Population Structure of the Chenchu and Other South Indian Tribal Groups: Relationships between Genetic, Anthropometric, Dermatoglyphic, Geographic, and Linguistic Distances

S.M. SIRAJUDDIN,[1,2] R. DUGGIRALA,[1] AND M.H. CRAWFORD[1]

Abstract   We describe the genetic structure and interrelationships of nine south Indian tribal groups (seven from Andhra Pradesh and two from the adjoining states of Tamil Nadu and Kerala) using seven polymorphic loci (ABO, MN, RH, PGM, ACP, PGD, and LDH). *R* matrix analysis indicates that the Andhra Pradesh tribes are clustered and that the Kadar and Irula are genetically isolated from them. This dispersion of populations has been explained by the combination of relatively high frequencies of the alleles RH *D* and MN *M* in the Kadar and the relatively high proportions of the allele *PGM*2* in the Irula. The Mahaboobnagar Chenchu subgroup is isolated from other Telugu-speaking groups because of high frequencies of the *PGM*1* and *ACP*A* alleles. The regression of mean per locus heterozygosity (*H*) on distance from the gene frequency centroid (r_{ii}) reveals considerable levels of external gene flow among the Lambadi, the Yerukula, and the two Chenchu subgroups and more homogeneity in the Kolam, Koya, Yanadi, Irula, and Kadar. Mantel statistics were used to assess the relative effects of nonbiological processes (i.e., language and geography) on the morphological and genetic patterns of these subdivided populations. The significance of correlations was determined between different data sets (genetic, dermatoglyphic, anthropometric, geographic, and linguistic) at three levels involving nine, six, and five populations. Although multiple correlation analysis reveals significant combined effects of geography and language on genetics, anthropometrics, and dermatoglyphics, highly significant partial correlations suggest strong effects of geography on both anthropometry and genetics. Our analysis indicates that geographic factors have an overwhelming effect on the genetic differentiation of the south Indian tribal groups.

Anthropologically, the most distinguishing feature of Indian society is its broad subdivision into tribal and caste populations. The tribes constitute 7.5% of the total Indian population. The numerous caste groups

---

(about 2000 in total) have complex hierarchies and practice marital endogamy and the Hindu religion. Several anthropological and genetic investigations have demonstrated significant genetic and phenetic differences between tribal and caste populations, both within and between (Chakraborty and Yee 1973; Balakrishnan 1978; Malhotra 1978; Char et al. 1989; Pathak and Kaul 1991; Roychoudhury 1984a, 1992). The amount of intragroup heterogeneity is enormous, and the tribal populations are believed to be the original inhabitants of India. Labie et al. (1989) recently postulated a unicentric origin of the tribal populations of India on the basis of their beta-globin DNA haplotypes.

The contemporary tribal populations of India can be broadly divided into five geographic regions: northeastern; northern and northwestern; eastern and central; western; and southern Indian tribal groups (Roychoudhury 1984a). These five geographic subdivisions are based on language, culture, and biophenetic characteristics. In each geographic region affinities among the subgroups are appreciable (Guha 1955). Tribal groups are predominantly confined to the hilly tracts and valleys, and the languages spoken by these tribal groups are similar in most cases to the major regional languages of the region or state. Of the many intriguing questions about the southern Indian tribal groups, two especially significant problems involve the origin of the so-called Negritoid and/or proto-Australoid element and its relationship to the Veddas and the Australian aborigines (Guha 1944; Sarkar 1954; Kirk et al. 1962; Simmons 1976; Bhalla 1984; Roychoudhury 1984b).

Population genetic studies on the tribal groups of south India have failed to reveal a correspondence between ethnological similarities and genetic relationships (Kirk 1976; Simmons 1976; Roychoudhury 1984a; Ghosh et al. 1977; Balakrishnan 1978; Saha et al. 1974). The action of systematic versus stochastic processes on the population subgroups has been examined separately using independent data sets (Chakravarti and Mukherjee 1964; Ghosh 1975; Saha et al. 1976; and others). The exact genetic relationships between the Nilgiri hill tribes (Irula, Kota, Toda, Kurumba), the Chenchu, and the Veddas of Sri Lanka are still ambiguous (Haimendorf 1943, 1982; Raghavaiah 1962; Balakrishnan 1978; Bhalla 1984). Reddy et al. (1982) and Sirajuddin and Balakrishnan (1991) tested the purported ethnic affinities of the Chenchu with the Yanadi, the Irula, and the Kadar and concluded that the Chenchu are biologically distinct from other tribal groups. However, these studies relied heavily on a few genetic markers with inadequate representation of the tribal groups, and thus the comparisons and significance between different data sets cannot be adequately assessed.

Given the paucity of data on southern Indian tribal groups in relation to factors such as territorial contiguity, historical reconstruction, and environmental factors, here we examine the genetic and phenetic

variation among nine south Indian tribal groups. These tribes are the Chenchu (from Kurnool and Mahaboobnagar); the Yanadi, Yerukula, Kolam, Koya, and Lambadi of Andhra Pradesh state; the Irula of Tamil Nadu state; and the Kadar of Kerala state. These tribal groups were selected for the study of their genetic composition and the probable action of microevolutionary forces and to retest the previously reported ethnological similarities between the Chenchu, Yanadi, Irula, and Kadar proposed by Haimendorf (1943) and Raghavaiah (1962). However, we use different data sets (genetic, anthropometric, dermatoglyphic, geographic, and linguistic) to test the possible genetic affinities. The relationships between genetic, morphological, geographic, and linguistic data were examined using Mantel statistics to obtain a comprehensive understanding of the evolution and genetic structure of the subdivided tribal groups.

Materials and Methods

The data for this study are based on an earlier study (Sirajuddin 1985) and from other published sources for the nine tribal groups. Although a battery of genetic markers (blood groups, serum proteins, and red cell enzymes) were collected for each tribal group, only 7 common polymorphic systems (ABO, MN, RH, PGM, ACP, LDH, and PGD) with 17 alleles were available for all populations. The seven genetic polymorphic systems and the anthropometric and dermatoglyphic variables used are presented in Table 1. The frequencies of the genes, location, and sample size of each population are given in the source publications of Table 1. The geographic locations of the nine tribal groups under study are shown in Figure 1.

Anthropometric data (males only) on five populations (both Chenchu subgroups, Yanadi, Irula, and Kadar) and dermatoglyphic data (males only) on six tribal groups (both Chenchu subgroups, Yerukula, Yanadi, Irula, and Kadar) were available. These data were used to calculate anthropometric and dermatoglyphic distances. The nine anthropometric variables used in this analysis are stature, head length, head breadth, minimum frontal breadth, bizygomatic breadth, total facial height, upper facial height, nasal height, and nasal breadth. Six common dermatoglyphic traits (subdivided into 20 variables), such as finger pattern types (3), palmar area patterns (5), monomorphic hands (1), mainline formulas (3), C-line polymorphism (3), and position of axial triradii (5), were available and used in the distance calculations. The details of the data and the analysis of anthropometric and dermatoglyphics data are included in the publications cited in Table 1.

