Microsatellite Diversity in Andhra Pradesh, India: Genetic Stratification Versus Social Stratification

B. MOHAN REDDY,[1] V. M. NAIDU,[1] V. KOMAL MADHAVI,[1] K. THANGARAJ,[2] VIKRANT KUMAR,[1] B. T. LANGSTIEH,[3] P. VENKATRAMANA,[4] A. G. REDDY,[2] AND LALJI SINGH[2]

Abstract DNA samples of 948 individuals belonging to 27 populations from southern Andhra Pradesh were analyzed for nine AmpF*l*STR Profiler Plus loci. The nature and extent of genomic diversity within and between these populations have been examined with reference to socioeconomic and geographic affiliations. The results suggest that the average heterozygosity is uniformly high in these populations (>0.80) and that the patterns of allele distributions are similar across the populations. The value of the coefficient of gene differentiation and the AMOVA and structure analysis results suggest that these populations are highly homogeneous. The neighbor-joining tree constructed using either D_A or F_{ST} distances suggests no intelligible pattern of population clusters based on ethnohistoric or geographic affiliations. All these observations suggest either a common recent origin of these populations or extensive gene flow across the populations that erased the original genetic differences. Given strict endogamy, the latter explanation can hold only if there has been unauthorized or unrecognized gene flow transecting the social boundaries. Nevertheless, the regression plot of average heterozygosity versus distance from the centroid (R_{ii}), based on Harpending and Ward's (1982) model, and the genetic distances computed between different hierarchical groups within Andhra Pradesh tend to support this conjecture. Overall, the results suggest lack of a significant degree of genetic stratification that is consistent with social stratification in Andhra Pradesh. Furthermore, the neighbor-joining tree based on comparative data from other Indian and continental populations brings out a single and compact cluster of all the Andhra populations that is clearly separated from the rest.

For historical reasons India presents enormous cultural, linguistic, and biological diversity (Karve 1961; Thapar 1966, 1995, 2003; Kosambi 1991; Gadgil et al.

[1]Biological Anthropology Unit, Indian Statistical Institute, Street No. 8, Habsiguda, Hyderabad 500 007, India.
[2]Centre for Cellular and Molecular Biology, Hyderabad, India.
[3]Department of Anthropology, Northeastern Hill University, Shillong, India.
[4]Department of Anthropology, S. V. University, Tirupati, India.

Human Biology, December 2005, v. 77, no. 6, pp. 803–823.

KEY WORDS: LOCAL VS. GLOBAL VARIATION, MICROSATELLITE DIVERSITY, SHORT TANDEM REPEATS, D3S1358, D8S1179, D5S818, VWA, D21S11, D13S317, FGA, D7S820, D18S5, ENDOGAMY, GENE FLOW, HOMOGENIZATION, INDIA.

1998; Krishnan and Reddy 1994; Majumder 1998; Reddy et al. 2004). Indian population structure is unique and characterized by the division of the population into strictly defined hierarchical castes, tribes, and religious groups, each of which are endogamous. By and large, these features of population structure apply to each linguistic region of the country, and languages form strong barriers to gene flow between such regions, even among members of apparently the same caste. It is therefore likely that populations that have the same caste identity but reside in extreme regions have had virtually no marital or cultural interaction between them and that populations from each such region might have evolved independently of populations from other regions.

The state of Andhra Pradesh, with its 70 million people organized into several endogamous castes, tribes, and religious groups, presents enormous variety in its populations, sociocultural patterns, and organization. Several anthropological investigations suggest that the populations of Andhra Pradesh that practice close consanguineous marriages prefer village endogamy and restrict marriage contacts to small distances and hence are highly inbred. This has probably led to a reduction in effective population size, creating breeding isolates within apparently single endogamous castes and subcastes (Reddy 2002).

These features of the Andhra Pradesh population structure are also unique and are restricted to each zone or region within the same linguistic state, such as Andhra Pradesh, which covers a large geographic area. The castes of different hierarchies from within each region have overlapping and/or contiguous distributions and interact closely in a symbiotic relationship with physical possibilities (historical or current) for exchange of genes; hence they provide an ideal study framework within which to test certain anthropological hypotheses, such as the alleged practice of hypergamy, and to study the patterns of gene flow among the hierarchical groups. These finer aspects of Indian population structure have not been hitherto adequately incorporated into the framework of earlier studies on the origin of Indian castes and the impact of Indian population structure on its genetic structure (Bamshad et al. 1996, 1998, 2001; Watkins et al. 1999; Clark et al. 2000; Kivisild et al. 1999a, 1999b, 2003; Cordaux et al. 2003, 2004; Basu et al. 2003).

We present here the results based on 9 autosomal short tandem repeat (STR) loci among the 27 populations, including 3 tribes inhabiting the southern region of Andhra Pradesh. In particular, we study the extent of genomic diversity and affinities between different hierarchical groups at the regional level in a linguistic area and examine whether the signatures of social stratification implicit among the populations are apparent in the observed genomic patterns. On the basis of published data, we also compare the affinities of the populations of Andhra Pradesh to those from other regions of India and elsewhere.

Materials and Methods

Intravenous blood samples (about 5 ml each) were collected, with the informed consent, from 948 individuals belonging to 27 endogamous groups distributed in the contiguous areas of the 6 southernmost districts of Andhra Pradesh

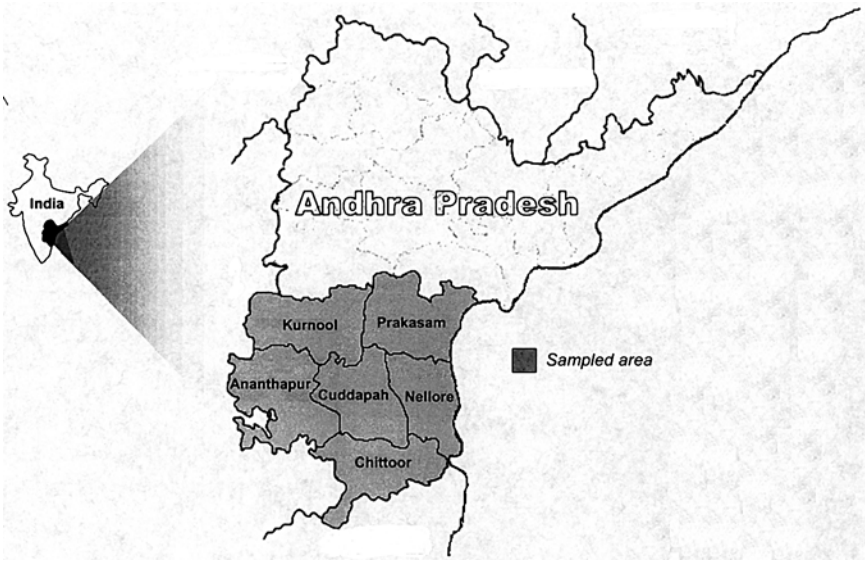

Figure 1. Map of Andhra Pradesh showing the study area and the six southern districts where the blood samples for the 27 populations were drawn.

(Chittoor, Cuddapah, Ananthapur, Kurnool, Prakasam, and Nellore) (Figure 1). The samples were drawn mostly from high school and college students, who represent a large number of surrounding villages and populations, and were supplemented with samples collected from the villages, particularly for tribes and lower castes. All the samples were anonymized and have no individual specific identity. The names of the populations studied, the number of samples drawn from each population, and their socioeconomic, geographic, and occupational backgrounds are furnished in Table 1. Although this study area can be considered culturally, linguistically, and geographically homogeneous, it is inhabited by a wide array of caste and tribal populations, representing almost the entire spectrum of socioeconomic variation in the state.

