

---

## ***Genomic Diversity at Thirteen Short Tandem Repeat Loci in a Substructured Caste Population, Golla, of Southern Andhra Pradesh, India***

B. MOHAN REDDY,<sup>1</sup> GUANGYUN SUN,<sup>2</sup> JAVIER RODRIGUEZ LUIS,<sup>3</sup> MICHAEL H. CRAWFORD,<sup>4</sup> NATABAR SHYAM HEMAM,<sup>5</sup> AND RANJAN DEKA<sup>2</sup>

**Abstract** Genomic diversity based on 13 short tandem repeat (STR) loci was studied in seven population groups of a substructured Golla caste from Chittoor district in southern Andhra Pradesh, India. These groups are traditionally pastoral, culturally homogeneous, and strictly endogamous. Blood samples were drawn from 317 individuals from 30 Golla villages. The 13 STR loci analyzed in five standard multiplex polymerase chain reactions were: (1) *CSF1R*, *TH01*, and *PLA2A*; (2) *F13A1*, *CYP19*, and *LPL*; (3) *D21S1446* and *D21S1435*; (4) *D20S481*, *D20S473*, and *D20S604*; and (5) *D5S1453* and *D6S1006*. The average heterozygosity was found to be low among the Golla subgroups (0.64–0.70) in comparison to that of groups at the upper levels of the hierarchy. The coefficient of gene differentiation was found to be moderate (average  $G_{ST} = 0.031$ ; range between 0.018 and 0.049 among the loci) when compared to that observed for a similar class of markers among populations with relatively higher levels of hierarchy, for example, among castes. It is, however, much higher when compared to the average observed for Indian caste and tribal populations, based on classical markers. Genetic distance measures revealed clusters of populations that are consistent with the known ethnohistorical and geographical backgrounds of the groups. We claim that these hypervariable markers are quite useful in understanding the process of substructuring within the Indian castes, leading to the formation of smaller breeding isolates, the basic Mendelian units within which microevolutionary forces operate.

One of the most remarkable features of the Indian population structure is its clear division into strictly defined endogamous groups. At the broadest level, the pop-

<sup>1</sup>Anthropology and Human Genetics Unit, Indian Statistical Institute, Calcutta, India.

<sup>2</sup>Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio.

<sup>3</sup>Departmento Anthropoloxia, University of Santiago, Santiago, Spain.

<sup>4</sup>Laboratory of Biological Anthropology, Department of Anthropology, University of Kansas, Lawrence, Kansas.

<sup>5</sup>Center for Studies in Demography and Ecology, University of Washington, Seattle, Washington.

*Human Biology*, April 2001, v. 73, no. 2, pp. 175–190.

Copyright © 2001 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: STR LOCI, SUBCASTES, GENETIC DIFFERENTIATION, SHORT EVOLUTIONARY HISTORY

ulation is subdivided into tribes and castes, while the castes are further divided along geographic, linguistic, and religious lines. However, unique and fundamental to the Indian population structure is the existence of endogamous subcastes within many of these castes in any particular region or linguistic area. These subcastes are usually characterized by a high degree of isolation, small effective population size, and high intensity of inbreeding. These conditions are both conducive and prerequisite to the process of rapid genetic microdifferentiation. In most cases, these subcastes have likely evolved from a common parental stock as a result of different processes of fission. For example, Basu (1969), Malhotra (1978a and 1978b; 1979), Reddy (1982), and Reddy and Gadgil (1999) have documented many such processes among the Indian castes. However, given that the caste system in India is only 2000 to 3000 years old, the evolutionary history of these subpopulations might have been relatively short. Nevertheless, it is basically at this level of Mendelian units that the forces of evolution operate, and hence it is imperative to study microevolution among such populations. Many earlier attempts have been made to understand microdifferentiation among such populations using traditional markers as well as other biological variables (Karve and Malhotra 1968; Basu 1969; Malhotra 1978a and 1978b; Malhotra et al. 1978; Reddy et al. 1987, 1989, 1995; Murty et al. 1993; Sirajuddin et al. 1994; Ramana et al. 1996). However, the resolving power of such markers is expected to be low and unreliable, especially at the level of endogamous subcastes.