Gene frequency data for the seven polymorphic systems for nine alleles (*ABO*A, *ABO*B, MN *M*, RH *D*, *PGM*1, *PGM*2, *ACP*A,

Table 1. Tribal Groups and Comparative Data Used in the Analysis

Tribal Group	Area	State	Data Used[a]	Source
Kurnool Chenchu	Kurnool	Andhra Pradesh	Genetic, anthropometric, dermatoglyphic	Sirajuddin (1985); Ramesh et al. (1980)
Mahaboobnagar Chenchu	Mahaboob Nagar	Andhra Pradesh	Genetic, anthropometric, dermatoglyphic	Sirajuddin (1985); Ramesh et al. (1980)
Irula	Nilgiri Hills	Tamil Nadu	Genetic, anthropometric, dermatoglyphic	Saha et al. (1976); Undevia et al. (1989); Chakravarti and Mukherjee (1964); Guha (1935)
Yanadi	Nellore	Andhra Pradesh	Genetic, anthropometric, dermatoglyphic	Reddy et al. (1982); Sreenath (1977); Reddy and Subramanyam (1985)
Kadar	Trichur Hills, Annamalai Hills	Kerala	Genetic, anthropometric, dermatoglyphic	Saha et al. (1974, 1976); Chakravarti and Mukherjee (1964); Sarkar et al. (1959)
Yerukula	Warangal Mahaboob Nagar	Andhra Pradesh	Genetic, dermatoglyphic	Goud and Rao (1979); Narahari (1982); Blake et al. (1981)
Kolam	Adilabad	Andhra Pradesh		Ramesh et al. (1979); Pingle et al. (1981)
Koya	Adilabad Khammam	Andhra Pradesh		Pingle et al. (1981); Blake et al. (1981)
Lambadi	Hyderabad	Andhra Pradesh		Roberts et al. (1980) Goud and Rao (1979)

a. Genetic data are on ABO, MN, RH, PGM, ACP, PGD, and LDH loci. Anthropometric data include stature, head length and breadth, minimum frontal breadth, bizygomatic breadth, total facial height, upper facial height, and nasal height and breadth. Dermatoglypic data are on palmar and finger dermatoglyphics.
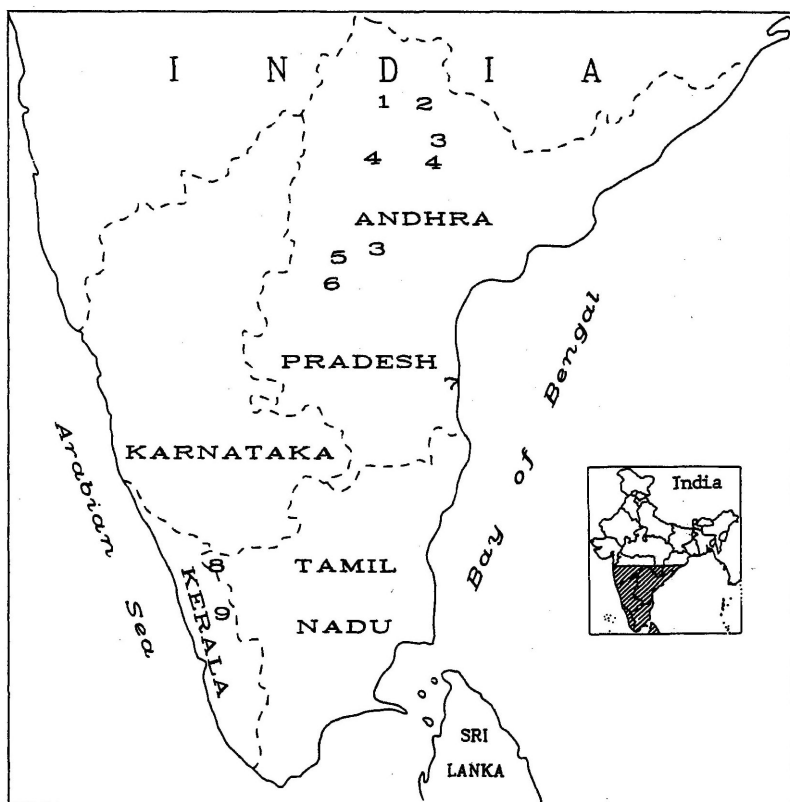
Figure 1. Geographic locations of the nine tribal groups from south India. Some tribal groups, such as the Yerukula, were represented by more than one sample. The multiple samples from a single tribe are indicated in this map by the same number. 1, Kolam; 2, Koya; 3, Yerukula; 4, Lambadi; 5, Mahaboobnagar Chenchu; 6, Kurnool Chenchu; 7, Yanadi; 8, Irula; 9, Kadar.

PGD*A, LDH*N) were subjected to R matrix analysis following the method of Harpending and Jenkins (1973). Of the total information on 17 alleles, 8 alleles (*ABO*O*, MN *N*, RH *d*, *PGM*1,6*, *ACP*B*, *ACP*C*, *PGD*C*, *PGD*K*, LDH *CAL1*) were not included in the R matrix analysis to avoid redundancy and to exclude the less informative rare variants. A variance-covariance matrix was constructed to compute the eigenvectors and the genetic distances. The weighted mean of the diagonal elements r_{ii} provides an estimate of F_{ST}. Because census or population sizes are not available for all the groups, uniform weights have been given to all populations. The eigenvectors, scaled by their respective eigenvalues, are plotted to reveal the genetic relationships between the subdivisions.

In addition, the method of Harpending and Ward (1982) was used to estimate the differential systematic pressure among the tribal groups. Under the assumption of uniform systematic pressure, we can predict a linear relationship between mean per locus heterozygosity and the distance of the population subdivision from the gene frequency centroid. The positive and negative departures from theoretical predictions suggest the probable effects resulting from external gene flow (systematic pressure) and genetic drift (nonsystematic pressure), respectively.

Anthropometric distances between five tribal groups (both Chenchu subgroups, Irula, Yanadi, Kadar) were obtained based on nine morphometric measurements following Mahalanobis's (1936) generalized distance. Dermatoglyphic distances were computed from 20 finger and palmar pattern variables for the 6 tribal groups (the previous five groups and Yerukula) following Balakrishnan and Sanghvi's (1968) B^2 method.

Geographic distances were measured in kilometers from a scaled map as a straight-line distance between the data collection centers for the populations. When several settlements were sampled from a single tribe, geographic distances were measured from a centroid equidistant to all the samples.