DNA was isolated from the samples following standard protocols (Sambrook et al. 1989). The AmpF*l*STR Profiler Plus kit (ABI, Foster City, California) was used to co-amplify nine STR loci (D3S1358, D8S1179, D5S818, VWA, D21S11, D13S317, FGA, D7S820, and D18S5) at a time, using an ABI Prism 377 automated DNA sequencer with the GeneScan and Genotyper software packages (Perkin Elmer, Wellesley, Massachusetts) to obtain the allele designations. All these loci are tetranucleotide repeats, and they are located on different chromosomes. To distinguish the alleles of different loci that have similar repeat lengths, we labeled each of the primers with one of the fluorescent dyes (5FAM, NED, and JOE), and the loci were amplified using a multiplex PCR.

Table 1. Studied Populations and Their Sample Size, Location, Occupation, and Socioeconomic Status

Population	*Sample Size*[a]	*Sample Locations (Districts)*	*Traditional Occupation(s)*	*Socioeconomic Status (Caste or Tribe)*[b]
Brahmin	56	Chittoor, Cuddapah, Kurnool	Service and religious cores	Upper
Kshatriya	84	Chittoor, Cuddapah,	Warriors	Upper
Vysya	40	Chittoor, Cuddapah, Nellore	Trade and business	Upper
Akuthota	58	Chittor (Chandragiri and Tirupati)	Land-owning agriculturists	Upper-middle
Kamma	98	Chittoor, Cuddapah	Land-owning agriculturists	Upper-middle
Kapu	40	Chittoor, Cuddapah, Kurnool	Land-owning agriculturists	Upper-middle
Pokanati	114	Chittoor, Cuddapah	Land-owning agriculturists	Upper-middle
Panta	82	Nellore	Land-owning agriculturists	Upper-middle
Vanne	64	Chittoor	Agriculturists	Upper-middle
Balija	76	Chittoor, Cuddapah, Kurnool	Traders and agriculturists	Lower-middle1
Ekila	62	Chittoor	Shepherds, agriculturists	Lower-middle1
Kurava	66	Chittoor, Cuddapah	Shepherds, agriculturists	Lower-middle1
Thogata	58	Chittoor, Cuddapah	Weavers and agriculturists	Lower-middle1
Yadava	52	Chittoor, Cuddapah, Kurnool, Nellore	Shepherds	Lower-middle1
Ediga	58	Chittoor, Cuddapah	Toddy tappers	Lower-middle2
Gandla	34	Chittoor, Cuddapah	Oil pressers	Lower-middle2
Jangam	48	Chittoor	Trade in pearls, beads, etc.	Lower-middle2
Chakali	42	Chittoor, Cuddapah	Washermen	Lower1
Mangali	38	Chittoor, Cuddapah, Kurnool	Barbers	Lower1
Vaddi	80	Chittoor, Cuddapah	Diggers and stone workers	Lower1
Madiga	94	Chittoor, Cuddapah, Kurnool, Anantapur	Scavengers and leather workers	Lower2/Pancham
Mala	190	Chittoor, Cuddapah, Kurnool, Ananthapur, Nellore	Scavengers, grave diggers, and menial workers	Lower2/Pancham
Erukala	62	Chittoor, Nellore, Ananthapur	Nomadic, basket weaving, hunter gatherers, pig herders	Proto-Australoid tribe
Sugali	116	Chittoor, Cuddapah, Ananthapur	Artisanry, cattle robbers	Europoid tribe
Yanadi	36	Chittoor, Nellore, Prakasam	Hunter-gatherers	Proto-"Australoid" tribe
Dudekula	84	Chittoor, Cuddapah	Cotton thread making	Muslims
Sheik	64	Chittoor, Cuddapah, Kurnool, Anantapur	Miscellaneous	Muslims

a. Number of chromosomes.

b. The socioeconomic categories are in decreasing order of hierarchy, except the two Muslim groups.

Statistical Methods. The allele frequencies were estimated using a direct gene-counting procedure. Average heterozygosity and the coefficient of gene differentiation (G_{ST}) along with their standard errors were obtained following the method of Nei (1987). Genetic distances were computed using the modified Cavalli-Sforza distance (D_A) of Nei et al. (1983). The distances were also obtained using the stepwise weighted genetic distance measure (D_{SW}) of Shriver et al. (1995) and Reynolds's θ_{st}/F_{ST} (Reynolds et al. 1983). Although D_A is not linear with evolutionary time, it is the most efficient way to obtain the most correct phylogenetic relationships among closely related populations (Takezaki and Nei 1996). On the other hand, F_{ST} or θ_{st} (Reynolds et al. 1983) is a modified form of Cavalli-Sforza's chord distance, with the assumptions that there is no new mutation and that all gene frequency changes are due to genetic drift. Constant and equal population sizes were not assumed. Therefore this may be the most appropriate measure of genetic distance for the regional populations of Andhra Pradesh, which probably have a short evolutionary history.

The neighbor-joining algorithm (Saitou and Nei 1987) was used to construct the phylogenetic trees. Computations were performed using the NJBAFD program (supplied by N. Takezaki, National Institute of Genetics, Mishima, Japan), DISPAN, and PHYLIP, version 3.573.

The multidimensional scaling plots of the populations, which are based on the genetic distances, were created with the help of SPSS, version 7.5. Hardy-Weinberg equilibrium (Guo and Thompson 1992) and population genetic structure, as inferred from the analysis of molecular variance (AMOVA) (Excoffier et al. 1992), were performed using Arlequin, version 2.00 (Schneider et al. 1997). The AMOVA used in Arlequin is essentially similar to other approaches based on the analysis of variance of the gene frequencies but takes into account the number of mutations between molecular haplotypes. A hierarchical analysis of variance partitions the total variance into covariance components resulting from intra-individual differences, interindividual differences, interpopulation differences, and/or intergroup differences. The significance of these differences is tested using a nonparametric permutation approach (Excoffier et al. 1992) by permuting haplotypes, individuals, or populations among individuals, populations, or groups of populations, respectively. The population structure was hierarchically defined into seven groups considering the socioeconomic criteria and the popular perception in the area based on the traditional occupations and relative ranking of different caste groups (see Table 1). However, the two Muslim groups were kept in a separate and additional category.

We have also performed a Markov chain Monte Carlo analysis of population structure using the program Structure, version 2 (available at http://pritch.bsd.uchicago.edu), which implements a model-based clustering method for grouping individuals in populations and for identifying migrants and admixed individuals. This approach assumes a model with K populations, each of which is characterized by a set of allele frequencies at each locus; individuals are assigned to a population or jointly to two or more populations if their genotypes

indicate that they are admixed. Individuals are grouped into populations in such a way so as to achieve Hardy-Weinberg equilibrium and linkage equilibrium (Pritchard et al. 2000). Because the populations analyzed in this study were closely related in their linguistic affiliations and geographic distribution, we used the improved model of allele-frequency correlations elucidated by Falush et al. (2003) without a priori population information. The burn-in lengths (the duration of a simulation run, before collecting data, to minimize the effect of the starting configuration) for the analysis were chosen in such a way so that the summary statistics would converge, and the simulation run was quite long (8×10^6 runs) to get accurate estimates of parameters.