With the identification of highly variable DNA markers, particularly the short tandem repeat (STR) sequences, and with the development of rapid screening techniques using polymerase chain reaction (PCR) amplification, the possibilities of understanding genetic structure and microdifferentiation accurately among the local subdivided populations have brightened enormously. STRs are a subclass of microsatellite loci whose repeat motifs are 2 to 6 base pairs (bp) long. The properties associated with STR loci, such as high mutation rates ( $\sim 1/1000$  gametes; Weber and Wong 1993), large number of alleles, high heterozygosities, and frequent occurrence in the genome, make them useful markers for studying genetic microdifferentiation among such populations. In fact, many successful attempts have been made to use this class of markers for understanding the evolutionary relationships among global populations, both at the racial and the continental levels (Deka et al. 1995; Nei and Takezaki 1996; Jorde et al. 1997; Shriver et al. 1997). An attempt has also been made recently to decipher the patterns of population substructure at the continental and global levels using STR markers (Eller 1999). However, to the best of our knowledge, their use in understanding the short evolutionary history at the level of subcastes has yet to be tested. It is with this primary objective that we analyzed 13 STR loci in the seven Golla subcastes. We report here the findings on the nature and extent of gene diversity among the subpopulations in comparison to those at the higher levels of caste hierarchy.

## **Materials and Methods**

**Population.** The Gollas are the great pastoral caste of Telugu people of Andhra Pradesh. Tradition traces their descent from Lord Krishna (Thurston 1909). Their traditional occupation has been tending sheep and cattle and selling milk, although many of them have now acquired land and engage in agriculture as well. The Gollas are distributed throughout Andhra Pradesh. People belonging to this tradition are also found in other parts of India, in different linguistic and geographical areas. They are known by different names and probably have different origins.

Extensive interviews with the elders of this community suggest that there are supposed to be 12 endogamous subcastes within the Gollas. Thurston (1909) noted only nine such sects, probably because the distributions of some of these subcastes are highly localized and have escaped the attention of the ethnographer. We encountered only seven such groups in our exploratory survey covering a large area, and none of the elders of this community could give a complete list of all 12 subcastes. However, these people still retain their traditional patterns intact and maintain a high degree of endogamy within the subcastes. The exchange of mates between Golla subcastes has just started, so it is restricted to the youngest generation. It is still found to be <1% per generation, as the preliminary investigations during the blood collection revealed. Like all other traditional populations of southern India, Gollas prefer consanguineous marriages and village endogamy, with marriage contacts usually restricted to a small radius. Crude estimates of the inbreeding coefficient, based on the consanguinity status of the parents of subjects from whom blood samples were collected, range between 0.0175 among the Karnam to 0.0354 among the Pokanati, with an average of 0.0272 for the Golla. We therefore restricted our study to a culturally homogeneous area called Chittoor district in Andhra Pradesh and collected samples from all the subgroups found in this area. We found seven subcastes in this area, including the Kurava, whose position among the Gollas is currently disputed. Each subcaste is distinguished by a unique variant of their traditional occupation, as well as by certain microcultural variations in dress pattern and marriage ceremonies. Barring Thurston's (1909) ethnographic account, there is very little that can be drawn from literature on the ethnography and origin of these people. The impressions gathered from the elders of the community also did not provide any clear answers.

**Blood Samples.** With the help of trained lab technicians 5–10 mL of intravenous blood was drawn into 15-mL centrifuge tubes, using EDTA as anticoagulant, from a total of 317 unrelated adults. These subjects belonged to seven subcastes (Table 1) and were drawn from 30 villages distributed in 9 taluks (an administrative or revenue unit below the level of a district) of Chittoor district (Figure 1). The blood samples were stored in Thermocol boxes filled with suffi-

**Table 1.** Allele Frequencies for the 13 STR Loci in the 7 Golla Populations (Alleles Are Designated by their Repeat Numbers Except in Case of D21S1446)

