{ST}) across loci was noted. The highest G_{ST} value (0.0348) is an order of magnitude higher than the lowest value (0.0027), although the absolute values are all low. The G_{ST} value for all eight loci combined was 0.0134. Studies by Batzer et al. (1996) on the *Alu* loci exhibited G_{ST} values that ranged from 0.039 to 0.132. An earlier study by Batzer et al. (1994) reported slightly higher G_{ST} values (0.097–0.283) for the same set of *Alu* insertion loci. The lower G_{ST} values reported by Batzer et al. in

Table 6. Nucleotides at Various Positions in the HVS1 Segment Drawn from Four Endogamous Groups of Punjab: Brahmin (PBR), Khatri (KHT), Jat Sikh (JSK), and Scheduled Caste (SCH)

	NUCLEOTIDE POSITION – 16000
POPLN.	*00000000000000111111111111111111111111111111111111*
CODE-	*33456678889999000112223344444455556666666777778888*
SL.NO.	*78816951262345248014694602457834682346789236890123*
CRS	*AAGAACTACTTTTCTCCGCTTGCTTCTGCCGTGAAAAACCCTCCTCAAAA*
PBR-055	....
PBR-058	....
PBR-059	....
PBR-060	....
PBR-061	....
PBR-062	....
PBR-063	....
PBR-064	....
PBR-067	..............................A...................
PBR-068	....................C......A.............C........
PBR-069	....
KHT-051	....
KHT-054	....
KHT-055	.....................A............................
KHT-056	....
KHT-057	....................C....................C........
KHT-058	....
KHT-060	.....................A.......T....................
KHT-061	....................C.............................
KHT-064	....
KHT-065	....
KHT-066	.............T....................................
KHT-069	....................C.............................
KHT-070	....
KHT-073	....
KHT-074	....
JSK-051	....
JSK-052	.....................A............................
JSK-053	....
JSK-055	....

JSK-057	. . . .
JSK-058	. . . .-. . . . .
JSK-059	. . . . .A. . . . .
JSK-060	. . . .
JSK-061	. . . . .A. . . . .
JSK-062	. . . .
JSK-063	. . . .
JSK-065	. . . .
JSK-066	. . . . .A. . . . .
JSK-067	. . . .
JSK-068	. . . .
SCH-003	. . . . .T. . . . .
SCH-004	. . . . .C. . . . .
SCH-005	. . . .
SCH-006	. . . . .C. . . . . . . .C. . . . .
SCH-007	. . . . .A. . . . .
SCH-008	. . . .
SCH-009	. . . .
SCH-010	. . . .
SCH-011	. . . . .A. . . . .
SCH-012	. . . .
SCH-013	. . . .
SCH-014	. . . . .C. . . . .
SCH-016	. . . . .A. . . . . .T. . . . . .
SCH-017	. . . . .C. . . . .
SCH-018	. . . . .C. . . . .

Table 6. (Continued)

	1111111112222222222222222222222222222222222222222222222222222222222222333333333333333333333333333333333
	8888889990001111122222233333444444444555555666666677777788888999999999900000011111222222334445555555666
	4567890236793457802345701459012356789046789013456902456846789012345678901245912689012457052342345679028
CRS	*CCCCCTCCCAATGCATCACCTCAATCACAACTCAACTCACCACCCTCACACAGATCACCTACCCACCCTTAACATAATAAAGCCATTCTATACTCCCTTTCTT*
PBR-055	...............C...C.
PBR-058	......................G...G...T......................
PBR-059	...................T..G..
PBR-060	...................TC..C.........................
PBR-061	....................C..C.........................
PBR-062	......................G...G...T......................
PBR-063	....G...T..........G...........
PBR-064	....
PBR-067	....G.............................
PBR-068	..................T.........................T..
PBR-069	...................T..............................T.A..............T.............A...........C.........
KHT-051	....
KHT-054	....C.
KHT-055	...................T...
KHT-056	.T.................T..G..C.
KHT-057	...........................T.....................................T.T..................C................
KHT-058	...................T..............................T..............................A...........C.........
KHT-060	...................T..T..
KHT-061	.T.................T...C.........................
KHT-064	....T.........................C..C...................C.....
KHT-065	...................T..G..
KHT-066	...................T................C..C...
KHT-069	...................T...................................T...
KHT-070	...................T...C.
KHT-073	...................T..................................C.....G..
KHT-074	..........G...G...C......................
JSK-051	...................T................................A..C.
JSK-052	...................T...
JSK-053	....
JSK-055	....................C...............C..C.........................
JSK-057	.......T...............................T..........T..C.
JSK-058	...................T...
JSK-059	...................T...