The regression model of Harpending and Ward (1982) was also used to study the possible effects of genetic drift and admixture on the substructured Telugu populations. In this model the average heterozygosity of the *i*th population (H_i) should be equal to the overall mean heterozygosity of the entire population (in this case, populations of Andhra Pradesh) H_t, multiplied by $(1 - R_{ii})$, where R_{ii} is the genetic distance of a particular population from the gene frequency centroid. If gene flow from outside the region varies in amount from population to population, then this linear relationship no longer holds. Isolated groups will be less heterozygous than the linear prediction and hence will lie below the expected regression line, whereas populations receiving more gene flow from outside will be more heterozygous and therefore will lie above the regression line.

Results

For the sake of brevity the allele-frequency tables for the 27 populations are not presented here and are being published elsewhere (Reddy et al. 2005). The following salient features of the pattern of allele-frequency distributions can be highlighted here.

All the loci were highly polymorphic in these populations, with the number of alleles ranging from 9 for D3S1358, VWA, D5S818, and D13S317 to as many as 30 for D21S11. Overall, the average heterozygosity is high ($\geq$0.80), and the loci with the largest number of alleles are the ones that show the highest values of heterozygosity in different populations (Table 2). The exact test probabilities for Hardy-Weinberg equilibrium suggest significant departures in certain locus-population combinations. Except for D5S818, all the other loci show departures from Hardy-Weinberg equilibrium in some populations. Forty-one out of the 243 loci-population combinations show such departures. Both a high degree of inbreeding among these populations and a relatively small sample size in some cases may account for most of these departures, because in only two cases was the observed departure a result of excess heterozygosity.

Genetic Diversity and Relationships. The spectrums of allele-frequency distributions are fairly uniform across the 27 populations (graphs not presented),

Table 2. Locus and Average Heterozygosity and G_{ST} Values for Different Andhra Pradesh Populations

Population	*D3S1358*	*VWA*	*FGA*	*D8S1179*	*D21S11*	*D18S51*	*D5S818*	*D13S317*	*D7S820*	*Average*
Brahmin	0.744	0.847	0.884	0.857	0.903	0.876	0.725	0.813	0.774	0.825
Kshatriya	0.728	0.798	0.884	0.847	0.903	0.871	0.763	0.807	0.764	0.818
Vysya	0.724	0.754	0.794	0.858	0.918	0.865	0.709	0.815	0.795	0.804
Akuthota	0.722	0.798	0.917	0.834	0.896	0.897	0.773	0.825	0.825	0.832
Kamma	0.727	0.824	0.898	0.838	0.891	0.863	0.710	0.823	0.761	0.815
Kapu	0.718	0.773	0.879	0.779	0.853	0.867	0.708	0.824	0.804	0.801
Pokanati	0.762	0.796	0.934	0.788	0.910	0.884	0.740	0.831	0.820	0.829
Panta	0.74	0.814	0.873	0.813	0.892	0.865	0.716	0.825	0.779	0.813
Vanne	0.75	0.816	0.896	0.874	0.847	0.898	0.699	0.835	0.835	0.828
Balija	0.760	0.825	0.926	0.828	0.896	0.833	0.762	0.841	0.789	0.829
Ediga	0.778	0.838	0.851	0.792	0.813	0.888	0.695	0.815	0.797	0.807
Ekila	0.714	0.802	0.874	0.810	0.866	0.879	0.736	0.855	0.708	0.805
Gandla	0.661	0.745	0.807	0.838	0.907	0.856	0.752	0.743	0.784	0.788
Jangam	0.786	0.796	0.835	0.837	0.887	0.839	0.735	0.829	0.789	0.815
Kurava	0.699	0.763	0.924	0.852	0.866	0.850	0.705	0.795	0.797	0.806
Thogata	0.762	0.803	0.862	0.805	0.913	0.848	0.710	0.795	0.793	0.810
Yadava	0.737	0.838	0.891	0.837	0.896	0.882	0.668	0.802	0.760	0.812
Chakali	0.771	0.783	0.904	0.835	0.868	0.826	0.740	0.838	0.770	0.815
Madiga	0.752	0.813	0.891	0.853	0.865	0.838	0.748	0.811	0.785	0.817
Mala	0.732	0.809	0.913	0.84	0.904	0.852	0.740	0.822	0.777	0.821
Mangali	0.718	0.688	0.882	0.817	0.926	0.898	0.679	0.787	0.775	0.797
Vaddi	0.73	0.826	0.881	0.850	0.844	0.840	0.751	0.836	0.816	0.819
Erukula	0.794	0.783	0.927	0.816	0.896	0.814	0.669	0.822	0.782	0.812
Sugali	0.751	0.816	0.928	0.808	0.878	0.869	0.743	0.810	0.789	0.821
Yanadi	0.695	0.822	0.873	0.815	0.873	0.865	0.622	0.862	0.736	0.796
Dudekula	0.771	0.800	0.881	0.793	0.899	0.880	0.687	0.827	0.803	0.816
Sheik	0.702	0.793	0.909	0.850	0.911	0.797	0.775	0.852	0.815	0.823
Number of alleles	9	9	25	12	30	22	9	9	10	–
G_{ST}	0.020	0.028	0.047	0.028	0.044	0.035	0.022	0.025	0.028	0.031 ± 0.00317

suggesting genetic homogeneity of the populations from this region. Twenty-five of the 27 populations (excluding the two Muslim groups, which do not fit strictly into the scheme of the hierarchical caste system) are grouped into 5 broad socioeconomic categories, pooling the lower-middle1 and lower-middle2 and the lower1 and lower2 categories as single groups (see Table 1). The allele-frequency distributions are presented in Figure 2. The shape of the distribution of allele frequencies is fairly uniform across most socioeconomic groups for most of the loci, barring the few most polymorphic ones. The relatively greater population heterogeneity in the pattern of allele distribution is evident at the loci with the larger numbers of alleles, resulting in much greater G_{ST} values for FGA (0.047), D21S11 (0.044), and D18S51 (0.035). The G_{ST} values for the remaining loci are low and range from 0.020 for D3S1258 to 0.028 for VWA, D8S1179, and D7S820. The average G_{ST} value is low (0.031 ± 0.0032), reflecting relative homogeneity of the Andhra Pradesh populations. Among the five major socioeconomic groups, upper castes ($G_{ST} = 0.015$) and tribes ($G_{ST} = 0.024$) show relatively greater population homogeneity compared to the upper-middle ($G_{ST} = 0.036$), lower-middle ($G_{ST} = 0.036$), or lower ranking caste groups ($G_{ST} = 0.032$) within them.

Although the neighbor-joining tree was constructed using both the D_A and the F_{ST} distances, to visualize the pattern of population relationships, given the significant similarity in the resultant trees, we present here only the F_{ST}-based neighbor-joining tree (Figure 3). Four distinct clusters of populations are apparent in the tree, but no clear separation of the populations based on social hierarchy or geographic affiliation can be discerned. For example, the first major cluster, representing 14 of the 27 Andhra Pradesh populations, consists of two distinct subclusters. One of these subclusters includes all three upper-caste groups (Brahmin, Kshatriya, and Vysya), but it also represents two upper-middle ranking castes (Akuthota and Kamma), a lower-middle ranking group (Kurava), and a tribal group (Sugali), which has a Europoid ethnic background. Similarly, the second major subcluster is a conglomeration of tribes (Yanadi and Erukela), Muslims (Sheik and Dudekula), a lower-middle ranking caste (Thogata), and two lower castes (Mangali and Madiga). The two-dimensional plot of the populations (Figure 4) based on multidimensional scaling of the F_{ST} distances also did not bring out any clear constellations of populations, based on any rational criteria, suggesting a lack of clear differentiation.