The linguistic distances were calculated between the nine tribal populations following Voegelin and Voegelin's (1977) classification and index of world languages, which is based on Krishnamurti's (1969) list of 22 Dravidian languages. Except for the Lambadi, who speak an Indo-European language, all the other eight tribal groups speak Dravidian languages, which are again divided into northern, central, and southern Dravidian branches. The central Dravidian branch is again bifurcated into major and minor groups (e.g., Kolami), and the southern Dravidian branch has Tamil and Malayalam, the two major languages spoken by the Irula and the Kadar, respectively.

Based on the hierarchical structure of the branching of Dravidian and non-Dravidian groups, scores were assigned. If a tribal group belongs to a similar subbranch of a major branch, it was given a score of 1; a score of 2 was given if the languages differ at the subbranch level. A score of 3 was given if the language difference is at the branch level. If the language difference is at the Dravidian versus non-Dravidian level, a score of 4 was assigned in the calculation of linguistic distance between the tribal groups.

Mantel statistics were used to examine the congruence between genetic, dermatoglyphic, anthropometric, linguistic, and geographic distance matrices. Given two distance matrices A and B, the Mantel test (1967) examines whether an association Z_{AB} exists between the elements of the two matrices. Following the methods of Dow and Cheverud (1985), Smouse et al. (1986), and Dow et al. (1987a,b), we computed Mantel correlations, partial correlations, and multiple correlations to examine

simultaneously the relationships between the various types of distance matrices. Significance of a given correlation is obtained through Mantel permutation procedures using the Mantel program (Relethford 1990). In this study, for the nine-population analysis 1000 permutations were used, whereas for the five-population analysis (anthropometric and related distances) and the six-population analysis (dermatoglyphic and related distances) all possible permutations were employed.

Results

The R matrix analysis of genetic variation among nine south Indian tribal groups is based on seven polymorphic loci representing nine allele frequencies ($ABO*A$, $ABO*B$, MN M, RH D, $PGM1*1$, $PGM1*2$, $PGD*A$, $ACP*A$, and $LDH*N$). The dispersal of the nine south Indian tribal groups along the first two scaled eigenvectors is shown in Figure 2. The first eigenvector, which explains 52.4% of the total variation, separates all seven Andhra Pradesh tribal groups from the Kadar of Kerala (Malayalam speakers) and the Irula of Tamil Nadu (Tamil speakers). The second axis explains 22.2% of the total variation and distinguishes between Andhra Pradesh and non–Andhra Pradesh tribes. It is interesting to note that the Mahaboobnagar Chenchu, although from Andhra Pradesh state, are a distinct group with respect to the other Andhra Pradesh tribal groups.

All the Telugu- (and Telugu-related) speaking groups align with the Lambadi, a population with known north Indian roots (non-Dravidian speakers) but inhabiting Andhra Pradesh for the last five centuries. Such a differentiation between the central (Telugu and Telugu-related groups and Kolami) and south Dravidian branches (Tamil and Malayalam) and the alignment of the Lambadi with the Andhra Pradesh tribal groups signify the historical and geographic-genetic distinction of the Irula and the Kadar from the Andhra Pradesh tribes. The F_{ST} value obtained for the nine tribal groups is 0.0412, indicating a moderately high level of microdifferentiation between these groups. However, when only the Dravidian-speaking Andhra Pradesh tribes (both Chenchu subgroups, Yanadi, Yerukula, Kolam, Koya) are considered, the F_{ST} value is 0.0218, and with the addition of the Lambadi tribe to the Andhra Pradesh tribes, F_{ST} decreases to 0.0212. These F_{ST} values further support the linguistic explanations.

The distribution pattern of nine alleles that underlie the separation of the nine subgroups along the first two scaled eigenvectors is shown in Figure 3. The relatively high frequency of the $PGM1*2$ allele contributes to the separation of the Irula from the other groups, whereas the isolation of the Kadar from the other groups is due to the combination
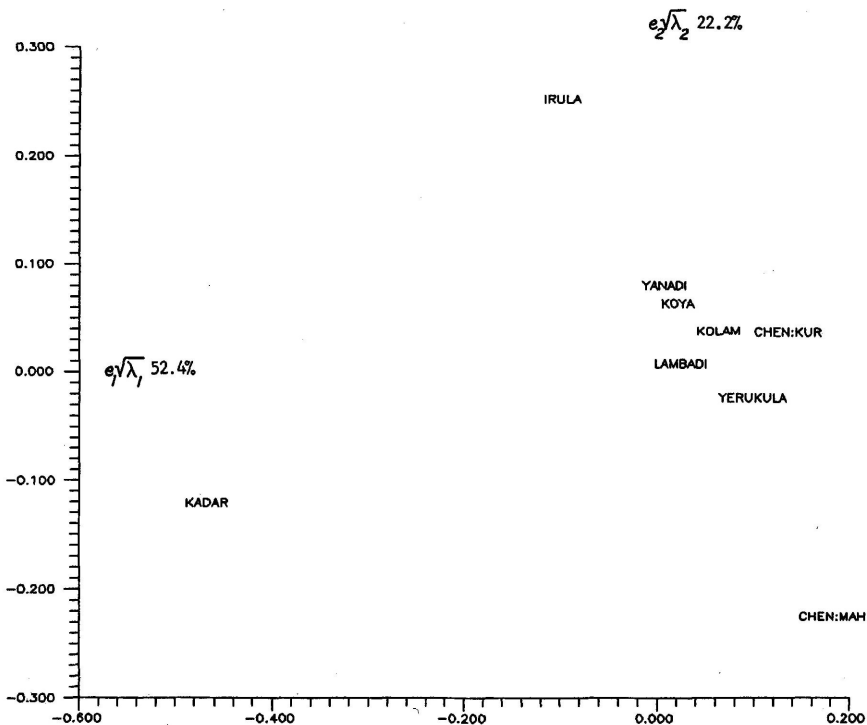
Figure 2. Least-squares reduction map of the nine South Indian populations based on seven polymorphic loci and nine allelic frequencies (*ABO*A*, *ABO*B*, MN *M*, RH *D*, *PGM1*1*, *PGM1*2*, *PGD*A*, *ACP*A*, and *LDH*N*).

of relatively high frequencies of the RH *D* and MN *M* alleles. The distinction of the Mahaboobnagar Chenchu group from other Andhra Pradesh groups is primarily due to high frequencies of the *PGM1*1* and *ACP*A* alleles. In fact, Ramesh et al. (1980) reported that the Mahaboobnagar Chenchu and the Kota tribe have a high *ACP*A* gene frequency in south India.