JSK-060	.................T..T.....................
JSK-061	.................T.......................................T.....C................................
JSK-062	.................T..T.....................................
JSK-063	....T.....................................
JSK-065	...T.............T....C...C.
JSK-066	.................T.......................................T......................................
JSK-067	....T.......................C..C.................C......
JSK-068	....C...........................
SCH-003	.................T...T....................
SCH-004	....
SCH-005	....
SCH-006	.........G..T...........G..T......................
SCH-007	.................T.......................................T.....C................................
SCH-008	.................T..G..T......................
SCH-009	.................T..C.
SCH-010	..T..............T.....................................G..
SCH-011	.................T.......................................T......................................
SCH-012	..T..............T.....................................G..
SCH-013	.................T............................T............................A..........C.........
SCH-014	....G..T......................
SCH-016	.................T...C........................
SCH-017	....C...........................
SCH-018	.................T..

Note: A dot (.) in a particular nucleotide position indicates that the nucleotide is the same as the nucleotide at this position occurring in the Cambridge Reference Sequence (CRS). The CRS is provided in the first line of the table.

Table 7. Observed Nucleotide Diversity (π) and Mean Number of Pairwise Nucleotide Differences (k) in Four Endogamous Groups of Punjab

Population	π	k
Brahmin	0.0138 ± 0.0081	5.07 ± 2.64
Khatri	0.0137 ± 0.0079	5.04 ± 2.59
Jat Sikh	0.0107 ± 0.0063	3.92 ± 2.08
Scheduled Castes	0.0126 ± 0.0073	4.62 ± 2.40
Combined	0.0127 ± 0.0075	4.67 ± 2.32

1996 are probably because of the inclusion of a number of closely related populations of European descent. A study by Novick et al. (1998) on 30 American populations of native and Asian origin on five *Alu* loci observed G_{ST} values in the range of 0.09 to 0.16. They attributed the reduced genetic diversity in these unrelated populations to bottleneck effect. The low overall G_{ST} value (0.0134) observed in the population groups included in the present study may be a reflection of their recent common ancestry or bottleneck effects. The overall G_{ST} value estimated in the present study is also lower than that (0.068) reported by Majumder et al. (1999). This is probably because the populations studied by Majumder et al. were ethnically and geographically more diverse.

Majumder (1998) and Majumder et al. (1999) reported that in Indian populations chosen from widely separated geographical areas, those in closer geographical proximity exhibited closer genomic affinity irrespective of the social rank of the populations. Thus, for example, their studies showed that Brahmins and other upper castes residing along with lower castes in close geographical locations showed greater genomic affinities than caste groups of a similar social rank inhabiting different geographical locales. While the present study cannot provide a validation or refutation of these observations because the populations in the present study are all drawn from a relatively small geographical region, the results of the genetic distance and phylogenetic analyses based on autosomal and mtDNA polymorphisms do not reveal a congruent picture of evolutionary relationships among the populations studied. From Table 4 and the branch lengths separating the populations in the NJ tree based on the autosomal InDel loci (Figure 1), it is clear that the Scheduled Castes and the Brahmins are genetically very dissimilar, which is also what is seen in the tree based on mtDNA polymorphisms (Figure 2). However, while the Scheduled Castes group is also genetically very distant from both the Khatris and the Jat Sikhs based on autosomal DNA data, this feature is not observed from the mtDNA data. Further, while with respect to the autosomal DNA data, the Brahmins, Khatris, and Jat Sikhs appear to form a cluster with respect to mtDNA data, the Khatris are clearly separated from the Brahmins and Jat Sikhs. These findings based on our autosomal DNA data are consistent with those from a previous study by Singh et al. (1974) using red cell enzyme polymorphisms, which were all autosomal loci, in which Brahmins, Khatris, and Jat Sikhs were also found to show no significant allele frequency differ-