Affinities of the Andhra Pradesh Populations with Other Indian and Continental Populations. Compared to other Indian and world populations, the populations of Andhra Pradesh form a distinct cluster clearly separated from the rest (Figure 5). The other populations in the tree seem to be aligned on broad geographic, ethnic, or linguistic affiliations. For example, Asian populations from northeastern India form a distinct cluster, as do the other Asian populations from sub-Himalayan India and East Asia. Populations from western India, central

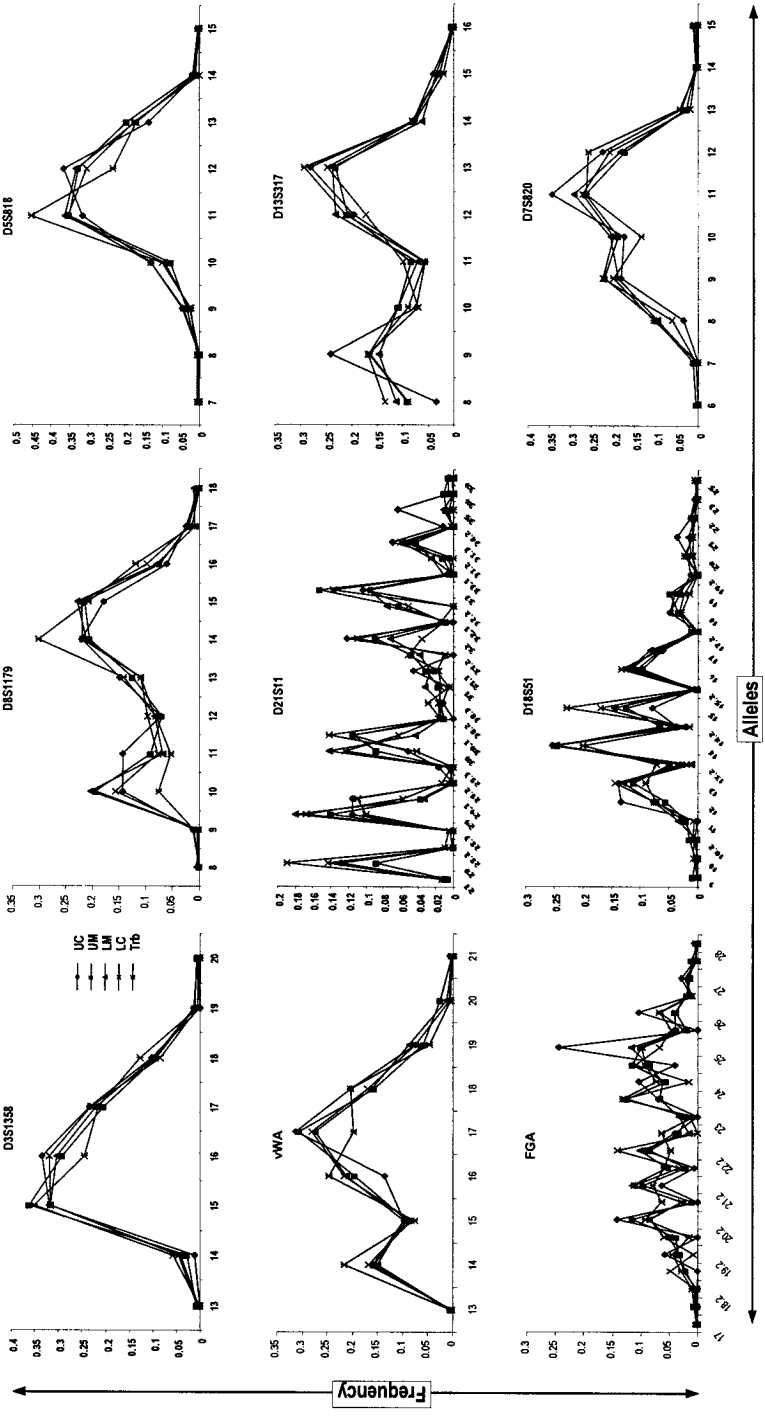

Figure 2. The frequency polygons showing the pattern of allele-frequency distributions across the five broad socioeconomic groups of Andhra Pradesh.

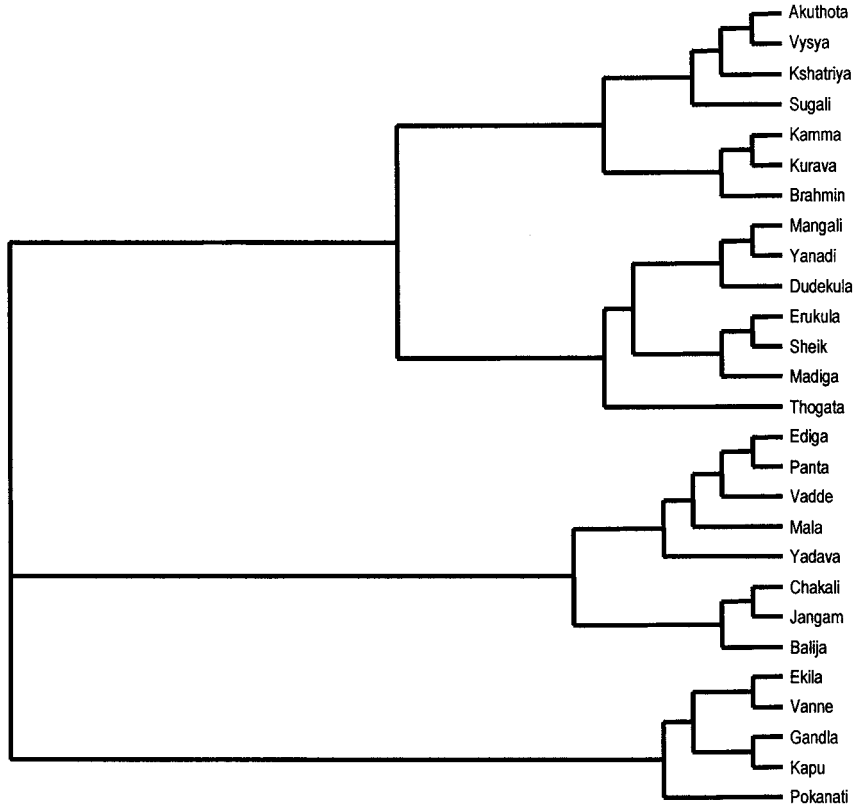

Figure 3. The neighbor-joining tree (constructed on the basis of F_{ST} distances) depicting phylogenetic relationships among the 27 populations of Andhra Pradesh.

India, and North India also form distinct subclades, although geographic contiguity is apparent with their placement as neighboring clades. However, some of the populations from South India that have Europoid physical features (Iyenger, Lingayat, Gowda, and Muslims) form a subclade along with the three American groups (US whites, US Hispanics, and African Americans).

