

Negligible Male Gene Flow Across Ethnic Boundaries in India, Revealed by Analysis of Y-Chromosomal DNA Polymorphisms

Nitai Pada Bhattacharyya,¹ Priyadarshi Basu,¹ Madhusudan Das,² Srimanta Pramanik,¹ Rajat Banerjee,² Bidyut Roy,³ Susanta Roychoudhury,² and Partha P. Majumder^{3,4}

¹Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, Calcutta, India; ²Department of Human Genetics, Indian Institute of Chemical Biology, Calcutta, India; ³Anthropology and Human Genetics Unit, Indian Statistical Institute, Calcutta 700 035, India

From the historically prevalent social structure of Indian populations it may be predicted that there has been very little male gene flow across ethnic boundaries. To test this finding, we have analyzed DNA samples of individuals belonging to 10 ethnic groups, speaking Indo-European or Austroasiatic languages and inhabiting the eastern and northern regions of India. Eight Y-chromosomal markers, two biallelic and six microsatellite, were studied. All populations were monomorphic for the deletion allele at the YAP (DYS287) locus and for the 119-bp allele at the DYS288 locus. Y-chromosomal haplotypes were constructed on the basis of one RFLP locus and five microsatellite loci. The haplotype distribution among the groups showed that different ethnic groups harbor nearly disjoint sets of haplotypes. This indicates that there has been virtually no male gene flow among ethnic groups. Analysis of molecular variance revealed that there was significant haplotypic variation between castes and tribes, but nonsignificant variation among ranked caste clusters. Haplotypic variation attributable to differences in geographical regions of habitat was also nonsignificant.

Although several early studies (Jakubiczka et al. 1989; Malaspina et al. 1990; Spurdle and Jenkins 1992; Dorit et al. 1995; Hammer 1995; Whitfield et al. 1995) pointed to a low level of variation in the Y chromosome, it has now been established beyond doubt that there are many Y-chromosomal markers that are highly polymorphic in all global populations (Deka et al. 1996; Ruiz-Linares et al. 1996; Santos and Pena 1996; Hammer et al. 1997; Karafet et al. 1997; Rodriguez-Delfin et al. 1997; Zerjal et al. 1997). Because the Y chromosome, except for its telomeric regions, is transmitted uniparentally (paternally) as a linkage group, it has turned out to be extremely useful in population genetic studies for establishing paternal lineages (Deka et al. 1996). Studies on Y-chromosomal variation permit the interesting possibility of contrasting male-specific histories of populations to female-specific ones, which are revealed by mitochondrial DNA (mtDNA) studies.

Population differentiation with respect to the Y chromosome has been studied in many regions of the world, and India represents one of the most ethnically and genetically diverse regions (Majumder 1998). Socially, the vast majority (~80%) of the Indian population belong to the Hindu religious fold and are orga-

nized into ~2000 caste groups, each of which belongs to a socially ranked (broadly, upper, middle, and lower) caste cluster. The social rank is dependent on occupation, certain beliefs of purity and pollution, and continued settlement in a particular geographical location (Thapar 1992). The tribal populations of India are organized into clan groups; there are ~400 tribes in India. Additionally, there are several religious communities, such as Sikhs, Muslims, Christians, Jews, etc. Marriages between different religious groups are extremely infrequent. The caste structure is also fairly rigid, and each caste remains as an endogamous unit, although the levels of endogamy can vary substantially (Malhotra and Vasulu 1993). The extent of admixture among caste groups of the same social rank is higher than among those belonging to different social ranks. Boundaries of middle caste groups have been the most fluid; these groups have admixed with both upper and lower caste groups. Despite the admixture between caste groups, the genetic implication of the approved social rule of hypergamy, by which a man can marry a woman belonging to a caste of lower social rank and continue to retain his caste affiliation (the woman is absorbed in her husband's caste subsequent to marriage), is that crossings of Y chromosomes across ranked caste-cluster boundaries have been negligible in historical times. We note that the converse union of a woman marrying a man of a lower social status and

retaining her caste affiliation (hypogamy) has been discouraged, historically. In this extremely infrequent type of marriage, the woman moves to the husband's caste, which is of a lower social rank. Although the rules governing marriage, that is, admixture of genes among castes, are clearly delineated, the practices leading to intermixture of genes between castes and tribes or other religious groups have not been consistent. Often such marriages result in social ostracization and excommunication, forcing the spouses to move to other geographical areas; thereafter, they become absorbed by a local group, generally of a low social rank. In view of the interesting social norms governing marriage, it is of considerable interest to examine the degree of differentiation of population groups (including castes, tribes, and other religious groups) of India with respect to male lineages. While the present study was in progress, a similar study, comprising 12 caste groups inhabiting a restricted geographical area (Andhra Pradesh) within India, recently has been completed (Bamshad et al. 1998). This study has indicated a relative lack of male gene flow among castes compared to female gene flow. In this paper we report results of a study conducted among 10 population groups (8 caste and 2 tribal) inhabiting a much wider geographical area (three different states—West Bengal and Orissa in the eastern region and Uttar Pradesh in the northern region—of the Republic of India) and covering two linguistic families (Indo-European and Austroasiatic), with respect to eight Y-chromosomal markers (two bi-allelic and six short-tandem repeat markers). Ethnic descriptions, sampling locations, and sample sizes of

the 10 study populations are given in Table 1. Our study, in addition to permitting estimation of male gene flow across caste boundaries, also permitted estimation of such gene flow across caste and tribal boundaries and has a wider geographical coverage than the study of Bamshad et al. (1998).