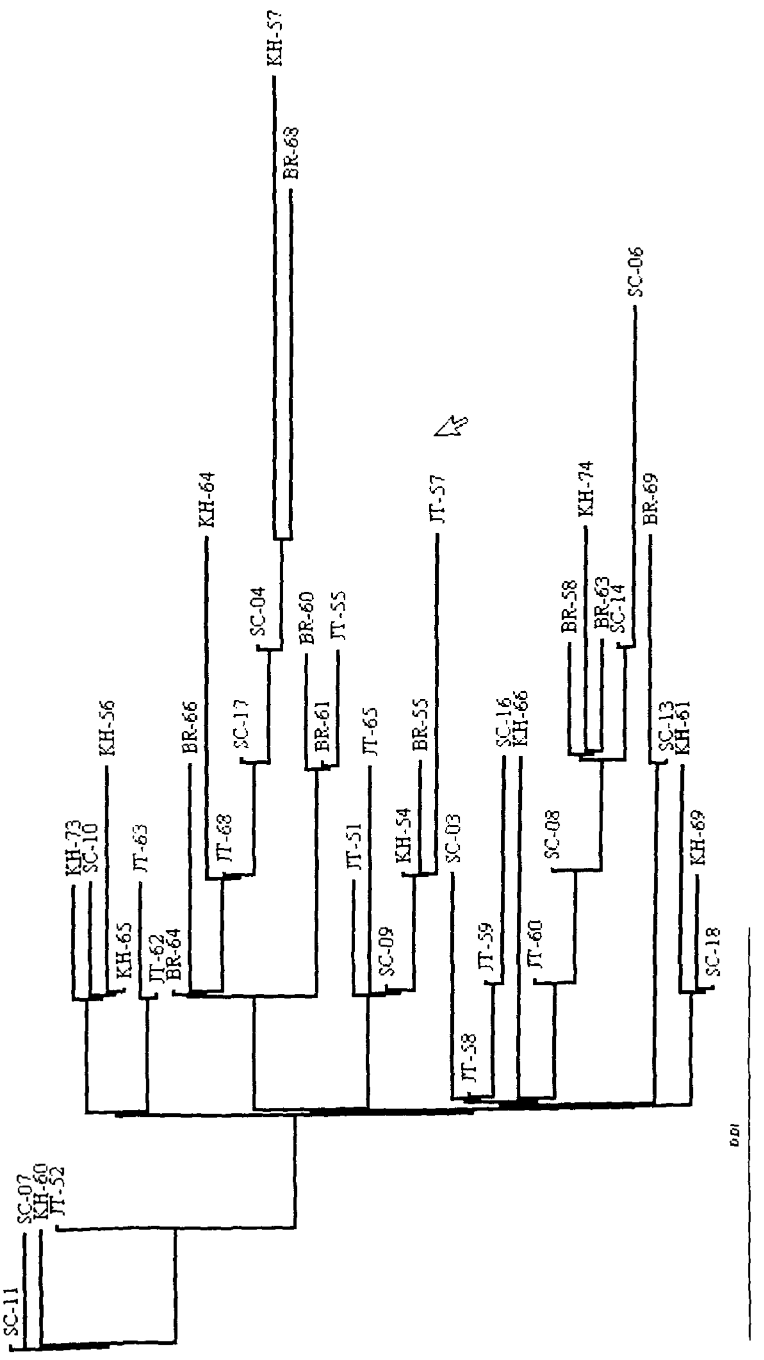

Figure 4. Maximum-likelihood tree depicting relationships among 44 distinct mtDNA HVS1 sequences drawn from four endogamous groups of Punjab. (BR = Brahmin, KH = Khatri, JT = Jat Sikh, SC = Scheduled Castes).