<i>Locus</i>	<i>Allele</i>	<i>Doddi (DO)</i>	<i>Erra (ER)</i>	<i>Karnam (KA)</i>	<i>Pokanati (PK)</i>	<i>Puja (PU)</i>	<i>Punugu (PN)</i>	<i>Kurava (KU)</i>
<i>CSF1R</i>	8	0.008	0.000	0.000	0.009	0.000	0.000	0.000
	9	0.030	0.012	0.000	0.019	0.000	0.016	0.000
	10	0.276	0.286	0.255	0.189	0.135	0.339	0.222
	11	0.224	0.286	0.309	0.387	0.269	0.371	0.333
	12	0.418	0.345	0.372	0.264	0.539	0.274	0.444
	13	0.045	0.060	0.064	0.123	0.058	0.000	0.000
	14	0.000	0.012	0.000	0.009	0.000	0.000	0.000
<i>n</i>		134	84	94	106	52	62	18
<i>THO1</i>	6	0.188	0.337	0.231	0.276	0.212	0.224	0.500
	7	0.116	0.058	0.164	0.147	0.077	0.259	0.150
	8	0.080	0.198	0.135	0.130	0.058	0.069	0.250
	9	0.486	0.291	0.298	0.319	0.519	0.362	0.100
	9,3	0.109	0.105	0.173	0.129	0.135	0.086	0.000
	10	0.022	0.000	0.000	0.000	0.000	0.000	0.000
<i>n</i>		138	86	104	116	52	58	20
<i>PLA2A</i>	10	0.000	0.000	0.000	0.000	0.019	0.000	0.000
	11	0.471	0.442	0.450	0.535	0.539	0.625	0.611
	12	0.094	0.140	0.050	0.079	0.039	0.036	0.111
	13	0.007	0.012	0.040	0.035	0.019	0.000	0.056
	14	0.138	0.163	0.190	0.175	0.231	0.054	0.111
	15	0.217	0.128	0.160	0.088	0.115	0.214	0.056
	16	0.073	0.116	0.110	0.070	0.039	0.071	0.056
	17	0.000	0.000	0.000	0.018	0.000	0.000	0.000
<i>n</i>		138	86	100	114	52	56	18
<i>F13A1</i>	3	0.008	0.000	0.000	0.000	0.000	0.000	0.000
	3,2	0.087	0.181	0.138	0.208	0.212	0.250	0.389
	4	0.087	0.028	0.011	0.009	0.000	0.000	0.000
	5	0.548	0.486	0.394	0.387	0.519	0.462	0.278
	6	0.159	0.194	0.330	0.217	0.135	0.231	0.278
	7	0.103	0.111	0.128	0.179	0.115	0.058	0.056
	12	0.000	0.000	0.000	0.000	0.019	0.000	0.000
	13	0.008	0.000	0.000	0.000	0.000	0.000	0.000
<i>n</i>		126	72	94	106	52	52	18
<i>CYP19</i>	5	0.237	0.282	0.141	0.278	0.250	0.185	0.188
	6	0.542	0.539	0.587	0.546	0.577	0.500	0.625
	7	0.017	0.013	0.054	0.019	0.058	0.000	0.000
	8	0.017	0.000	0.000	0.000	0.000	0.000	0.000
	9	0.000	0.000	0.011	0.000	0.000	0.259	0.000
	10	0.153	0.167	0.207	0.157	0.115	0.056	0.188
	11	0.017	0.000	0.000	0.000	0.000	0.000	0.000
	12	0.017	0.000	0.000	0.000	0.000	0.000	0.000
<i>n</i>		118	78	92	108	52	54	16
<i>LPL</i>	9	0.008	0.000	0.000	0.000	0.000	0.000	0.000
	10	0.375	0.369	0.402	0.417	0.404	0.386	0.222
	11	0.250	0.238	0.244	0.176	0.250	0.136	0.222
	12	0.313	0.202	0.232	0.333	0.308	0.364	0.444