Figure 4 describes the functional relationship between the mean per locus heterozygosity and the distance r_{ii} of the region from the gene frequency centroid and shows the effects of differential systematic pressure among the nine south Indian tribal groups. In Figure 4 the Kolam, Yanadi, and Kadar are identified with negative departures from the theoretical predictions, thus showing their relatively increased isolation from external gene flow. The Koya and Irula exhibit similar trends, although with minor deviations; the other four groups of Andhra Pradesh lie above the theoretical regression line, indicating relative excess of gene flow from the "outside world."
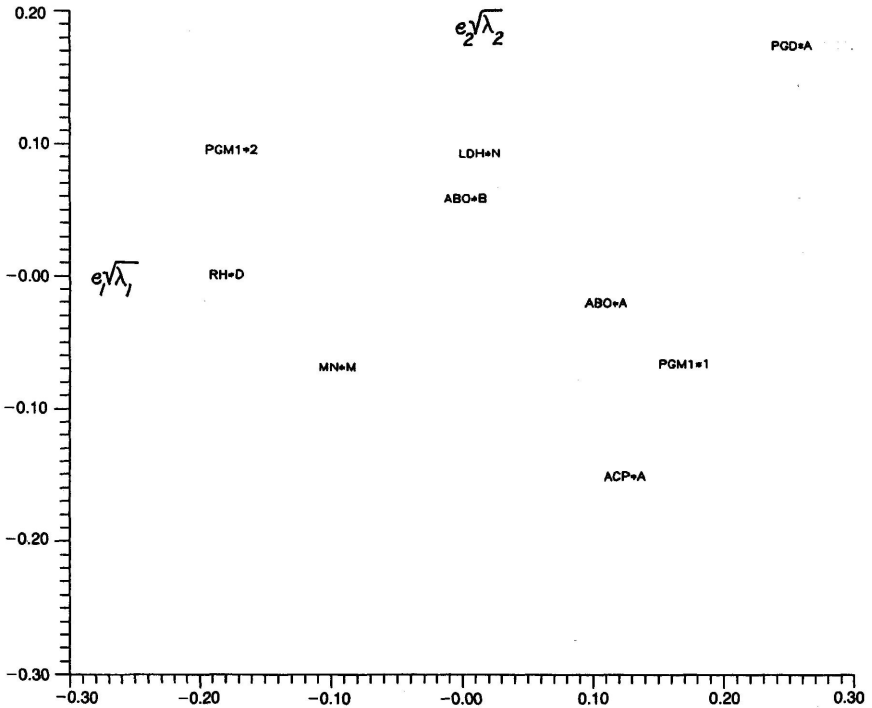
Figure 3. Distribution pattern of nine alleles that underlie the separation of the nine populations along the first two scaled eigenvectors.

The Kadar, Irula, and Mahaboobnagar Chenchu, with their increasing distances from the centroid, can be explained in the following way. The Kadar are believed to be one of the most ancient autochthons of the Annamalai and Trichur forests of southern India. With a small population size of about 1000 individuals (Ehrenfels 1952; Saha et al. 1974; Bhatia and Rao 1986), the Kadar are culturally and genetically isolated from other neighboring groups. By analyzing the genetic distances between the south Indian tribal populations, Saha et al. (1976) and Balakrishnan (1978) hypothesized that the Irula of the Nilgiri Hills were one of the more ancient groups. The isolation of the Mahaboobnagar Chenchu from the other Andhra Pradesh tribes can be explained by the considerable amount of gene flow between them and the neighboring populations (Macfarlane 1940; Ramesh et al. 1980).

The Koya speak Kui and the Kolam speak Kolami language, and they are isolated from their neighboring groups because of geography and language. The isolation of the Yanadi is also not surprising, given their settlement patterns along the coastal region, breeding pattern, mi-
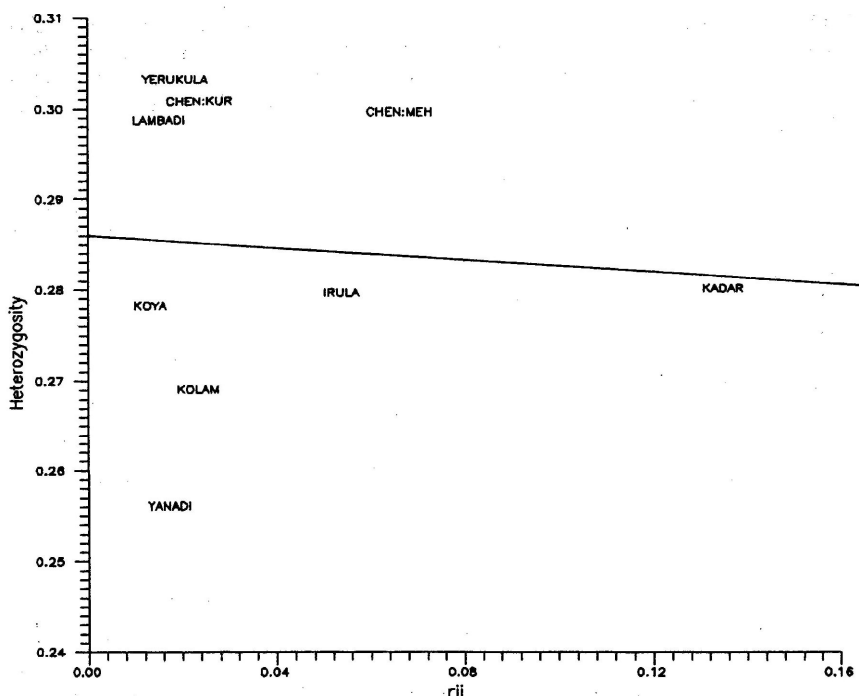
Figure 4. Regression of mean per locus heterozygosity H on distance from the gene distance centroid r_{ii}.

gration history, and low social standing among the tribal groups (Reddy et al. 1982; Vasulu and Pal 1989). The seminomadic lifestyles of the Yerukula and the Lambadi together with their social proximity to the caste populations suggest a greater amount of gene flow among them. The Kurnool Chenchu, who align with the Yerukula and Yanadi with regard to the level of heterozygosity, appear to have had genetic contact with their neighboring populations.

The genetic distances between the nine populations also corroborate the earlier findings of the relative distinctive features of the Kadar and the Irula from the other Andhra Pradesh tribes. The Mahaboobnagar Chenchu and the Kadar have the highest distance (0.2858), followed by the Kurnool Chenchu and the Kadar (0.2460). Comparatively higher distances were also observed between the Mahaboobnagar Chenchu and the Irula (0.220) and between the Yerukula and the Kadar (0.2197).