RESULTS

All populations were monomorphic for the Y *Alu* polymorphic [YAP – (deletion)] allele at the DYS287 locus and for the 119-bp allele at the DYS288 locus. [Correspondences between repeat numbers and allele sizes (bp) at all STR loci were obtained from Kayser et al. (1997).] The remaining six loci were polymorphic in all populations.

Haplotypes constructed on the basis of data of the six polymorphic loci and their frequencies observed in the study populations are presented in Table 2. The 125 sampled individuals harbored 81 distinct haplotypes, indicating extensive Y-chromosomal diversity. It was also observed that the haplotypes were generally population-specific; that is, the sets of haplotypes observed in the different study populations were largely disjoint. Only 12 (15%) of the 81 distinct haplotypes were shared between populations. Among the northern Indian populations of Uttar Pradesh (Brahmin, Chamar, and Rajput), the total number of distinct haplotypes was 35, of which only 3 (8.6%) were shared among the populations (1 haplotype was shared between the upper caste Brahmin and middle caste Rajput; 2 were shared between Rajput and lower caste Chamar). Among the eastern Indian populations of

Table 1. Names of Study Populations, Locations of Sampling, Ethnological and Linguistic Descriptions, and Sample Sizes

Population name (code)	Ethnic cluster	Sampling region and location	Language	Occupational details	Sample size
Brahmin (BR1)	upper caste	northern region, Lucknow (Uttar Pradesh)	Indo-European	traditionally priests, now various occupations	11
Brahmin (BR2)	upper caste	eastern region, Calcutta (West Bengal)	Indo-European	traditionally priests, now various occupations	16
Chamar (CHA)	lower caste	northern region, Lucknow (Uttar Pradesh)	Indo-European	menial and agricultural laborers	9
Lodha (LOD)	tribal	eastern region, Medinipur (West Bengal)	Austroasiatic	traditionally hunter-gatherers, now mostly agricultural laborers	17
Santal (SAN)	tribal	eastern region, Medinipur (West Bengal)	Austroasiatic	traditionally hunter-gatherers, now mostly agriculturists and agricultural laborers	15
Bagdi (BAG)	lower caste	eastern region, Hooghly (West Bengal)	Indo-European	mostly fishermen and agricultural laborers	10
Mahishya (MAH)	lower caste	eastern region, Hooghly (West Bengal)	Indo-European	mostly agriculturists and agricultural laborers	12
Agharia (AGH)	middle caste	eastern region, Sundergarh (Orissa)	Indo-European	various occupations	8
Rajput (RAJ)	middle caste	northern region, Lucknow (Uttar Pradesh)	Indo-European	various occupations	21
Tanti (TAN)	lower caste	eastern region, Sundergarh (Orissa)	Indo-European	weavers and agricultural laborers	6

Table 2. Y-Chromosomal Haplotypes and Their Frequencies in 10 Ethnic Populations of India