**Table 1.** Continued

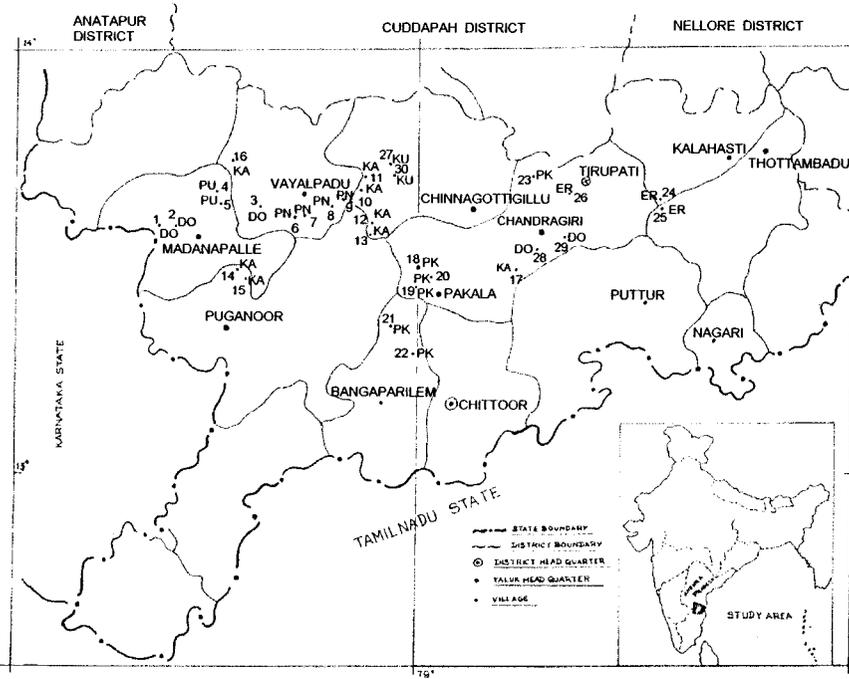
<i>Locus</i>	<i>Allele</i>	<i>Doddi (DO)</i>	<i>Erra (ER)</i>	<i>Karnam (KA)</i>	<i>Pokanati (PK)</i>	<i>Puja (PU)</i>	<i>Punugu (PN)</i>	<i>Kurava (KU)</i>	
<i>n</i> <i>D20S481</i>	13	0.055	0.179	0.110	0.074	0.039	0.114	0.111	
	14	0.000	0.012	0.012	0.000	0.000	0.000	0.000	
		128	84	82	108	52	44	18	
	9	0.000	0.012	0.000	0.000	0.000	0.000	0.000	
	10	0.000	0.000	0.000	0.000	0.096	0.000	0.000	
	11	0.065	0.058	0.082	0.026	0.000	0.086	0.050	
	12	0.073	0.000	0.092	0.061	0.077	0.069	0.050	
	13	0.029	0.058	0.051	0.026	0.019	0.000	0.000	
	14	0.087	0.221	0.082	0.123	0.019	0.121	0.000	
	15	0.159	0.255	0.092	0.184	0.192	0.155	0.000	
<i>n</i> <i>D20S473</i>	16	0.341	0.279	0.388	0.298	0.327	0.259	0.500	
	17	0.196	0.106	0.174	0.237	0.250	0.241	0.350	
	18	0.044	0.012	0.041	0.044	0.000	0.069	0.050	
	19	0.007	0.000	0.000	0.000	0.019	0.000	0.000	
		138	86	98	114	52	58	20	
	8	0.000	0.000	0.000	0.009	0.019	0.019	0.000	
	9	0.023	0.042	0.000	0.000	0.019	0.000	0.056	
	10	0.182	0.222	0.228	0.218	0.231	0.346	0.167	
	11	0.659	0.597	0.609	0.591	0.635	0.423	0.444	
	12	0.129	0.125	0.152	0.182	0.077	0.212	0.278	
<i>n</i> <i>D20S604</i>	13	0.008	0.014	0.011	0.000	0.019	0.000	0.056	
		132	72	92	110	52	52	18	
	8	0.000	0.012	0.000	0.000	0.000	0.000	0.000	
	10	0.000	0.000	0.000	0.000	0.000	0.035	0.000	
	11	0.029	0.012	0.021	0.000	0.000	0.000	0.000	
	12	0.036	0.071	0.052	0.009	0.019	0.000	0.050	
	13	0.051	0.036	0.063	0.009	0.019	0.035	0.000	
	14	0.319	0.393	0.417	0.500	0.173	0.207	0.350	
	15	0.348	0.274	0.292	0.304	0.558	0.552	0.500	
	16	0.167	0.179	0.115	0.161	0.212	0.173	0.100	
<i>n</i> <i>D6S1006</i>	17	0.015	0.024	0.042	0.018	0.019	0.000	0.000	
	18	0.036	0.000	0.000	0.000	0.000	0.000	0.000	
		138	84	96	112	52	58	20	
	13	0.157	0.256	0.265	0.186	0.154	0.050	0.200	
	14	0.179	0.128	0.102	0.186	0.058	0.100	0.250	
	15	0.664	0.616	0.633	0.627	0.789	0.850	0.550	
	<i>n</i> <i>D5S1453</i>		134	86	98	118	52	60	20
		4	0.269	0.134	0.229	0.114	0.212	0.333	0.250
		6	0.000	0.012	0.000	0.035	0.000	0.000	0.100
		7	0.022	0.085	0.042	0.105	0.019	0.033	0.100
8		0.485	0.573	0.448	0.588	0.577	0.500	0.200	
9		0.067	0.037	0.052	0.053	0.096	0.083	0.150	
10		0.105	0.110	0.125	0.061	0.077	0.017	0.100	
11		0.045	0.024	0.104	0.026	0.000	0.033	0.100	
12		0.008	0.024	0.000	0.018	0.019	0.000	0.000	
<i>n</i>		134	82	96	114	52	60	20	