Mantel Statistics.   The degrees of concordance between various distance measures using Mantel statistics were evaluated at three levels: (1) genetic, geographic, and linguistic distances involving nine tribal pop-

Table 2. Correlations for Genetic (GEN), Geographic (GEO), and Linguistic (LIN) Distance Matrices; Partial Correlations between Two Matrices Controlling the Third Matrix; and Multiple Correlation obtained from Multiple Regression of Genetic Distance Matrix against Both Geographic and Linguistic Distance Matrices among Nine South Indian Tribal Populations

Distances Compared	Correlation (r)	Significance[a]
Correlations		
GEN × GEO	0.520	0.002
GEN × LIN	0.270	0.052
GEO × LIN	0.314	0.033
Partial correlations		
GEN × GEO (LIN)	0.476	0.005
GEN × LIN (GEO)	0.132	0.225
Multiple correlation		
GEN × GEO, LIN	0.532	0.003

a. Mantel test of probability (p).

ulations, (2) genetic, dermatoglyphic, geographic, and linguistic distances using six populations (both Chenchu subgroups, Yanadi, Irula, Kadar, Yerukula), and (3) genetic, anthropometric, geographic, and linguistic distances with reference to the five tribal groups (the previous six groups minus Yerukula).

Table 2 summarizes the results of various Mantel tests for the nine tribal groups. The correlation between genetics and geography is relatively high (0.520) and highly significant compared with the correlation between genetics and language (0.270), where significance is slightly more than 5%. Between geography and language a significant correlation of 0.314 is observed. The strong relationship between geography and genetics is further confirmed by the significant partial correlation (0.476) when language is kept constant. The level of correlation between genetics and language when geography is kept constant is low (0.132) and not significant. However, the relationship between the combined effects of geography and language (i.e., the correlation between geography and language is 0.314 and significant) and genetics is highly significant, with a multiple correlation value of 0.532. Only 28.3% of the variance in genetics is explained by the joint effects of language and geography.

The correlations relating to the dermatoglyphic, genetic, geographic, and linguistic distance measures for the six-population analysis are reported in Table 3. All the pairwise correlations between various distance measures are significant with variable magnitudes, ranging from moderate to high correlations, except the correspondence between dermatoglyphics and geography, where the significance is above the conventional level.

Table 3. Correlations for Genetic (GEN), Dermatoglyphic (DER), Geographic (GEO), and Linguistic (LIN) Distance Matrices; Partial Correlations between Two Matrices Controlling the Third Matrix; and Multiple Correlations Obtained from the Multiple Regressions of Genetic and Dermatoglyphic Distances on Geographic and Linguistic Distances among Six South Indian Tribal Groups

Distances Compared	Correlation (r)	p^a
Correlations		
DER × GEN	0.842	0.000
DER × GEO	0.423	0.061
DER × LIN	0.578	0.032
GEN × GEO	0.591	0.033
GEN × LIN	0.729	0.026
GEO × LIN	0.790	0.015
Partial correlations		
DER × GEO (LIN)	−0.066	0.602
DER × LIN (GEO)	0.439	0.063
GEN × GEO (LIN)	0.037	0.433
GEN × LIN (GEO)	0.530	0.021
Multiple correlations		
DER × GEO, LIN	0.580	0.028
GEN × GEO, LIN	0.730	0.021

The six populations examined in this analysis are the Mahaboobnayar Chenchu, Kurnool Chenchu, Yanadi, Irula, Kadar, and Yerukula.
a. Mantel test of probability.

The congruence between dermatoglyphics and genetics is very high ($r = 0.842$) and is highly significant ($p = 0.000$). Similar high correlations have been reported for female Eskimos (0.854) (Crawford and Duggirala 1992) and for Tlaxcaltecans (0.958) (Enciso 1983). However, comparatively lower correlations were observed by Neel et al. (1974) for the Yanomama (0.34) and by Dow et al. (1987a) for Solomon Islanders (males: 0.224; females: 0.125).

The partial correlations between dermatoglyphics and geography and between genetics and geography when language is kept constant, are negligible and nonsignificant, but the partial correlation between dermatoglyphics and language (0.439) and between genetics and language (0.530) when geography is kept constant are moderate and the latter is significant.

The interrelated effects of geography and language (i.e., the correlation between geography and language is 0.790 and significant) on dermatoglyphics and genetics are appreciable, as shown by the significant multiple correlation values of 0.580 and 0.730. Geography and language together explain 33.7% of the variation in dermatoglyphics, whereas the proportion of variance explained in genetics by these two predictors is moderately high (53.2%).

Table 4. Correlations for Genetic (GEN), Anthropometric (ANTH), Geographic (GEO), and Linguistic (LIN) Distance Matrices; Partial Correlations between Two Matrices Controlling the Third Matrix; and Multiple Correlations Obtained from the Multiple Regressions of Genetic and Dermatoglyphic Distances on Geographic and Linguistic Distances among Five South Indian Populations

Distances Compared	Correlation (r)	p[a]
Correlations		
ANTH × GEN	0.587	0.025
ANTH × GEO	0.712	0.000
ANTH × LIN	0.488	0.042
GEN × GEO	0.707	0.025
GEN × LIN	0.657	0.042
GEN × LIN	0.843	0.008
Partial correlations		
ANTH × GEO (LIN)	0.640	0.017
ANTH × LIN (GEO)	−0.297	0.857
GEN × GEO (LIN)	0.378	0.177
GEN × LIN (GEO)	0.159	0.311
Multiple correlations		
ANTH × GEO, LIN	0.742	0.008
GEN × GEO, LIN	0.720	0.042

The five populations examined in this analysis are the Mahaboobnagar Chenchu, Kurnool Chenchu, Yanadi, Irula, and Kadar.

a. Mantel test of probability.

Table 4 presents the correlations between anthropometrics, genetics, language, and geography for the five-population analysis. All the bivariate correlations are significant, and the magnitude of the correlations ranges from moderate to fairly high. The correspondence between anthropometrics and geography is substantial and highly significant ($r = 0.712$, $p = 0.000$). Most earlier studies of South Amerindian tribes report high correspondence between anthropometrics and geography, and the Yanomama Indians exhibit a particularly high correlation of 0.801 (Spielman 1973). The partial correlations fail to provide any additional information in this data set, except that the correlation between anthropometrics and geography remains high (0.640) and significant when the effect of language is controlled. Multiple correlation analysis of both anthropometric (0.742) and genetic distances (0.720) show highly significant relationships with the combined effects of geography and language, where the correlation between geography and language is 0.843 ($p = 0.008$). Fairly substantial components of variances in anthropometrics (55%) and genetics (51.2%) are explained by the interrelated effects of geography and language, respectively.