Haplotype No.	chHapl(f)	DYS19	DYS389I	DYS390	DYS391	DYS393	BR1	BR2	C1A	LOD	SAN	BAG	MAH	AGH	RAJ	TAN	Total
1	1	14	10	23	11	12	2										2
2	0	17	10	26	11	14	1										1
3	0	14	10	25	10	13	1										1
4	1	15	10	22	10	12	1										1
5	1	15	10	22	11	12	2										2
6	0	15	10	23	10	15	1										1
7	0	16	10	24	10	15	1					1					2
8	1	14	10	22	11	12	1										1
9	0	16	11	25	10	13	1	3							1		5
10	0	17	11	25	11	13		4									4
11	0	15	10	23	10	14		1	1					1			3
12	0	17	11	24	11	13		1									1
13	0	17	11	25	12	13		1									1
14	0	16	10	25	10	13		1									1
15	0	14	10	23	10	14		2									2
16	0	15	11	23	10	15		1									1
17	0	17	11	25	10	14		1									1
18	1	16	11	21	10	13		1	1								2
19	0	17	10	24	11	14		1									1
20	0	16	10	25	10	14		1			2						3
21	0	17	10	24	10	13		1									1
22	1	16	11	22	10	12		1		1			1		2		6
23	1	17	11	22	10	14		1									1
24	0	16	10	22	10	14		1									1
25	0	16	10	25	11	13		1							1		2
26	1	16	10	23	10	12				5				1			6
27	0	15	10	23	11	15				2							2
28	0	16	11	23	10	14				3							3
29	1	15	9	23	12	14				1							1
30	1	16	10	23	10	13				1							1
31	0	15	11	24	11	15				1							1
32	1	16	10	22	11	12				1							1
33	0	16	10	22	10	13				1				1	1		3
34	1	13	9	23	11	14				1							1
35	0	16	10	24	10	15				1							1
36	1	17	10	25	10	14				1							1
37	1	17	10	25	12	14				1							1
38	1	15	9	23	10	14				1							2
39	1	15	9	23	11	13				1							1
40	1	18	10	25	10	15				1							1
41	1	16	10	24	10	14				1							1
42	1	16	10	23	10	14				2				1			3
43	1	16	10	24	15	14				1							1
44	0	17	9	23	10	13						5					5
45	1	17	9	23	10	13						1					1
46	1	15	10	23	10	12						1					1
47	1	15	11	24	10	14						1					1
48	1	16	11	21	11	14						1					1
49	0	16	10	24	10	13							1	1			2
50	0	16	11	21	10	12							3				3
51	1	15	11	23	10	14							1				1
52	1	16	10	22	10	13							1		1		2
53	1	14	11	25	10	13							1				1
54	0	16	10	21	10	12							1	1			1
55	0	16	10	23	9	14							1				1
56	0	15	11	23	10	14							1				1
57	1	16	10	25	11	12								1			1
58	1	17	10	22	11	12								1			1
59	0	17	10	26	10	14								1			1
60	0	16	10	24	11	12								1			1
61	0	15	11	24	10	13									1		1
62	1	17	9	25	10	12									1		1
63	0	17	10	22	11	13									1		1
64	1	16	11	21	11	12									1		1
65	1	17	10	22	10	13									1		1
66	1	16	11	25	11	13									1		1
67	1	17	9	24	10	13									1		1
68	0	18	10	25	10	12									1		1
69	1	15	10	24	9	12									1		1
70	1	16	10	26	10	14									1		1
71	1	14	9	24	10	12									1		1
72	1	15	9	24	10	12									1		1
73	0	16	10	25	11	12									1		1
74	0	15	10	23	10	12									1		1
75	0	15	10	23	10	13									1		1
76	0	17	10	25	10	13									1		1
77	0	15	10	24	10	14									1		1
78	0	17	10	24	11	13									1		1
79	0	15	11	23	11	14									1		1
80	0	16	11	23	10	13									1		1
81	1	16	11	25	11	14									1		1

West Bengal and Orissa (Brahmin, Agharia, Bagdi, Mahishya, Tanti, Lodha, and Santal), the total number of distinct haplotypes observed was 60, of which only 7 (11.7%) were shared among the populations. The upper caste Brahmin shared one haplotype with middle caste Agharia and another with Santal tribals; the Agharia also shared one other haplotype with lower caste Mahishya and two other haplotypes with Lodha tribals; the Mahishyas shared a haplotype with the Santal tribals; and one haplotype was shared by three groups—the lower caste Mahishya and Tanti and the two tribal groups of Lodha and Santal. Therefore, there was minimal sharing of haplotypes among the ethnic groups studied.

However, because the sample sizes of the individual ethnic populations were small and because our hypothesis largely pertained to ranked caste cluster boundaries, we decided to carry out further analyses by pooling data of ethnic groups belonging to separate ranked caste clusters (upper, middle, and lower) and also of the two tribal groups. The sample sizes of these pooled categories were upper caste = 27, middle caste = 29, lower caste = 37, and tribal cluster = 32. The numbers of distinct haplotypes observed among upper, middle, and lower castes were, respectively, 18, 27, and 30. This number among the tribes was 20. The upper castes shared two haplotypes with the middle castes and three with the lower castes, but none with the tribals. The middle castes additionally shared five haplotypes with lower castes and three with tribals. The lower castes also shared three haplotypes with tribals. We noted that of these shared haplotypes, one haplotype was shared by all three caste clusters and another was shared by the middle and lower caste groups with the tribals. The haplotypes shared by these four clusters

of populations are presented in Table 3. Therefore, even when individual ethnic populations are grouped in ranked clusters, there is very little haplotype sharing among clusters, indicating that there has been minimal gene flow even across these ethnic clusters. It is, however, noteworthy that the upper castes, while sharing haplotypes with middle and lower castes, do not share any haplotype with the tribes.

We sought to examine whether the observed extent of haplotype sharing among these ranked clusters of populations was statistically significantly greater than expected by chance. Because the frequencies of distinct haplotypes vary among clusters, to test this hypothesis, it is not sufficient to take only number of distinct haplotypes shared, but to actually count the numbers of individuals between pairs of clusters that have identical haplotypes. For the six pairs of clusters, the numbers of individuals with identical haplotypes were upper caste and middle caste = 5, upper caste and lower caste = 3, upper caste and tribal cluster = 0, middle caste and lower caste = 8, middle caste and tribal cluster = 11, and lower caste and tribal cluster = 8. Based on 500 simulation runs (details provided in Methods), the upper 95% cutoff values of the distribution of the numbers of individuals who possess identical haplotypes by chance were, for the six pairs of clusters listed above, 14, 17, 16, 19, 16, and 19, respectively. Because all of the observed values are smaller than these cutoff values, it is clear that the observed sharing of haplotypes between clusters of populations is not statistically significant.