Genetic Structure Versus Social Structure. The analysis of molecular variance (AMOVA) considering the 27 Andhra Pradesh populations in 8 rational groups, based on socioeconomic criteria, suggests virtually no variation among the groups (Table 3). Although the variation among populations within the groups is statistically significant, it is rather small (1.37%). Most of the variation is observed to be between individuals within populations. This reiterates the relative homogeneity of the populations shown by the neighbor-joining trees based on

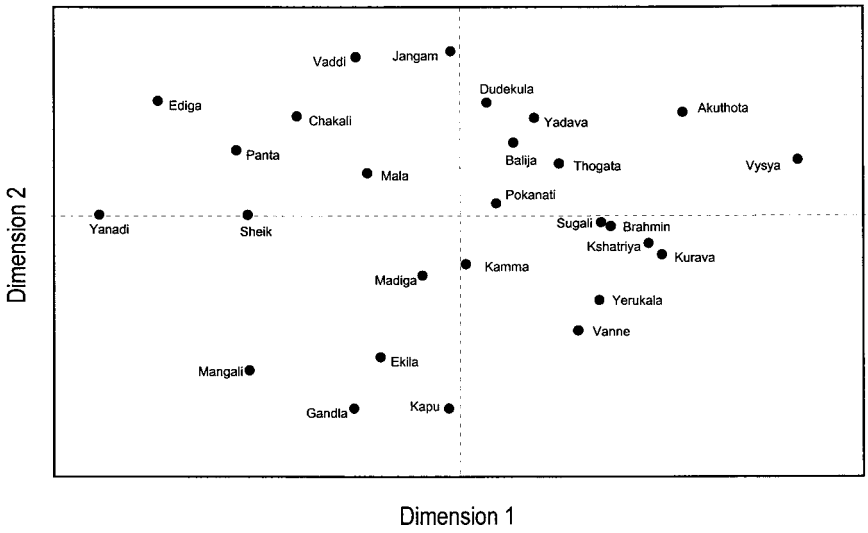

Figure 4. Two-dimensional plot based on multidimensional scaling of the F_{ST} distance matrix of the 27 Andhra Pradesh populations.

different distance measures. The AMOVA results remained almost identical even after the number of and criteria (e.g., geographic contiguity and/or socioeconomic similarity) for constituting the rational groups was changed.

Because the AMOVA results did not conform to the population structure based on socioeconomic hierarchy, we further attempted a structural analysis of the populations without a priori information using an admixture model, and we repeated the run for different values of K ($K = 1$ to 20) to ascertain whether any genetic structure other than that based on the socioeconomic criteria exists. We found that the estimates of Pr(genotypic data $|$ K) are similar for $K = 1$ to 11, suggesting that there is no genetic structure among these groups that could be revealed by the nine STR loci. Beyond $K = 11$ the estimates of the probability decrease substantially. Although, the estimates of the probability were similar for $K = 1$ to 11, we chose $K = 8$ as the optimum number of clusters to be inferred to compare with the results of the AMOVA. Figure 6, based on the proportion of membership of each predefined population in each cluster, suggests that none of the inferred clusters have a significant proportion of membership from any of the populations. In fact, all the populations exhibit a uniform degree of presence in each of the inferred clusters, which is consistent with the AMOVA results. Structural analysis of the data on the five subgroups of the Reddy caste (results not presented) did not bring out any genetic substructure bearing signatures of the subcaste endogamy and differentiation.

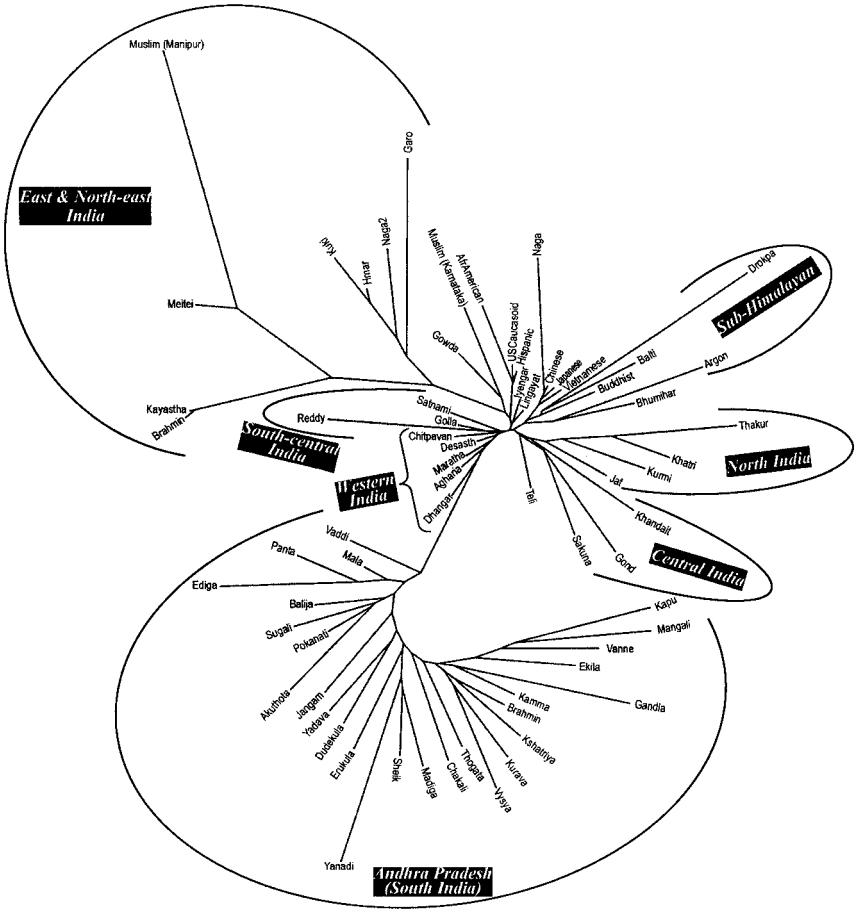

Figure 5. Cladogram depicting the relationship of the populations of Andhra Pradesh to other Indian and continental populations. Sources for comparative data: AfrAmerican (African American), USCaucasoid (US whites), and Hispanic (US Hispanics) (*AmpFlSTR Identifiler PCR Amplification Kit User's Manual* 2000); Vietnamese (Borys et al. 1999a); Japanese (Borys et al. 1999b); Chinese (Fung et al. 2001); Golla (Reddy et al. 2001b); Iyengar, Gowda, Lingayat, and Muslim (Karnataka) (Rajkumar and Kashyap 2002); Agharia, Satnami, Gond, and Teli (Sarkar and Kashyap 2002); Buddhist, Argon, Drokpa, and Balti (Trivedi et al. 2002); Maratha, Desasth Chitpavan, and Dhangar (Gaikwad and Kashyap 2002); Bhumihar, Reddy, Sakuna, Naga, and Khandait (Kashyap et al. 2002); Thakur, Khatri, Kurmi, and Jat (Tandon et al. 2002); Garo, Naga2, Kuki, and Hmar (Chattopadhyay et al. 2001); Brahmin, Kayastha, Muslim (Manipur), and Meitei (Dutta et al. 2000).