Discussion and Conclusion

In the absence of a complete archeological record, tracing genetic relationships and estimating genetic microdifferentiation of contemporary populations from ancestral groups remain speculative. Earlier investigations based on anthroposcopic traits (Guha 1937; Sarkar 1954; Malhotra 1978; and others) and genetic markers have documented the distinctive features and genetic differentiation of tribal and nontribal groups in all regions of India (Balakrishnan 1978; Roychoudhury 1984a, 1992). These studies have also suggested that the tribal groups are composed of differential and variable amounts of European, Asian, and Australian aborigine admixture, resulting from successive migrations and subsequent genetic amalgamations (Malhotra 1978; Balakrishnan 1984; Roychoudhury 1992). The question of whether or not south Indian tribal groups are autochthons is still controversial, and Labie et al. (1989) proposed a unicentric evolutionary theory for their origin.

Most present-day tribal groups still dwell in relatively isolated forests and valleys. As a result of this isolation, the southern, western, and eastern tribal groups are relatively more heterogeneous (Balakrishnan 1978; Roychoudhury 1984a, 1992). By using multiple genetic characters, Simmons (1976), Kirk (1976), and Roychoudhury (1984c) examined the question of the origin of the so-called proto-Australoid elements and the relationship of south Indian tribal groups with the Veddas and Australian aborigines. They documented the distinctiveness of south Indian tribal groups and came to the conclusion that these tribal groups are not related to the Australian aborigines.

We have examined the genetic structure and genetic affinities of nine south Indian tribal groups (two Chenchu subgroups, Yanadi, Yerukula, Kolam, Koya, Lambadi, Irula, and Kadar) using multivariate analyses (at three levels) of genetic, morphological, geographic, and linguistic data. Furthermore, we have explored how effective geography and language were in the geographic distribution of genes and the structure of these groups.

Because of the paucity of comparable data on genetic loci (seven), number of populations for dermatoglyphics (six), and number of populations for anthropometric variables (five), our analyses were limited. The diverse sources of our data require cautious interpretation of these results. As shown by the genetic relationship analysis, the Andhra Pradesh tribes are distinct from the non–Andhra Pradesh tribes of the Kadar and the Irula. The Lambadi are linguistically different and are recent migrants to their present locale. As a result of this ethnohistory, these groups show genetic affinities with their geographic neighbors, probably as a result of gene flow. The clustering of six Andhra tribes (both Chenchu subgroups, Yanadi, Yerukula, Kolam, Koya) and the smaller genetic

differentiation compared with the F_{ST} of all nine populations indicate that the between-group diversity of the Andhra Pradesh tribes is not enormous. Murthy et al. (1993), who measured the level of heterozygosity of Andhra Pradesh tribal groups at different levels, reached similar conclusions.

The relationship between heterozygosity and distance from the centroid indicates that the Kadar are the most distant group from the centroid, followed by the Mahaboobnagar Chenchu and the Irula. The isolation of the Kadar and the Irula is not surprising because tribes of the Nilgiri Hills, such as the Kurumba, Irula, and Kadar, are considered the earliest occupants of this region (Balakrishnan 1978, 1984; Saha et al. 1976). The Kadar show some rare allele variants, such as *PGM*1,6*, and LDH *CAL1*, and there is little to negligible gene flow from the Irula to the Kadar in the Annamalai forest hills (Saha et al. 1976; Sirajuddin and Balakrishnan 1991). The isolation of the Kadar can also be attributed to stochastic processes when their small population size (approximately 1000 persons) is considered.

Our analyses reveal a greater amount of gene flow among the Yerukula and the Lambadi and in the two Chenchu subgroups. The Yerukula and the Lambadi are seminomadic and are plains dwellers with closer contact with the other caste populations.

The Mahaboobnagar Chenchu in the Mannanoor, Farahabad, and Appaipally settlements have been observed to have a high incidence of intertribal marriages (Sirajuddin 1985). Admixture among the Mahaboobnagar Chenchu has also been shown by Macfarlane (1940) and Ramesh et al. (1980), whereas Murthy et al. (1993) have pointed out more gene flow among the plains-dwelling tribes of Andhra Pradesh. Comparatively, the Yanadi appear to be more homogeneous than the other tribal groups. Although the Yanadi are mostly plains dwellers (because of their low social standing and because their habitation is mostly restricted to islands and coastal regions), they are reproductively isolated from their neighbors (Reddy et al. 1982; Vasulu and Pal 1989). Thus they may also experience the effects of stochastic processes.

The three-tier correlation analysis with different data sets adds further dimension to the understanding of the genetic structure of south Indian tribal groups. Because genetics is associated substantially with anthropometrics and dermatoglyphics, all three data sets are valid indicators of population structure, although with differential magnitudes. Earlier studies comparing genetics and anthropometrics have shown a range of correlations, from -0.061 in the Solomon Islands (Dow et al. 1987a) to 0.83 among the Caingang (Salzano et al. 1980), whereas dermatoglyphics and genetics have correlations that range from -0.069 in Hvar (Šimić and Rudan 1990) to 0.970 among Tlaxcaltecans (Enciso 1983). It is im-

portant to remember that the choice of characters and the methods for computing correlations vary from comparison to comparison.

Given that geography and language are correlated, we have observed consistently high significant correlations of geography and language on the genetic, dermatoglyphic, and anthropometric structure of these groups. Such significant overall congruence between genetics and dermatoglyphics with language and geography has been observed in Eskimo and Amerindian populations (Crawford and Duggirala 1992). But Dow et al. (1987b) found insignificant multiple correlations in Solomon Island groups with regard to anthropometrics and blood genetics but substantial association involving dermatoglyphics.

The partial correlation results point out that, when language effects are removed, both genetics (i.e., nine-population analysis) and anthropometrics are significantly associated with geography, as seen from the bivariate correlations between anthropometrics and geography and between genetics and geography. This contrasts with Dow et al.'s (1987b) observations on the Solomon Island groups and the suggestion that, relative to monogenetic traits, anthropometrics depict better interpopulation relationships within the context of geography (Dow and Chevrud 1985). However, dermatoglyphics do not indicate significant association with geography, as shown by both bivariate and partial correlation. Although language is appreciably associated with dermatoglyphics, anthropometrics, and genetics (in the six- and five-group analyses) individually in different distance comparisons, only genetics correlates well with language in the absence of geographic effects (six-population analysis).

Sokal (1988), in his study of European populations, found that geographic differentiation influences genetic structure more than linguistic differentiation does. Our dermatoglyphic results contrast with other results that suggest that dermatoglyphics exhibit superiority over other distance measures in understanding linguistic structures of populations (Dow et al. 1987a,b). A word of caution must be added here because our linguistic distance analysis is based on the robust and broad classification of Dravidian language into three main categories: central, northern, and southern branches. We have subclassified the southern Dravidian branch into Tamil speakers (Irula) and Malayalam speakers (Kadar), and the central branch has been bifurcated into major and minor categories following Krishnamurti (1969). A different approach with fewer subgroups may elevate correlations. On the other hand, the contemporary linguistic features of some of the tribal groups may have been influenced by the ancient literary traditions of the neighboring caste populations.