The allele frequencies and haplotype diversities are presented in Figures 1 and 2, respectively. Considerable variation in allele frequencies has been observed at many loci among these clusters of populations. For

Table 3. Y-Chromosomal Haplotypes Shared by Various Ethnic Clusters of India

Haplotype no. (as in Table 2)	α <i>HindIII</i>	DYS19	DYS389I	DYS390	DYS391	DYS393	Haplotype shared by
7	0	16	10	24	10	13	upper and lower caste
9	0	16	11	25	10	13	upper and middle caste
11	0	15	10	23	10	14	upper, middle, and lower caste
18	1	16	11	23	10	12	upper and lower caste
20	0	16	10	25	10	14	middle and lower caste, tribal
22	1	16	11	22	10	12	lower caste and tribal
25	0	16	10	25	11	13	middle and lower caste
26	1	16	10	23	10	12	middle caste and tribal
32	1	16	10	22	11	12	middle caste and tribal
42	1	16	10	25	10	14	lower caste and tribal
46	1	15	10	23	10	12	middle and lower caste
49	0	16	10	24	10	12	middle and lower caste
52	1	16	10	22	10	12	middle and lower caste
46	1	15	10	23	10	12	middle and lower caste
74	0	15	10	23	10	12	middle and lower caste

Haplotypes 46 and 74 have the same microsatellite haplotype (15-10-23-10-12) but are on α *HindIII*+ and α *HindIII*− backgrounds, respectively. Thus, this sharing is based on microsatellite haplotype only (i.e., ignoring the difference at the RFLP site).

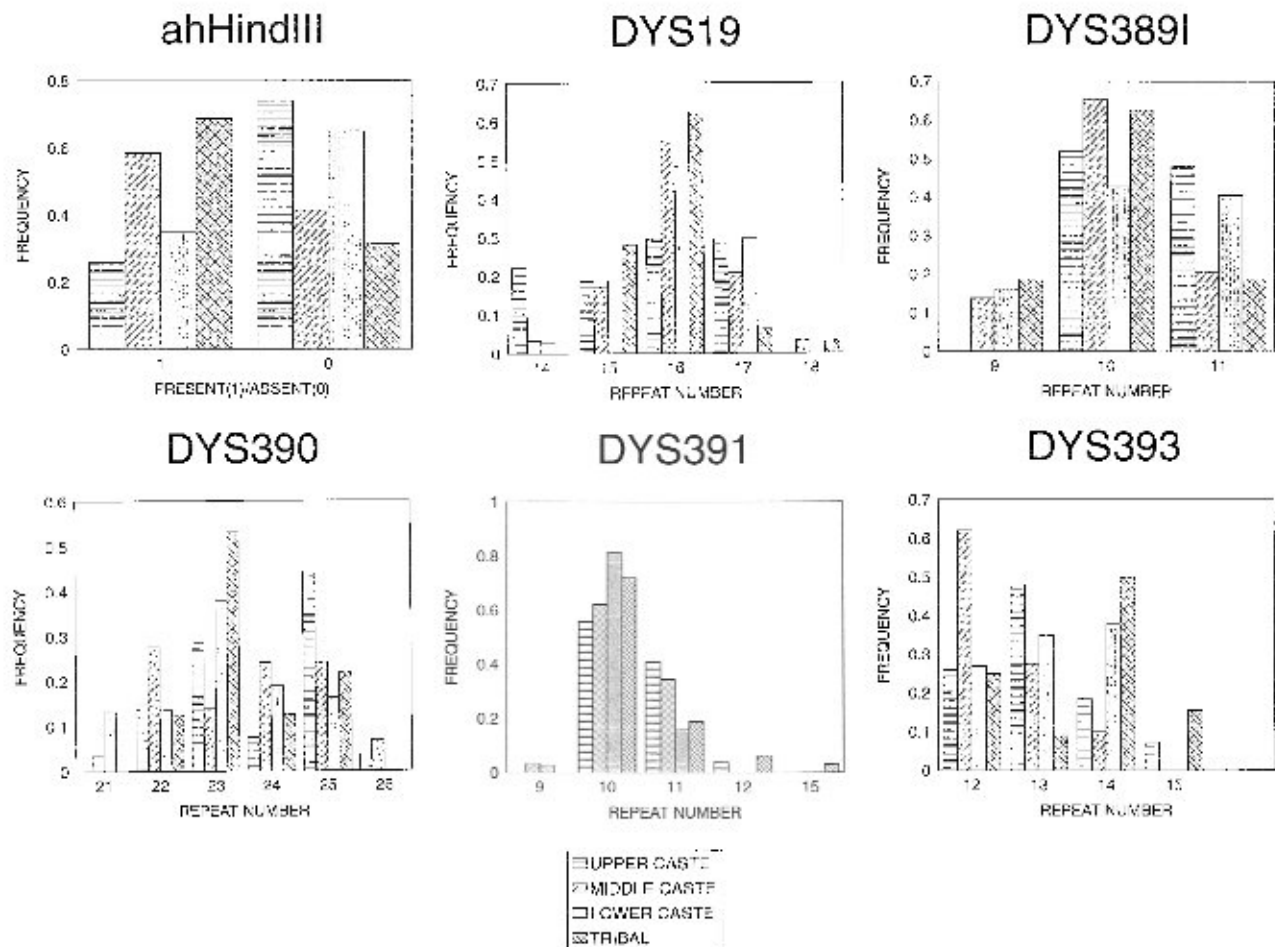


Figure 1 Allele frequency distributions at six polymorphic Y-chromosomal loci in four ethnic clusters of India.