**Table 1.** Continued

<i>Locus</i>	<i>Allele</i>	<i>Doddi</i> ( <i>DO</i> )	<i>Erra</i> ( <i>ER</i> )	<i>Karnam</i> ( <i>KA</i> )	<i>Pokanati</i> ( <i>PK</i> )	<i>Puja</i> ( <i>PU</i> )	<i>Punugu</i> ( <i>PN</i> )	<i>Kurava</i> ( <i>KU</i> )
<i>D21S1446</i>	205	0.023	0.000	0.000	0.018	0.000	0.071	0.000
	209	0.469	0.598	0.438	0.455	0.269	0.411	0.333
	213	0.069	0.098	0.188	0.173	0.096	0.214	0.167
	215	0.008	0.000	0.000	0.000	0.000	0.000	0.000
	217	0.023	0.037	0.010	0.000	0.058	0.000	0.056
	219	0.062	0.049	0.031	0.036	0.115	0.071	0.000
	221	0.062	0.000	0.042	0.027	0.019	0.018	0.000
	223	0.239	0.207	0.229	0.246	0.385	0.179	0.444
	225	0.008	0.000	0.000	0.000	0.000	0.000	0.000
	227	0.039	0.012	0.063	0.036	0.058	0.036	0.000
	229	0.000	0.000	0.000	0.009	0.000	0.000	0.000
<i>n</i>		130	82	96	110	52	56	18
<i>D21S1435</i>	7	0.000	0.000	0.047	0.000	0.000	0.000	0.000
	8	0.009	0.000	0.000	0.000	0.000	0.000	0.000
	9	0.110	0.138	0.081	0.157	0.096	0.129	0.000
	10	0.339	0.338	0.279	0.259	0.365	0.307	0.556
	11	0.263	0.238	0.326	0.222	0.308	0.210	0.389
	12	0.246	0.188	0.233	0.287	0.058	0.307	0.056
	13	0.034	0.063	0.023	0.074	0.173	0.032	0.000
	14	0.000	0.025	0.012	0.000	0.000	0.000	0.000
	15	0.000	0.013	0.000	0.000	0.000	0.016	0.000
<i>n</i>		118	80	86	108	52	62	18

cient ice and transported every second day to the Human Genetics Laboratory at the Medical College of Madras University where buffy coats were extracted. DNA was extracted from the buffy coat spots on filter paper using the protocol *Chelex 100* developed by the US Federal Bureau of Investigation.

**DNA Analysis.** The 13 STRs were analyzed in five standard multiplex PCR reactions: (1) *CSF1R*, *TH01*, and *PLA2A*; (2) *F13A1*, *CYP19*, and *LPL*; (3) *D21S1446* and *D21S1435*; (4) *D20S481*, *D20S473*, and *D20S604*; and (5) *D5S1453* and *D6S1006*. Twenty-five to fifty nanograms of DNA were amplified in a total volume of 25  $\mu$ L reaction mixture containing standard PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mM KCl; 1.5 mM MgCl<sub>2</sub>), 200  $\mu$ M each dNPT, 1 unit of *Taq* DNA polymerase, and 1  $\mu$ M of each primer. The amplified products were separated on 6% denaturing polyacrylamide gels. Following electrophoresis, the gels were stained using a silver staining kit (Promega). Known size standards were run in four lanes in each gel for determination of allele sizes. The repeat sizes were determined by sequencing a subset of alleles (*CSF1R*, *TH01*, *PLA2A*, *F13A1*, *CYP19*, *LPL*, *D20S473*) or inferring from GenBank sequences (*D13S124*, *D21S1435*, *D20S481*, *D20S604*, *D5S1453*, and *D6S1006*). Alleles at these loci have been designated by the size (in base pairs) of their PCR product and their corresponding repeat numbers. Fragment sizes at the *D21S1446* (a tetranu-



**Figure 1.** Map showing the location of the study area in Andhra Pradesh and the 30 villages from which the blood samples were collected. The village numbers and the abbreviated population names are given in the map. The explanations of the population abbreviations can be found in Table 1.

cleotide repeat) locus do not always correspond to a 2-bp and 4-bp increment, respectively. Therefore, alleles at this locus could not be assigned repeat numbers. Whereas *PLA2A1*, *D20S473*, *D5S1453*, and *D6S1006* are trinucleotide repeats, the remaining nine loci are tetranucleotide repeats.