Geographic factors appear to be prime determinants of the genetic differentiation of south Indian tribal groups. This conclusion has been supported by the studies of Balakrishnan (1984) and Roychoudhury (1984a,c, 1992), even though geographic variables per se were not used

by these researchers for the analysis both at the regional and the all India levels. Although language has limited effect, the geographically patterned genetic differentiation is important among south Indian tribal groups. Because geography limits migration, one expects high correlation between geography and biological distance at some intermediate level of migration (Jorde 1980). Our study suggests that historical processes of gradual expansion and microdifferentiation are geographically patterned among south Indian tribal groups. This study supports the earlier preliminary findings on the Chenchu reported by Reddy et al. (1982) and Sirajuddin and Balakrishnan (1991). However, we would like to expand this study to other states and increase the number of variables and populations to test the veracity of our current findings. In the future we would also like to examine the genetic diversity of Indian tribal groups in relation to the genetic diversity of the neighboring caste populations to test tribal autochthony.

Literature Cited

Balakrishnan, V. 1978. A preliminary study of genetic distances among some populations of the Indian subcontinent. *J. Hum. Evol.* 7:67–75.

Balakrishnan, V. 1984. Admixture as an evolutionary force in populations of the Indian subcontinent. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Human Genetics and Adaptation*, K.C. Malhotra and A. Basu, eds. Calcutta, India: Indian Statistical Institute, 103–145.

Balakrishnan, V., and L.D. Sanghvi. 1968. Distance between populations on the basis of attribute data. *Biometrics* 24:859–865.

Bhalla, V. 1984. Genetic diversity in tribal populations of India: Illustrative maps showing the distribution of ABO, MNSs and Rh polymorphisms. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Human Genetics and Adaptation*, K.C. Malhotra and A. Basu, eds. Calcutta, India: Indian Statistical Institute, 420–447.

Bhatia, H.M., and V.R. Rao. 1986. *Genetic Atlas of Indian Tribes*. Parel, Bombay, India: Institute of Immunohemotology, Indian Council of Medical Research.

Blake, N.M., A. Ramesh, M. Vijayakumar et al. 1981. Genetic studies on some tribes of the Telangana region, Andhra Pradesh, India. *Acta Anthropogenet.* 5:41–56.

Chakraborty, R., and S. Yee. 1973. Five tribes of Orissa, India: Anthropometry and kinship. *Hum. Hered.* 23:301–307.

Chakravarti, M.R., and D.P. Mukherjee. 1964. Dermatoglyphic affinities of the tribes and castes or Nilgiri hills. *Z. Morphol. Anthropol.* 55:335–356.

Char, K.S.N., P. Lakshmi, K.B. Gopalam et al. 1989. Genetic differentiation among some endogamous populations of Andhra Pradesh, India. *Am. J. Phys. Anthropol.* 78:421–429.

Crawford, M.H., and R. Duggirala. 1992. Digital dermatoglyphic patterns of Eskimo and Amerindian populations: Relationships between geographic, dermatoglyphic, genetic, and linguistic distances. *Hum. Biol.* 64(5):683–704.

Dow, M.M., and J.M. Cheverud. 1985. Comparison of distance matrices in studies of population structure and genetic microdifferentiation: Quadratic assignment. *Am. J. Phys. Anthropol.* 68:367–373.

Dow, M.M., J.M. Cheverud, and J.S. Friedlaender. 1987a. Partial correlations of distance matrices in studies of population structure. *Am. J. Phys. Anthropol.* 72:343–352.

Dow, M.M., J.M. Cheverud, J.G. Rhoads et al. 1987b. Statistical comparison of patterns of biological and cultural-historical variation. In *The Solomon Islands Project: A Long Term Study of Health, Human Biology, and Culture Change*, J.S. Friedlaender, ed. Oxford: Clarendon Press, 265–280.

Ehrenfels, U.R. 1952. *Kadar of Cochin*. Madras, India: University of Madras.

Enciso, V.B. 1983. Estimates of genetic distances and population structure from blood and dermatoglyphic data. M.A. thesis, University of Kansas, Lawrence.

Ghosh, A.K., R.L. Kirk, S. Joshi et al. 1977. A population genetic study of the Kota in the Nilgiri Hills, south India. *Hum. Hered.* 27:225–241.

Ghosh, G.C. 1975. Towards dermatoglyphic study of the Dravidian speaking tribes of central India. In *Bio-Anthropological Research in India*, H.K. Rakshit, ed. Calcutta, India: Anthropological Survey of India, 185–188.

Goud, J.D., and P.R. Rao. 1979. Genetic distance among the five tribal populations of Andhra Pradesh, south India. *Anthropol. Anz.* 37:1–9.

Guha, B.S. 1935. *The Racial Affinities of the People of India*. New Delhi, India: Census of India.

Guha, B.S. 1937. *An Outline of the Racial Ethnology of India*. Calcutta, India: Indian Science Congress Association.

Guha, B.S. 1944. *Racial Elements in the Population*. Bombay, India: Oxford University Press.

Guha, B.S. 1955. Indian aborigines and who they are. In *The Tribal People of India*. New Delhi, India: Publications Division, Ministry of Information and Broadcasting, Government of India, 28–35.

Haimendorf, V.C. 1943. *The Chenchus*. London, England: Macmillan.

Haimendorf, V.C. 1982. *The Tribes of India*. Los Angeles, CA: University of California Press.

Harpending, H., and T. Jenkins. 1973. Genetic distance among southern African populations. In *Methods and Theories of Anthropological Genetics*, M.H. Crawford and P.L. Workman, eds. Albuquerque, NM: University of New Mexico Press, 137–165.

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

Jorde, L.B. 1980. The genetic structure of subdivided populations: A review. In *Current Developments in Anthropological Genetics*, v. 1, *Theory and Methods*, J.H. Mielke and M.H. Crawford, eds. New York: Plenum, 135–208.

Kirk, R.L. 1976. Serum protein and enzyme markers as indicators of population affinities in Australia and western Pacific. In *The Origin of the Australians*, R.L. Kirk and A.G. Thorne, eds. Canberra, Australia: Australian Institute of Aboriginal Studies, 329–346.

Kirk, R.L., L.Y.C. Lai, G.H. Vos et al. 1962. The blood and serum groups of selected populations in south India and Ceylon. *Am. J. Phys. Anthropol.* 20:485–497.

Krishnamurti, B.H. 1969. Comparative Dravidian studies. In *Current Trends in Linguistics*, v. 5, Thomas A. Sebeok, ed. Paris, France: Mouton, 309–333.