example, at DYS391, not all alleles are observed in all the clusters. Although all clusters harbor very high levels of haplotype diversity (range, 0.9658–0.9975), the highest level of haplotype diversity (0.9975 ± 0.0099) is observed among middle caste populations.

We have performed analyses of molecular variance (Excoffier et al. 1992) to quantitatively establish that there is much greater haplotypic variation within clusters of populations than between them. Because the analysis of molecular variance (AMOVA) takes into account not only the relative frequencies of haplotypes but also the number of mutational steps separating pairs of haplotypes, we excluded the *ahHindIII* RFLP locus and analyzed haplotypes defined by the five polymorphic microsatellite loci only. The individual ethnic populations were suitably grouped to enable examination of the effect on linguistic, geographical, and social rank on Y-chromosomal haplotype differentiation.

It may be noted from Table 1 that there is a complete confounding of linguistic and caste/tribal affiliations of populations. All caste populations speak lan-

guages that belong to the Indo-European family, whereas both tribal populations are Austroasiatic speakers. AMOVA results showed that although 95.75% of the Y-chromosomal microsatellite haplotypic variation was within populations belonging to the two language families (or within populations belonging to caste or tribal clusters), the extent of varia-

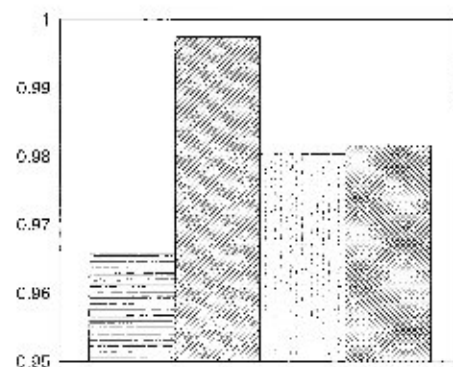


Figure 2 Y-chromosomal haplotype diversities in four ethnic clusters of India.

tion between these groups of populations was 4.25% [$F(ST) = 0.0425$]. This $F(ST)$ value was statistically significant, indicating that significant Y-chromosomal structuring due to differences in language, or analogously, for the present data set, due to caste-tribal differences. When we subdivided our data further by ranked caste categories (i.e., when the caste populations were grouped into the three separate ranked clusters), the additional variance thus explained was not statistically significant.

To examine whether Y-chromosomal variation was significantly structured because of differences in geographical locations of habitat of the populations, we grouped the populations as "northern Indian inhabitants" and "eastern Indian inhabitants". Only 8 (13.3%) of the 60 distinct haplotypes (defined by five polymorphic microsatellite loci) were shared between populations inhabiting these two geographical regions. (It may be noted that of the 81 observed haplotypes defined on the basis of six polymorphic loci, several haplotypes defined by five microsatellite loci occurred on both α HindIII+ and α HindIII- backgrounds; hence, the number of distinct haplotypes dropped from 81 to 60.) AMOVA results indicated that haplotypic variation attributable to geographical differences was not statistically significant.

Because microsatellite loci are known to have higher mutation rates than biallelic RFLP loci, we sought to examine whether insights into the evolutionary histories of these populations can be obtained by examining variations at the microsatellite loci separately for chromosomes that possess the α HindIII restriction site and for chromosomes that do not possess this site. In the pooled set of chromosomes from all populations, there were 59 (47.2%) chromosomes that possessed the HindIII site and 66 (52.8%) that did not. This difference was not statistically significant at the 5% level. Although these two groups of chromosomes shared only 5 (8.33%) haplotypes out of 60 distinct haplotypes (defined by the 5 polymorphic microsatellite loci), the ranges and frequency distributions of repeat numbers at the microsatellite loci between these two groups of chromosomes were, however, strikingly similar (Table 4). The variances among individuals of repeat numbers at the microsatellite loci in these two groups of chromosomes (α HindIII+, α HindIII-) were DYS19 = (0.74, 0.72), DYS389I = (0.46, 0.36), DYS390 = (1.45, 1.48), DYS391 = (0.69, 0.26), and DYS393 = (0.90, 0.72). Therefore, except for the DYS391 locus at which the α HindIII+ chromosomes showed greater variability, the extent of variability at all the other microsatellite loci were similar in both groups of chromosomes. Furthermore, the numbers of distinct microsatellite haplotypes in α HindIII+ and α HindIII- chromosomes were 39 and 42, respectively. The haplotype diversities in these two groups