**Statistical Analysis.** Allele frequencies were computed using the gene-counting method (Li 1976). Pairwise genetic distances were computed using the modified Cavalli-Sforza distance ( $D_A$ ) of Nei et al. (1983). Although this genetic distance measure is not linear with evolutionary time, it is observed to be most efficient for obtaining correct phylogenetic trees under various evolutionary conditions, and also is least affected by small sample size (Takezaki and Nei 1996). The distances were also obtained using the stepwise-weighted genetic distance measure ( $D_{SW}$ ) of Shriver et al. (1995) developed exclusively to suit microsatellite loci. The  $G_{ST}$  values and heterozygosities were computed and a neighbor-joining (NJ) tree (Saitou and Nei 1987) and an unweighted-pair-group-with-arithmetic-mean (UPGMA) tree were drawn using the program supplied by Dr. Takezaki of the National Institute of Genetics, Mishima, Japan. Since we are

dealing with subgroups of a single caste with a relatively recent history of separation, the allele size variance was not considered in the estimation of  $G_{ST}$ , because the error of estimating such a value would be very large (Takezaki and Nei 1996).

The regression model of Harpending and Ward (1982) was applied to the STR data to study the possible effects of stochastic processes and admixture on these populations. According to this model, the average heterozygosity of the  $i^{\text{th}}$  population ( $H_i$ ) should be equal to the overall mean heterozygosity of the entire population (in this case, of all the Golla populations)  $H_r$ , 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, this linear relationship no longer holds. Isolated groups will be less heterozygous than the linear prediction and hence lie below the expected regression line, whereas populations receiving more gene flow from outside will be more heterozygous and therefore lie above the regression line.

## Results

**Allelic Distributions.** Allele frequency distributions at the 13 STR loci in the seven Golla subpopulations are presented in Table 1. As mentioned in the Methods section, we have designated the alleles in terms of their exact repeat number. Although a detailed, comprehensive, locus-by-locus presentation of allele frequencies is not the objective, some salient features are noteworthy. Of the 13 loci, *CSF1R*, *PLA2A*, *CYP19*, *LPL*, and *D20S473* present a generally uniform pattern of allelic distributions and could be categorized as least variable. As expected, the  $G_{ST}$  values are the lowest at these five loci, with a range of 1.8% to 2.5%. *F13A1*, *D20S481*, *D6S1006*, *D5S1453*, *D21S1446*, and *D21S1435* show a moderate level of interpopulation variation. In general, although the seven Golla subpopulations share the same common alleles, and also in most cases the most predominant allele is the same, there are wider differences in allele frequencies at these loci. The  $G_{ST}$  values range between 3.1% and 3.8%. *TH01* and *D20S604* are the most variable among the studied loci and significant interpopulation differences are observed. Notably, the  $G_{ST}$ s are the highest at these two loci, 4.9% and 4.1%, respectively. In spite of such local differences, however, the Gollas share the same common alleles with other global populations (Deka et al. 1999).

Locus and average population heterozygosities computed based on allele frequencies are presented in Table 2, along with the  $G_{ST}$  values that indicate the degree of differentiation among the subpopulations. The average heterozygosities that reflect within-population heterogeneity are very similar among the subpopulations and vary between 0.64 and 0.70. Among the loci, *D20S481* shows the highest levels of heterozygosity in the populations (0.653–0.834) and *D6S1006* the lowest (0.27–0.626). The coefficient of gene differentiation is moderate ( $G_{ST} = 0.031 \pm 0.0025$ ) at 3.1%. However, as stated earlier, there is considerable het-