Labie, D., R. Srinivas, O. Dunda et al. 1989. Haplotypes in tribal Indians bearing the sickle gene: Evidence for the unicentric origin of B^s mutation and the unicentric origin of the tribal populations of India. *Hum. Biol.* 61(4):479–491.

Macfarlane, W. 1940. Blood grouping in Deccan and Eastern Ghats. *J. Asiat. Soc. Bengal* 6:39–49.

Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* 2:49–55.

Malhotra, K.C. 1978. Morphological composition of the people of India. *J. Hum. Evol.* 7:45–63.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.

Murthy, J.S., B. Muralidhar, J.D. Goud et al. 1993. Hierarchial gene diversity and genetic structure of tribal populations of Andhra Pradesh, India. *Am. J. Phys. Anthropol.* 90:169–183.

Narahari, S. 1982. A genetic study among the Yerukulas of Andhra Pradesh. PhD thesis, Sri Venkateswara University, Tirupathi, India.

Neel, J.V., R. Rothammer, and J.C. Lingoes. 1974. The genetic structure of a tribal population, the Yanomama Indians. X. Agreement between representations of village distances based on different sets of characteristics. *Am. J. Hum. Genet.* 26:281–303.

Pathak, R.K., and S.S. Kaul. 1991. Nonmetric cranial variation and biological distance between samples of six nontribal populations of India. *Int. J. Anthropol.* 6(1):81–87.

Pingle, U., B.N. Mukherjee, and S.K. Das. 1981. A genetical study of five tribal groups of Andhra Pradesh, India. *Z. Morphol. Anthropol.* 72:339–348.

Raghavaiah, V. 1962. *The Yanadis*. New Delhi, India: Adimjati Sevak Sangh.

Ramesh, A., J.S. Murthy, and N.M. Blake. 1979. Genetic studies on the Kolams of Andhra Pradesh, India. *Hum. Hered.* 29:339–348.

Ramesh, A., N.M. Blake, M. Vijayakumar et al. 1980. Genetic studies on Chenchu tribe of Andhra Pradesh, India. *Hum. Hered.* 30:291–298.

Reddy, A.P., B.N. Mukherjee, K.C. Malhotra et al. 1982. A serological and biochemical genetic study among the coastal and plateau Yanadis, a tribal population of Andhra Pradesh. *Homo* 33(4):174–182.

Reddy, V.R., and K.N. Subramanyam. 1985. Hand clasping and arm folding behavior and somatometric traits of the Yanadis of southeastern Andhra Pradesh. *J. Ind. Anthropol. Soc.* 15(2–3):245–248.

Relethford, J. 1990. Mantel: A microcomputer program for computing the Mantel probability between distance matrix elements. Oneonta, NY: Department of Anthropology, State University of New York, College at Oneonta.

Roberts, D.F., S.S. Papiha, G.N. Rao et al. 1980. A genetic study of some Andhra Pradesh populations. *Ann. Hum. Genet.* 24:680–685.

Roychoudhury, A.K. 1984a. Genetic relationships of Indian populations. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Human Genetics and Adaptation*, K.C. Malhotra and A. Basu, eds. Calcutta, India: Indian Statistical Institute, 146–174.

Roychoudhury, A.K. 1984b. Genetic relationships between Indian populations and their neighbors. In *The People of South Asia*, J.R. Lukacs, ed. New York: Plenum, 283–291.

Roychoudhury, A.K. 1984c. Genetic relationship between Indian tribes and Australian aboriginals. *Hum. Hered.* 34:314–320.

Roychoudhury, A.K. 1992. Genetic relationships of the populations in eastern India. *Ann. Hum. Biol.* 19(5):489–501.

Saha, N., R.L. Kirk, S. Shanbhag et al. 1974. Genetic studies among the Kadar of Kerala. *Hum. Hered.* 24:198–218.

Saha, N., R.L. Kirk. S. Shanbhag et al. 1976. Population genetic studies in Kerala and Nilgiris (southwest India). *Hum. Hered.* 26:175–197.

Salzano, F.M., S.M. Callegari-Jacques, M.H.L.P. Franco et al. 1980. The Caingang revisited: Blood genetics and anthropometry. *Am. J. Phys. Anthropol.* 53:513–524.

Sarkar, S.S. 1954. *The Aboriginal Races of India*. Calcutta, India: Bookland.

Sarkar, S.S., G.S. Ray, M.R. Chakravarti et al. 1959. *A Physical Survey of the Kadar of Kerala*. Memoir 6. Calcutta, India: Department of Anthropology, Government of India.

Šimić, D., and P. Rudan. 1990. Isolation by distance and correlation analysis of distance measures in the study of population structure: Examples from the island of Hvar. *Hum. Biol.* 62:113–130.

Simmons, R.T. 1976. The biological origin of Australian aboriginals. In *The Origin of the Australians*, R.L. Kirk and A.G. Thorne, eds. Canberra, Australia: Australian Institute of Aboriginal Studies, 307–328.

Simmons, R.T., J.J. Graydon, N.M. Semple et al. 1953. A genetical survey in Chenchu, south India: Blood, taste, and secretion. *Med. J. Austr.* 1:497–503.

Sirajuddin, S.M. 1985. Human biology of the Chenchus of Andhra Pradesh: A demo-morpho-genetic study. PhD thesis, Karnatak University, Dharwad, India.

Sirajuddin, S.M., and V. Balakrishnan. 1991. Multivariate analysis of dermatoglyphic data and their correspondence with anthropometric and serogenetic data: The Chenchus. In *Dermatoglyphics Today*, B.M. Reddy, S.B. Roy, and B.N. Sarkar, eds. Calcutta, India: IBRAD, ASI, and ISI, 300–310.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35(4):627–632.

Sokal, R.R. 1988. Genetic, geographic, and linguistic distances in Europe. *Proc. Natl. Acad. Sci. USA* 85:1722–1726.

Spielman, R.S. 1973. Differences among Yanomama Indian villages: Do the patterns of allele frequencies, anthropometrics, and map locations correspond? *Am. J. Phys. Anthropol.* 39:461–480.

Sreenath, J. 1977. A study of palmar dermatoglyphics among the Yanadis in and around Nellore, A.P. M.Sc. dissertation, Sri Venkateswara University, Tirupati, India.

Undevia, J.V., R.K. Gulati, P.K. Sukumaran et al. 1981. Genetic variation in Tamil Nadu. In *Biology of the People of Tamil Nadu*, L.D. Sanghvi, V. Balakrishnan, and I. Karve, eds. Calcutta, India: Indian Society of Human Genetics and Indian Anthropological Society.

Vasulu, T.S., and M. Pal. 1989. Size and shape components of anthropometric differences among Yanadis. *Ann. Hum. Biol.* 16:449–462.

Voegelin, C.F., and F.M. Voegelin. 1977. *Classification and Index of the World's Languages*. New York: Elsevier.