Table 4. Allele Frequencies at Five Y-Chromosomal Microsatellite Loci for Chromosomes on α HindIII+ and - Backgrounds

Locus	Allele (no. of repeats)	α HindIII+ chromosomes (n = 59)	α HindIII- chromosomes (n = 66)
DYS19	14	0.084 ± .037	0.046 ± .026
	15	0.237 ± .056	0.182 ± .048
	16	0.525 ± .066	0.470 ± .062
	17	0.136 ± .045	0.288 ± .056
	18	0.017 ± .017	0.015 ± .015
DYS389I	9	0.186 ± .051	0.076 ± .033
	10	0.542 ± .058	0.561 ± .062
	11	0.271 ± .065	0.364 ± .060
DYS390	21	0.034 ± .024	0.061 ± .030
	22	0.305 ± .060	0.046 ± .026
	23	0.339 ± .062	0.349 ± .059
	24	0.119 ± .042	0.197 ± .049
	25	0.186 ± .051	0.318 ± .058
DYS391	26	0.017 ± .017	0.030 ± .021
	9	0.017 ± .017	0.015 ± .015
	10	0.661 ± .062	0.712 ± .056
	11	0.271 ± .058	0.258 ± .054
DYS393	12	0.034 ± .024	0.015 ± .015
	15	0.017 ± .017	
	12	0.559 ± .065	0.151 ± .044
	13	0.119 ± .042	0.454 ± .062
	14	0.305 ± .060	0.303 ± .057
	15	0.017 ± .017	0.090 ± .044

were, respectively, 0.97 ± 0.01 and 0.98 ± 0.01 . Therefore, it appears that the antiquities of both of these groups of chromosomes are roughly equal. All populations harbor microsatellite haplotypes on both α HindIII+ and α HindIII- backgrounds. Differentiation of the populations, therefore, seems to have taken place after this locus became polymorphic.

We have also examined the relationships among the haplotypes defined by the five microsatellite loci. The haplotype tree (not shown) comprised four major clusters of haplotypes; one haplotype (Table 1, haplotype 43) formed a single-point cluster. [This haplotype, observed among the tribal Santals, possessed a 15-repeat allele at the DYS391 locus, which was not observed in any other population.] Contrary to our expectations, however, the clusters of haplotypes did not correspond to the ethnic clusters.

DISCUSSION

The prevailing social customs of hypergamy and hypogamy in India restrict male gene flow among ethnic groups of India. We have tested this prediction, using several biallelic and microsatellite Y-chromosomal DNA markers.

The YAP element, which is found in varying frequencies in most global populations, is absent in all the populations included in the present study. The YAP+ frequency is very high among most African groups and low among European populations (see

Table 7 of Passarino et al. 1998). The absence of this element among Indian populations confirms that Indians show relatively more genetic similarities with the Caucasoids than with the Negroids (Majumder 1998). It is, however, noteworthy that our earlier studies have revealed that the Austroasiatic tribal populations of India have some of the human-specific *Alu* elements in the nuclear genome at frequencies that are similar to those found in many African populations (Majumder et al. 1999). The ranges of repeat numbers and allele frequencies at the polymorphic microsatellite loci in the study populations are consistent with global estimates (Kayser et al. 1997). At most of these loci, the most frequent allele is not the same across populations. This may indicate different origins of the study populations or may be due to effects of genetic drift. The locus DYS288 was found to be monomorphic. Comparable data at this locus are not available from many other populations (Kayser et al. 1997).

High levels of haplotype diversity were noted in all clusters of populations. Consistent with our earlier findings, based on serum protein and enzyme polymorphisms, the ethnic populations of India harbor higher levels of genetic diversity than most comparable global regions (Majumder 1998). The highest level of haplotype diversity was found among the middle castes. This finding is not unexpected in view of the fact that the social boundaries of the middle caste groups have, historically, been the most fluid.

Nearly disjoint sets of haplotypes were found among the study populations. Because the effective population size with respect to the Y chromosome is only one-quarter of the autosomal effective population size, this phenomenon may be due to drift effects but is consistent with the prevailing norms governing marriage that severely restrict male gene flow across ethnic boundaries. In the caste hierarchy, the middle caste groups are expected to be the most fluid genetically. The data on haplotype sharing presented in this paper are largely consistent with this expectation. However, although within a restricted geographical region (eastern or northern India in the present study) the caste groups belonging to the upper social rank do not share haplotypes with groups belonging to the lower social rank, there is such sharing of haplotypes across geographical regions. This may indicate that when there are unions, within or outside of marriage, between an upper caste man and a lower caste woman inhabiting a geographical area, there is a tendency for them to move away to distant geographical areas and then affiliate with a lower, not upper, caste in the new location. The fact that tribal clusters share haplotypes with middle and lower castes, but not with upper castes, is also interesting. There are documented instances of tribal groups that after relinquishing the hunter-gatherer life style and adopting agriculture, were con-

verted to castes—mostly lower castes (Bose 1953; Mandelbaum 1970). Sharing of haplotypes among clusters of populations was, however, not significantly higher than chance expectation.