**Table 2.** Locus and Average Heterozygosities and the  $G_{ST}$  Values

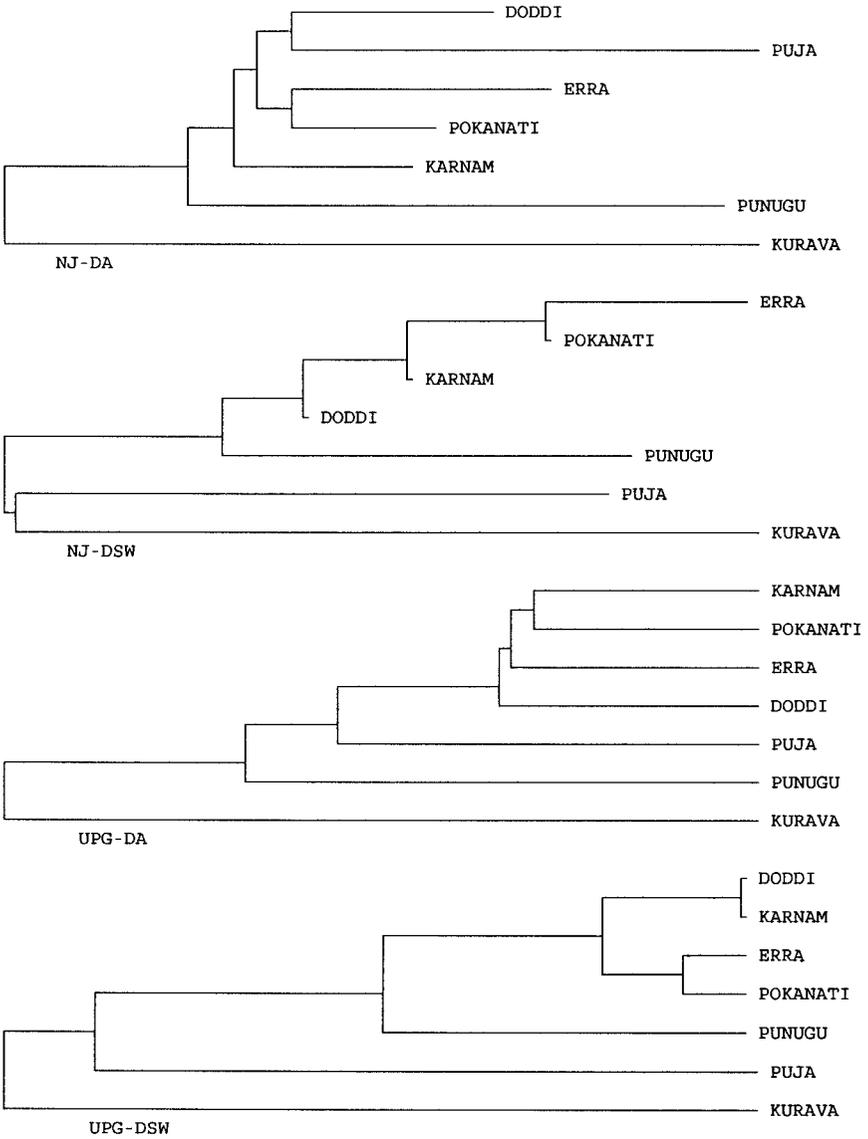
Locus	Doddi	Erra	Karnam	Pokanati	Puja	Punugu	Kurava	$G_{ST}$
CSF1R	0.701	0.722	0.704	0.736	0.628	0.683	0.680	0.024
THO1	0.702	0.757	0.791	0.774	0.671	0.753	0.689	0.049
PLA2A	0.703	0.737	0.727	0.668	0.652	0.564	0.627	0.020
F13A1	0.654	0.690	0.708	0.735	0.667	0.681	0.732	0.031
CYP19	0.630	0.610	0.596	0.605	0.600	0.658	0.575	0.024
LPL	0.702	0.743	0.721	0.685	0.692	0.703	0.732	0.018
D20S481	0.806	0.799	0.793	0.806	0.793	0.834	0.653	0.031
D20S473	0.519	0.585	0.560	0.575	0.548	0.669	0.732	0.025
D20S604	0.749	0.740	0.727	0.637	0.625	0.632	0.647	0.041
D6S1006	0.506	0.545	0.524	0.542	0.358	0.269	0.626	0.036
D5S1453	0.680	0.639	0.723	0.627	0.618	0.640	0.879	0.038
D21S1446	0.713	0.594	0.722	0.706	0.765	0.755	0.699	0.033
D21S1435	0.749	0.781	0.761	0.778	0.744	0.762	0.569	0.035
Average	0.678	0.688	0.697	0.683	0.643	0.662	0.680	0.031

Var. ( $G_{ST}$ ) = 0.000006; S.E. ( $G_{ST}$ ) = 0.002477

erogeneity in the degree of differentiation at different loci. While it is about 5% in the case of *THO1*, it is low at 1.8% in the case of *LPL*.