ences. Our findings reveal that the Brahmins and Jat Sikhs may have been founded by a small group of female lineages, but because of waves of migration during the historical period (Thapar 1966), which were possibly predominantly male, various new autosomal genes were differentially introduced in the populations studied. The origin of the Scheduled Castes appears to be quite different, because they stand apart from the other three groups both in respect to mitochondrial and autosomal DNA data. Our interpretation stated above seems reasonable because the haplotypes that are common in the endogamous populations of Punjab are also common among many other populations of India (Roychoudhury et al. 2000, 2001). While this interpretation may imply that there should also be extensive sharing across endogamous groups and population clustering of HVS1 sequences, neither of these is observed in this present study. The most likely reason for these observations is the high mutation rate in the HVS1 region, which can potentially scramble long-term evolutionary history.

The frequency of haplogroup *M* increases with decrease in the social rank of the groups, although the differences in frequencies of this haplogroup among the populations is not statistically significant. The pooled frequency (40%) of this haplogroup is lower than frequencies observed in many caste populations of other parts India, except among the Brahmins of Uttar Pradesh (Roychoudhury et al. 2000). The frequency of this so-called 'Asian specific' haplogroup shows a rough clinal decrease from south to north India. The frequency of the 'European' haplogroup *U* does not show any statistically significant difference across the study populations, but the pooled frequency (22%) is higher than most that of other caste populations in other parts of India, except those of Uttar Pradesh (Roychoudhury et al. 2000). However, it is interesting that haplogroup *H*, which is found in Europe in about 40% to 50% frequency, is present among 10% of Khatris, but at a lower (4% to 6%), but not significantly so, frequency among the other populations. This haplogroup is absent among many tribal populations of India (Roychoudhury et al. 2001), but has earlier been reported among northern Indian caste populations of Punjab and Uttar Pradesh (Passarino et al. 1996). These results agree with the interpretation of Passarino et al. (1996) that there has been a considerable inflow of genes from Indo-European–speaking populations from central, and possibly also from west, Asia into northern India, including Punjab.

The extents of haplotype and nucleotide diversities and average number of pairwise nucleotide differences do not differ significantly across the endogamous groups, although the values of these statistics are highest among the Brahmins. Since mtDNA is maternally inherited, a parsimonious explanation of these features, including the extensive sharing of a small number of haplotypes across the study populations, is that in spite of high levels of admixture in historical times in the endogamous groups of Punjab, there was no significant introduction of maternal lineages subsequent to the formation of these endogamous groups.

While Bamshad et al. (1998, 2001), using data on mitochondrial, Y-chromosomal, and autosomal DNA polymorphisms, found that genetic distances between the upper, middle, and lower Hindu caste groups correlated with their social ranks, we have not found such a clear trend. The results of this study when contrasted with those obtained by Majumder et al. (1999) and Bamshad et al. (1998, 2001) show that there are regional differences in the patterns of correlation between genomic and sociocultural affinities within India.

Acknowledgments This study was supported by Punjab State Council for Science and Technology, Chandigarh, through grant No. PSCST/MFA/2540 to JRS and AJSB, and in part by a grant from the Department of Biotechnology, Government of India, to PPM.

Received 18 March 2002; revision received 17 August 2002.

Literature Cited

Bamshad, M.J., T. Kivisild, W.S. Watkins et al. 2001. Genetic evidence on the origins of Indian caste populations. *Genome Res.* 11:994–1004.