The extent of molecular variance attributable to differences among socially ranked clusters was found to be statistically nonsignificant. This is striking and indicates that Y-chromosomal variation is not structured by social rank, consistent with the anthropological finding that there has been social mobility and variations in ranks of ethnic groups in India (Thapar 1992). However, our analysis has revealed that the extent of molecular variation at the Y-chromosomal microsatellite loci between castes, who are all Indo-European speakers in our sample, and tribes, who are all Austroasiatic speakers in our sample, is significant. No such significant variation was observed between geographical regions.

The comparison of chromosomes on α *Hind*III+ and α *Hind*III– backgrounds has also revealed an interesting aspect of the population differentiation in India. If one makes the reasonable assumption that the loss of a restriction site is more probable than its gain, it is clear that the α *Hind*III site loss was a very ancient mutation and that the two alleles at this locus had reached nearly equal frequencies before the differentiation of the study populations into separate ethnic groups. The finding that the most frequent alleles at all microsatellite loci, except at DYS393, are the same on chromosomes with α *Hind*III+ and α *Hind*III– backgrounds further corroborates this view. However, the fact that the clustering of haplotypes did not correspond to the ethnic clusters, is contrary to the simple expectation arising from the finding of highly disjoint sets of haplotypes among the populations. We are unable to provide a clear explanation of this lack of correspondence. We hypothesize that there was large Y-chromosomal haplotype diversity even before the people of India became organized into distinct social groups and that each social group was formed by a restricted number of male lineages. Collection of data on more biallelic loci and further study variation at microsatellite loci on the multiallelic locus backgrounds in these populations will contribute to a deeper understanding of the population history of India.

METHODS

Populations

One hundred twenty-five males belonging to 10 ethnic populations of India were studied. Their anthropological details and sample sizes are presented in Table 1. All individuals were not related at least at the first-cousin level.

DNA Isolation and Genotyping

From each selected individual, 5–10 ml of blood was drawn

with consent. DNA was isolated following the protocol of Miller et al. (1988). PCR primers and conditions used for screening DYS19, DYS287, DYS288, DYS389I, DYS390, DYS391, and DYS393 loci were essentially the same as those given in Jobling and Tyler-Smith (1995). However, for the DYS19 and DYS391 loci, PCR-amplified products were run on 6% sequencing gels, transferred to Hybond N+ membranes, probed with one of the end-labeled primers, blotted, and autoradiographed. Band sizes were determined by comparing against locus-specific allelic ladders. For the DYS288, DYS389I, DYS390, and DYS393 loci, product sizes were determined from electrophoretograms, using GeneScan Analysis version 2.02 in an ABI-377 automated DNA sequencer. The α HindIII site was screened, using the primers and protocols given in Santos et al. (1995).

Statistical Analysis

For estimating haplotype diversities and performing AMOVA, Arlequin version 1.1 (Schneider et al. 1997) was used. Significance of variance components was tested, using the nonparametric permutation procedure approach described in Excoffier et al. (1992).

For testing the significance of the observed numbers of shared haplotypes among individuals between pairs of the four ranked ethnic clusters, we carried out a permutation test as follows. The total number of sampled individuals in our study was 125; the numbers of individuals belonging to the four ranked clusters—upper caste, middle caste, lower caste, and tribal clusters—were, respectively, 18, 27, 30, and 20. We first randomly permuted the haplotype data of the 125 individuals and then partitioned the data into four subsets of sizes 18, 27, 30, and 20, respectively. To obtain the number of individuals who shared haplotypes between any two subsets I and J ($I < J$; $I, J = 1, 2, 3, 4$), we compared individuals i and j ($i \in I$, $j \in J$; $i < j$) and checked whether they had the same haplotypes. For all pairs of the four subsets, numbers of shared haplotypes were counted. This procedure was repeated 500 times. The upper 95% cutoff point of the frequency distribution (based on the 500 replications) of the number of haplotypes shared between subsets I and J was then calculated. If the actual number of shared haplotypes between the corresponding pair of clusters was less than this cutoff point, then the observed number of shared haplotypes was declared to be statistically nonsignificant at the 5% level.

For constructing a tree of observed haplotypes defined by the five polymorphic microsatellite loci, we computed pairwise distances between haplotypes, using the squared Euclidean distance measure, and then the UPGMA and neighbor-joining clustering algorithms.