Table 3. Results of AMOVA Based on the 27 Andhra Populations Rationalized into 8 Socioeconomic Groups Depicting the Population Structure[a,b]

Source of Variation	*df*	*Sum of Squares*	*Variance Component*	*Variation (%)*
Among groups	7	51.232	−0.00057 (Va)	−0.02
Among populations within groups	19	128.751	0.04912 (Vb)	1.37
Within populations	1,869	6609.988	3.53664 (Vc)	98.65
Total	1,895	6789.97	3.58519	

a. Fixation indexes: F_{ST}: 0.01354 (significant at 0.01); F_{SC}: 0.01370 (significant at 0.01); F_{CT}: −0.00016.

b. Group 1: Brahmin, Kshatriya, Vysya; Group 2: Pokanati, Vanne, Panta, Akuthota, Kapu, Kamma; Group 3: Balija, Yadava, Kurava, Ekila, Thogata; Group 4: Jangam, Gandla, Ediga; Group 5: Chakali, Mangali, Vaddi; Group 6: Mala, Madiga; Group 7: Sugali, Erukula, Yanadi; Group 8: Sheik, Dudekula.

To further gauge the pattern of relationships between different hierarchical caste groups and the tribes, we computed Nei's standard genetic distances (D_S) and standard errors (Nei 1972, 1978) among the five broad socioeconomic groups (upper, upper-middle, lower-middle, and lower castes and tribes) (Table 4). Although genetic distance tended to increase with increasing difference in the social hierarchy, the differences were not statistically significant. Even this meek trend

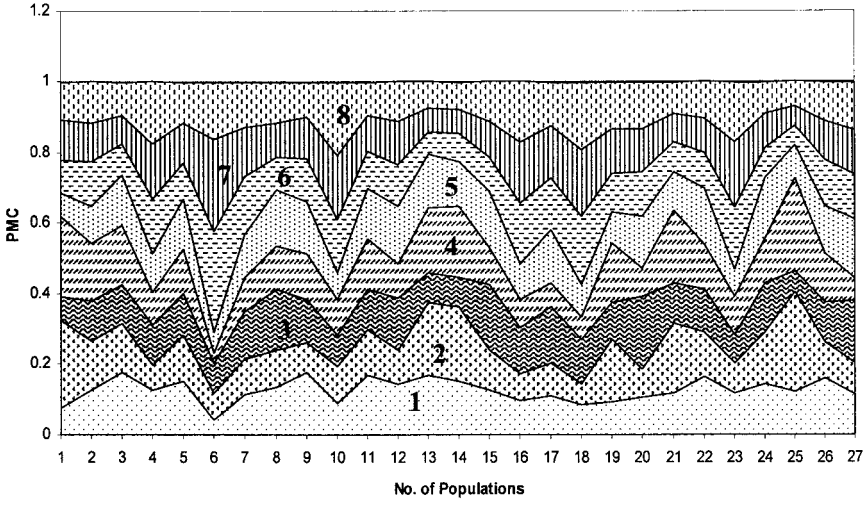

Figure 6. Proportion of membership coefficients (PMC) of the 27 Andhra Pradesh populations into the 8 inferred clusters.

Table 4. Standard Genetic Distances Between the Hierarchical Caste Groups of Andhra Pradesh and the Tribes (Below Diagonal) and Their Standard Errors (Above the Diagonal)

	Upper Castes	*Upper-Middle Castes*	*Lower-Middle Castes*	*Lower Castes*	*Tribes*
Upper castes	–	0.0129	0.0152	0.0235	0.0152
Upper-middle castes	0.0247	–	0.0051	0.0047	0.0097
Lower-middle castes	0.0295	0.0048	–	0.0040	0.0148
Lower castes	0.0365	0.0077	0.0034	–	0.0109
Tribes	0.0475	0.0304	0.0371	0.0303	–

disappears when we consider D_A distances or when we compute average distances for different pairs of populations between different hierarchical groups. Furthermore, the average distance between populations of the same socioeconomic group is not significantly different from or lower than the average distance between the populations of different groups. However, each of these hierarchical groups shows the largest genetic distance with the tribes compared to the mutual distances among them, suggesting genetic isolation and differentiation of the tribes and castes.

Gene Flow. The regression plot of mean heterozygosity versus the R_{ii} of the 27 Andhra Pradesh populations is depicted in Figure 7. It is interesting to note from the plot that the Akuthota Kapu, Vanne Kapu, and Pokanati (the three subgroups of the Reddy caste) appear as relatively more distinct outliers above the theoretical regression line, suggesting that external gene flow played a role in their differentiation. However, the position of the Akuthota Kapu as an outlier above the theoretical regression line is quite distinct, besides being far removed from the gene frequency centroid with a large value of R_{ii}. This may suggest that admixture as well as stochastic processes played a role in shaping their genetic composition. The Gandla, Yanadi, Mangali, and Kapu appear as outliers below the theoretical regression line, indicating that isolation and stochastic processes played a role in sculpting their genetic composition. The rest of the populations are scattered above and below but in the vicinity of the regression line, in conformity to the model and suggesting a uniform degree of gene flow among them.

Discussion

In recent years there has been increasing use of microsatellite loci to understand genetic relationships between closely related populations (Chu et al. 1998; Reddy et al. 2001a, 2001c). In the present study we used nine autosomal AmpF*l*-STR microsatellite markers to understand the population structure and patterns of variation of the 27 caste and tribal populations distributed in the contiguous

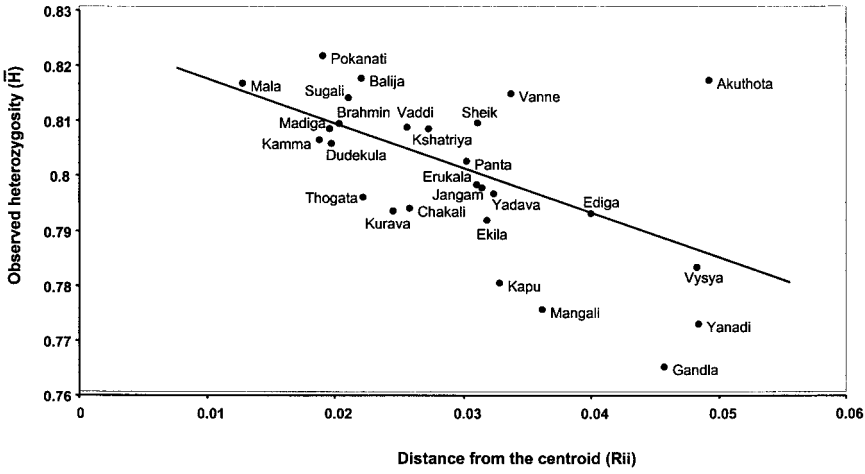

Figure 7. Regression plot of average heterozygosity (H) versus the distance from the centroid (R_{ii}) of the 27 Andhra Pradesh populations.

areas of the southern districts of Andhra Pradesh in India. Although most of these microsatellite loci are highly polymorphic within each of the Andhra Pradesh populations, the allele distributions are fairly uniform across the populations, suggesting relative homogeneity among them.

Concurrent with this, the coefficient of gene differentiation is low ($G_{ST} = 0.031 \pm 0.0032$) and is almost identical to the value obtained by Reddy et al. (2001c) (but using a different set of 13 STR loci) among 8 endogamous subcastes of a substructured single caste, the Golla, of Andhra Pradesh. This low coefficient of gene differentiation is reflected in the lack of clear differentiation and clustering pattern of the population based on either socioeconomic stratification or known ethnohistorical and geographic affiliations. Our analyses based on the three most informative markers (FGA, D21S11, and D18S51) did not qualitatively change the results. Consistent with the overall homogeneity among them, the 27 Andhra Pradesh populations form a single and compact cluster compared to other Indian and world populations with a high degree of support from bootstraps.