Bamshad, M.J., W.S. Watkins, M.E. Dixon et al. 1998. Female gene flow stratifies Hindu Castes. *Nature* 395:651–652.

Batzer, M.A., S.S. Arcot, J.W. Phinney et al. 1996. Genetic variation of recent *Alu* insertions in human populations. *J. Mol. Evol.* 42:22–29.

Batzer, M.A., and P.L. Deininger. 1991. A human-specific subfamily of *Alu* sequences. *Genomics* 9:481–487.

Batzer, M.A., C.M. Rubin, U. Hellmann-Blumberg et al. 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. *J. Mol. Biol.* 247:418–427.

Batzer, M.A., M. Stoneking, M. Alegria-Hartman et al. 1994. African origin of human-specific polymorphic *Alu* insertions. *Proc. Natl. Acad. Sci., USA* 91:12288–12292.

Cann, R.L. 2001. Genetic clues to dispersal of human populations: Retracing the past from the present. *Science* 291:1742–1748.

Deininger, P.L., and M.A. Batzer. 1993. Evolution of retroposons. *Evol. Biol.* 27:157–196.

Gill, P., J.E. Lygo, S.J. Fowler et al. 1987. An evaluation of DNA fingerprinting for forensic purposes. *Electrophoresis* 8:38–44.

Ibbetson, D. 1916a. The Jat, Rajput and Allied Castes, pp. 97–163; Religious, Professional, Mercantile and Miscellaneous Castes, pp. 214–265; Vagrant, Menial and Artisan Castes, pp. 266–338. *Panjab Castes*. New Delhi, India: Neeraj Publications.

Ibbetson, D. 1916b. *Panjab Castes*. New Delhi, India: Neeraj Publications. Pp. 97–163; 214–265; 266–338.

Kivisild, T., M.J. Bamshad, K. Kaldma et al. 1999. Deep common ancestry of Indian and Western Eurasian mitochondrial DNA lineages. *Curr. Biol.* 9:1331–1334.

Majumder, P.P. 1998. People of India: Biological diversity and affinities. *Evol. Anthropol.* 6:100–110.

Majumder, P.P., B. Roy, S. Banerjee et al. 1999. Human-specific insertion/deletion polymorphism in Indian populations and their possible evolutionary implications. *Eur. J. Hum. Genet.* 7:435–446.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci., USA* 70:3321–3323.

Novick, G.E., C.C. Novick, J. Yunis et al. 1998. Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum. Biol.* 70:23–39.

Passarino, G, O. Semino, L.F. Bernini et al. 1996. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am. J. Hum. Genet.* 59:927–934.

Perna, N.T., M.A. Batzer, P.L. Deininger et al. 1992. *Alu* insertion polymorphism: A new type of marker for human population studies. *Hum. Biol.* 64:641–648.

Roychoudhury, S., S. Roy, A. Basu et al. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* 109:339–350.

Roychoudhury, R., S. Roy, B. Dey et al. 2000. Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Curr. Sci.* 79:1182–1192.

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

Singh, S., K.N. Sareen, and H.W. Goedde. 1974. Investigation of some biochemical genetic markers in four endogamous groups in Punjab (N.W. India). II. Red cell enzyme polymorphisms. *Humangenetik* 22:133–138.

Stoneking, M., J.J. Fontius, S.L. Clifford et al. 1997. *Alu* insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res.* 7:1061–1071.

Thapar, R. 1966. *A History of India.* Vol. 1. Middlesex, UK: Penguin.

Torroni, A., K. Huoponen, P. Francalacci et al. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850.

Torroni, A., T.B. Schurr, M.F. Cabell et al. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563–590.

Vigilant, L.A., A.C. Wilson, and H. Harpending. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.

Zischler, H., H. Geisert, A. von Haeseler et al. 1995. A nuclear ‘fossil’ of the mitochondrial D-loop and the origin of modern humans. *Nature* 378:489–492.