ACKNOWLEDGMENTS

This work was supported by a grant from the Department of Biotechnology, Government of India. We are grateful to Badal Dey, Monami Roy, Madan Chakraborty, R.S. Balgir, and B.P. Dash for their participation in fieldwork for collection of samples. We are also grateful to Chris Tyler-Smith and Andres Ruiz-Linares for information and advice, and to Lynn Jorde and W. Scott Watkins for contributing some labeled primers for initiation of this work. Suggestions provided by an anonymous reviewer were extremely helpful.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be

hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bamshad, M.J., W.S. Watkins, M.E. Dixon, L.B. Jorde, B.B. Rao, J.M. Naidu, B.V.R. Prasad, A. Rasanayagam, and M.F. Hammer. 1998. Female gene flow stratifies Hindu castes. *Nature* **395**: 851–852.
- Bose, N.K. 1953. The Hindu method of tribal absorption. In *Cultural anthropology and other essays*. Calcutta, India.
- Deka, R., L. Jin, L.M. Yu, N. Saha, R. Barrantes, R. Chakraborty, and R.E. Ferrell. 1996. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res.* **6**: 1177–1184.
- Dorit, R., H. Akashi, and W. Gilbert. 1995. Absence of polymorphism at the ZFY locus on the human Y-chromosome. *Science* **268**: 1183.
- Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Hammer, M.F. 1995. A recent common ancestry for human Y chromosomes. *Nature* **378**: 376–378.
- Hammer, M.F., A.B. Spurdle, T. Karafet, M.R. Bonner, E.T. Wood, A. Novelletto, and P. Malaspina. 1997. The geographic distribution of human Y chromosome variation. *Genetics* **145**: 787–805.
- Jacubiczka, S., J. Arnemann, H.J. Cooke, M. Krawczak, and J. Schmidtke. 1989. A search for restriction fragment length polymorphisms on the human Y chromosome. *Hum. Genet.* **84**: 86–88.
- Jobling, M.A. and C. Tyler-Smith. 1995. Fathers and sons: The Y chromosome and human evolution. *Trends Genet.* **11**: 449–456.
- Karafet, T., S.L. Zegura, J. Vuturo-Brady, O. Posukh, L. Osipova, V. Wiebe, and F. Romero. 1997. Y-chromosome markers and trans-Bering Strait dispersals. *Am. J. Phys. Anthropol.* **102**: 301–314.
- Kayser, M., A. Caglia, D. Corach, N. Fretwell, C. Gehrige, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, et al. 1997. Evaluation of Y-chromosomal STRs: A multicenter study. *Int. J. Legal Med.* **110**: 125–133.
- Majumder, P.P. 1998. People of India: Biological diversity and affinities. *Evol. Anthropol.* **6**: 100–110.
- Majumder, P.P., B. Roy, S. Banerjee, M. Chakraborty, B. Dey, N. Mukherjee, M. Roy, P. Guha Thakurta, and S.K. Sil. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *European J. Hum. Genet.* **7**: 435–446.
- Malaspina, P., F. Persichetti, A. Novelletto, C. Iodice, L. Terrenato, J. Wolfe, M. Ferraro, and G. Prantera. 1990. The human Y-chromosome shows a low level of DNA polymorphism. *Ann. Hum. Genet.* **54**: 297–305.
- Malhotra, K.C. and T.S. Vasulu. 1993. Structure of human populations in India. In *Human population genetics: A centennial tribute to J.B.S. Haldane* (ed. P.P. Majumder). Plenum Press, New York, NY.
- Mandelbaum, D. 1970. *Society in India*. University of California Press, Berkeley, CA.
- Miller, S.A., D.D. Dykes, and H.F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**: 1215.
- Passarino, G., O. Semino, L. Quintana-Murci, L. Excoffier, M. Hammer, and A.S. Santachiari-Benerecetti. 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am. J. Hum. Genet.* **62**: 420–434.
- Rodriguez-Delfin, L., S.E.B. Santos, and M.A. Zago. 1997. Diversity of human Y-chromosome of South American Amerindians: A comparison with Blacks, Whites, and Japanese from Brazil. *Ann. Hum. Genet.* **61**: 439–448.
- Ruiz-Linares, A., K. Nayar, D.B. Goldstein, J.M. Hebert, M. Seielstad, P.A. Underhill, A.A. Lin, M.W. Feldman, and L.L. Cavalli-Sforza.

1996. Geographic clustering of human Y-chromosomal haplotypes. *Am. Hum. Genet.* **60**: 401–408.
- Santos, F.R. and S.D.J. Pena. 1996. Worldwide distribution of human Y chromosome haplotypes. *Genome Res.* **6**: 601–611.
- Santos, F.R., S.D.J. Pena, and C. Tyler-Smith. 1995. PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. *Gene* **165**: 191–198.
- Schneider, S., J.M. Kueffer, D. Roessli, and L. Excoffier. 1997. *Arlequin, v. 1.1: A software for population genetic data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Spurdle, A.B. and T. Jenkins. 1992. The search for Y chromosome polymorphism is extended to negroids. *Hum. Mol. Gen.* **1**: 169–170.
- Thapar, R. 1992. *Interpreting early India*. Oxford University Press, Delhi, India.
- Whitfield, L.S., J.E. Sulston, and P.N. Goodfellow. 1995. Sequence variation of the human Y chromosome. *Nature* **378**: 379–380.
- Zerjal, T., B. Dashnyam, A. Pandya, M. Kayser, L. Roewer, F.R. Santos, W. Schiefenhover, N. Fretwell, M.A. Jobling, S. Harihara, et al. 1997. Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**: 1174–1183.