In a recent study of southwestern Indian populations based on data from 15 microsatellites, Rajkumar and Kashyap (2004) observed similar trends. Bamshad et al. (2003) and Rosenberg et al. (2002) also observed that a large number of microsatellite loci are required to differentiate populations, even on the continental scale, implying a need for a much larger number of microsatellite loci to resolve the phylogeny on the regional level. However, the nine autosomal AmpF*l*-STR markers that are validated and widely used for forensic investigations have

also been found to be useful for unraveling the local population structure and for reconstructing evolutionary relationships at the level of ethnic, geographic, and linguistic categories in India (Dutta et al. 2002; Langstieh et al. 2004) and elsewhere (Sun et al. 2003). Furthermore, on the basis of 13 STR loci, Reddy et al. (2001c) observed that these markers help to reconstruct the short evolutionary history at the level of endogamous subcastes of an Indian caste, because these loci seem to have left signatures of subcaste endogamy.

Assuming that microsatellite loci are useful for clarifying evolutionary relationships of closely related populations (Takezaki and Nei 1996), how can one explain the uniformly high degree of polymorphism and heterogeneity within the populations and the reduced diversity between the populations? Most of the genetic variation accrued in the populations of Andhra Pradesh can best be explained in terms of founder effects in the formation of the subgroups and subsequent genetic drift over the generations. Therefore, if the barriers of caste and tribal boundaries are impermeable, the observed homogeneity of the Andhra Pradesh populations may imply a relatively recent history of separation or substructuring, so that the small differences that may have accumulated are not being captured by the nine loci used in this study.

With a relatively larger number of STR markers among the regional populations of China, Chu et al. (1998) interpreted similar findings as possibly reflecting single origin, although Chu acknowledged that the lack of resolution of the STR loci used could also be one of the reasons for the observed pattern. However, Langstieh et al. (2004) recently observed that the reduced microsatellite diversity among the Meghalaya populations is due to admixture, which is perpetuated by the system of matrilineal descent and matrilocal residence, coupled with the relatively short history of separation of these tribes. Can these explanations be plausibly extended to the present situation in the Andhra Pradesh populations? The populations of southern Andhra Pradesh have clearly defined hierarchical structure, including within them the upper castes, middle-ranking castes, lower castes, and tribes. There could be two plausible scenarios that could explain the observed genetic composition and structure of these Andhra populations: (1) a recent and common origin of these populations (hence the genetic differentiation is not significant enough to show a systematic pattern of population relationships); and (2) unrecognized gene flow, albeit at a low rate, over the generations among the coexisting and closely interacting caste populations from a homogeneous area.

Even though it is interesting to note the systematic pattern of the differences in genetic distances observed between the hierarchical groups, tempting one to surmise that this is not really a case of genetic stratification being consistent with the social stratification, given the large standard errors, this stratification does not reach statistical significance. However, Bamshad et al. (2001), using a large set of autosomal markers, found this stratification to be highly significant in the caste populations of Andhra Pradesh. When we computed distances based on the nine STR loci for the hierarchical caste groups, treating the three *varna* categories in the upper castes (Brahmin, Kshatriya, and Vysya) separately, no

particular pattern of genetic distances, adhering to the implicit hierarchy, emerged between them (distance matrix not presented). This suggests a lack of strong genetic signatures consistent with the traditional *varna* system (constituting only Brahmin, Kshatriya, Vyshya, and Sudra categories), although a semblance of genetic stratification was evident with respect to socioeconomic hierarchy (i.e., upper, middle, and lower castes).

It may be pertinent to note here that Bamshad et al. (1998) did not observe genetic stratification in the Y-chromosome-based markers, whereas Bhattacharyya et al. (1999) found no evidence of male gene flow across caste and ethnic boundaries in India. On the other hand, Ramana et al. (2001) and Cordaux et al. (2004) found evidence to support bidirectional male gene flow in southern India, particularly in the caste populations of Andhra Pradesh. Furthermore, based on traditional genetic markers, Kumar et al. (2004) demonstrated a tribe-caste continuum in the genetic structure of the populations of northeast India, suggesting gene flow between them.

Bamshad et al. (1998) interpreted the observed stratification in mtDNA among the populations of Andhra Pradesh as due to movement of females, because of the practice of hypergamy in the patrilocal Hindu society, and, conversely, the lack of stratification in Y-chromosome-based markers as due to a lack of movement of males across caste boundaries. No doubt, hypergamy, which is widely acknowledged in the literature to have existed in historical times, might have been practiced only by the minority in the higher echelons of the society, merely as a mechanism to accommodate and legitimize the multiple marriages that the men of the ruling class were accustomed to. The practice of hypergamy as an institution or as a traditional practice is certainly unknown in contemporary India to have caused such a systematic and significant genetic stratification. On the contrary, given the traditional pattern of symbiotic relationships among the cohabiting and contiguously distributed hierarchical caste groups with day-to-day intimate interactions, we speculate that there might have been continuous unauthorized or unrecognized gene flow, albeit at a low rate, across the social groups, despite strict adherence to the caste endogamy, resulting in relative homogenization of at least the caste groups. This is perhaps implicit in the observed homogeneity of caste populations in southern India for Y-chromosome-based markers (Cordaux et al. 2004).

A reflection of this can also be seen in the results of structure analyses with the admixture model, because the proportions of membership of different populations is fairly uniform across all the inferred clusters (see Figure 6). This is feasible only through male gene flow in a strictly endogamous and patrilocal society; there is no possibility of female genes transecting the caste boundaries, even in the event of unauthorized sexual interaction between caste groups. Therefore one may expect reduced Y-chromosome diversity across the populations and even stratification based on social hierarchy, given relatively greater interaction between groups of adjacent ranks. This may explain the mild gradient of genetic distances observed with the changing hierarchy of populations in the present

study. Given the small number of microsatellite loci used in this study, the conclusions remain tentative until our analysis of Y-chromosome-based markers provides more conclusive evidence on the probable processes behind the observed trend of genetic stratification vis-à-vis social stratification in the Indian populations.

Acknowledgments This study is part of the Indian Statistical Institute's plan project titled "DNA Polymorphisms in the Caste and Tribal Populations of Andhra Pradesh, India," which is being carried out in collaboration with the Centre for Cellular and Molecular Biology, Hyderabad. The project leader (B. M. Reddy) thanks the directors of both institutes for logistical support and the two anonymous reviewers for helpful comments and suggestions. We are grateful to the donors of the samples.

Received 10 May 2005; revision received 8 September 2005.

Literature Cited

AmpFlSTR Identifiler PCR Amplification Kit User's Manual. 2000. Foster City, CA: Applied Biosystems.

Bamshad, M., A. E. Fraley, M. H. Crawford et al. 1996. mtDNA variation in caste populations of Andhra Pradesh, India. *Hum. Biol.* 68:1–28.

Bamshad, M. J., T. Kivisild, W. S. Watkins et al. 2001. Genetic evidence on the origin 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.

Bamshad, M. J., S. Wooding, W. S. Watkins et al. 2003. Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* 72:578–589.

Basu, A., N. Mukherjee, S. Roy et al. 2003. Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res.* 13:2277–2290.

Bhattacharyya, N. P., P. Basu, M. Das et al. 1999. Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.* 9:711–719.

Borys, S., A. Eisenberg, G. Carmody et al. 1999a. Allele frequencies for 9 STR loci in African American, Chinese, Vietnamese, and Bangladesh populations. *J. Forensic Sci.* 44:1316–1318.

Borys, S., S. Iwaamoto, J. Miyakoshi et al. 1999b. Allele frequency distribution for 9 STR loci in the Japanese population. *J. Forensic Sci.* 44:1319.

Chattopadhyay, P., R. Dutta, and V. K. Kashyap. 2001. Allele frequency data at nine fluorescent labeled STR loci in four tribal population groups of India. *J. Forensic Sci.* 45:184–188.

Chu, J. Y., W. Huang, S. Q. Kuang et al. 1998. Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. USA* 95:11,763–11,768.

Clark, V. J., S. Sivendren, N. Saha et al. 2000. The 9-bp deletion between the mitochondrial lysine tRNA and COII genes in tribal populations of India. *Hum. Biol.* 72:273–285.

Cordaux, R., R. Aunger, G. Bentley et al. 2004. Independent origins of Indian caste and tribal paternal lineages. *Curr. Biol.* 4:231–235.

Cordaux, R., N. Saha, G. R. Bentley et al. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* 11:53–64.

Dutta, R., P. Chattopadhyay, and V. K. Kashyap. 2000. STR data for the AMPFlSTR profiler plus loci among four predominant populations of eastern India. *J. Forensic Sci.* 45:1353–1357.

Dutta, R., B. M. Reddy, P. Chattopadhyay et al. 2002. Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum. Biol.* 74:33–49.

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.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inferences of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.

Fung, W. K., J. Ye, L. Hu et al. 2001. Allele frequencies for nine STR loci in Beijing Chinese. *Forensic Sci. Int.* 121:207–209.

Gadgil, M., N. V. Joshi, S. Manoharan et al. 1998. Peopling of India. In *The Indian Human Heritage*, D. Balasubramanian and N. A. Rao, eds. Hyderabad, India: University Press, 100–129.

Gaikwad, S., and V. K. Kashyap. 2002. Polymorphism at fifteen hypervariable microsatellite loci in four populations of Maharashtra, India. *Forensic Sci. Int.* 126:267–271.

Guo, S., and E. Thompson. 1992. Performing exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.

Harpending, H. C., and R. Ward. 1982. Chemical systematics and human evolution. In *Biochemical Aspects of Evolutionary Biology*, M. Nitecki, ed. Chicago: University of Chicago Press, 213–256.

Karve, I. 1961. *Hindu Society: An Interpretation.* Poone, India: Deshmukh Prakashan.

Kashyap, V. K., N. Sarkar, and R. Trivedi. 2002. Allele frequencies for STR loci of the Powerplex 16 Multiplex system in five endogamous populations of India. *Forensic Sci. Int.* 126:178–186.

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

Kivisild, T., K. Kaldma, M. Metspalu et al. 1999b. The place of mitochondrial DNA variants in the global network of the maternal lineages and the peopling of the Old World. In *Genome Diversity: Applications in Human Population Genetics,* S. S. Papiha and R. Deka, eds. New York: Kluwer, 135–152.

Kivisild, T., S. Rootsi, M. Metspalu et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72:313–332.

Kosambi, D. D. 1991. *The Culture and Civilization of Ancient India in Historical Outline*. New Delhi, India: Vikas.

Krishnan, T., and B. M. Reddy. 1994. Geographical and ethnic variability of finger ridge counts: Biplots of male and female Indian samples. *Ann. Hum. Biol.* 21:155–169.

Kumar, V., D. Basu, and B. M. Reddy. 2004. Genetic heterogeneity in northeastern India: Reflection of tribe-caste continuum in the genetic structure. *Am. J. Hum. Biol.* 16:334–345.

Langstieh, B. T., B. M. Reddy, K. Thangaraj et al. 2004. Genetic relationships among the tribes of Meghalaya, India: Matriliny and reduced microsatellite diversity. *Hum. Biol.* 76:569–590.

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

Nei, M. 1972. Genetic distances between populations. *Am. Nat.* 106:283–292.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.

Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19:153–170.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Rajkumar, R., and V. K. Kashyap. 2002. Distribution of alleles of 15 STR loci of the Powerplex Multiplex system in four predominant population groups of South India. *Forensic Sci. Int.* 126:173–177.

Rajkumar, R., and V. K. Kashyap. 2004. Genetic structure of four socioculturally diversified caste populations of southwest India and their affinity with related Indian and global groups. *BMC Genet.* (online) 5:23.

Ramana, G. V., B. Su, L. Jin et al. 2001. Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* 9:695–700.

Reddy, B. M. 2002. Patterns of endogamy and consanguinity in India [abstract]. Invited talk delivered at the Symposium on Community Genetics in Developing Countries, Indian Institute of Science, Bangalore.

Reddy, B. M., D. A. Demarchi, S. Bharti et al. 2004. Patterns of ethnic, linguistic, and geographic heterogeneity of palmer interdigital ridge counts in the Indian subcontinent. *Hum. Biol.* 76:211–228.

Reddy, B. M., R. Dutta, B. Langstieh et al. 2001a. Diversity at three tetrameric STR loci in a substructured Golla caste population of southern Andhra Pradesh in comparison to other Indian populations. *Int. J. Hum. Genet.* 1:1–9.

Reddy, B. M., V. M. Naidu, V. K. Madhavi et al. 2005. Data for the 9 STR loci among 27 populations of different social hierarchy from southern part of Andhra Pradesh, India. *Forensic Sci. Int.* 149:81–97.

Reddy, B. M., G. Sun, and R. Dutta. 2001b. STR data for the AmpF/STR Profiler Plus loci among Golla population of southern Andhra Pradesh, India. *J. Forensic Sci.* 46:734–735.

Reddy, B. M., G. Sun, J. R. Luis et al. 2001c. Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Hum. Biol.* 73:175–190.

Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation for coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105:767–779.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber et al. 2002. Genetic structure of human populations. *Science* 298:2381–2385.

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

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sarkar, N., and V. K. Kashyap. 2002. Genetic diversity at two pentanucleotide STR and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of central India. *Forensic Sci. Int.* 128:196–201.

Schneider, S., D. Rosslie, and L. Excoffier. 1997. *Arlequin ver. 2.000: A Software for Population Genetics Data Analysis.* Geneva: Genetics and Biometry Laboratory, University of Geneva. Available at http://anthropologie.unige.ch/arlequin

Shriver, M. D., L. Jin, E. Boerwinkle et al. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* 12:914–920.

Sun, G., S. T. McGarvey, R. Bayoumi et al. 2003. Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. *Eur. J. Hum. Genet.* 11:39–49.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.

Tandon, M., R. Trivedi, and V. K. Kashyap. 2002. Genomic diversity at 15 fluorescent labeled short tandem repeat loci in few important populations of state of Uttar Pradesh, India. *Forensic Sci. Int.* 128:190–195.

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

Thapar, R. 1995. The first millennium B.C. in northern India (up to the end of Mauryan period). In *Recent Perspective of Early Indian History*, R. Thapar, ed. Bombay: Popular Prakshan, 80–141.

Thapar, R. 2003. *Early India: From the Origins to A.D. 1300.* Berkeley: University of California Press.

Trivedi, R., P. Chattopadhyay, B. Maity et al. 2002. Genetic polymorphism at nine microsatellite loci in four high altitude Himalayan desert human populations. *Forensic Sci. Int.* 127:150–155.

Watkins, W. S., M. Bamshad, M. E. Dixon et al. 1999. Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Am. J. Phys. Anthropol.* 109:147–158.