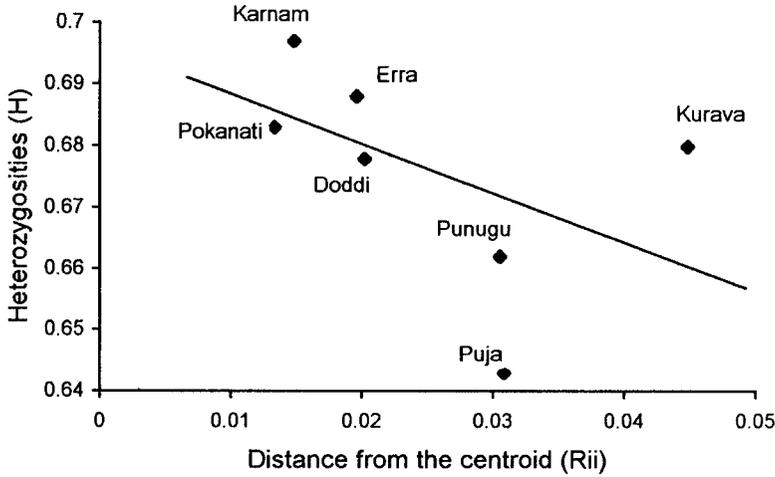
The NJ and UPGMA trees constructed using  $D_A$  and  $D_{SW}$  distances are given in Figure 2. Although there are certain minor differences in the clustering pattern based on  $D_A$  and  $D_{SW}$ , the correspondence between the two matrices is very high as revealed by the Mantel test ( $r = 0.814$ ,  $p = 0.0024$ ). However, all four tree diagrams consistently suggest much earlier separation, hence distinctness, of the Kurava, whose position among the Gollas is disputed, from the rest of Golla subpopulations. Branch length of the Kurava is much larger when compared to that of the other Golla subpopulations. The other feature that emerges is the fact that there is a major split of Punugu and/or Puja from the remaining four Golla subgroups among whom the subclustering was primarily on the basis of geographic proximity. While the two geographically contiguous Erra and Pokanati form a subcluster, the Doddi and Karnam, distributed mostly in the western parts of the district, form another subcluster. However, the Mantel test does not suggest any significant overall correspondence between the average geographic distance matrix and the genetic distance matrices,  $D_A$  ( $r = -0.224$ ,  $p = 0.76$ ) and  $D_{SW}$  ( $r = -0.065$ ,  $p = 0.56$ ), of the Golla populations. Given that the Golla populations live within a small geographical area, such a systematic pattern of association with the genetic distances cannot probably be realized.

The Harpending and Ward (1982) method of regression of mean-per-locus heterozygosities against the distance from the centroid of the allelic distribution provides insight into the relationship between systematic (primarily gene flow) and nonsystematic (genetic drift) pressures acting on subdivided populations. The regression plot of mean heterozygosity versus the  $r_{ii}$  of the seven Golla popula-



**Figure 2.** Neighbor-joining and UPGMA trees based on the  $D_A$  and  $D_{SW}$  distances among the seven Golla subgroups.

tions is presented in Figure 3, which indicates that Puja has the lowest heterozygosity and a relatively high  $r_{ii}$  value, suggesting that this population differentiated as a result of stochastic processes. While Kurava exhibits high distance from the centroid, it also has a relatively high heterozygosity level, suggesting that this population experienced both admixture and the effect of stochastic processes.



**Figure 3.** The plot of mean heterozygosity versus the distance from the gene frequency centroid of the Golla populations and the theoretical regression line.

Karnam and Erra are both above the regression line, and most likely the allelic frequencies were sculpted by gene flow.

## Discussion

India offers immense scope to study the processes of subdivision and their microevolutionary significance. However, this traditional structure of Indian populations is currently facing threat of disintegration as a result of urbanization, increasing communication, and the growing acceptance of marriages between subcastes. Therefore, there is an urgent need to understand this traditional population structure and study its role in shaping human genome diversity in India. The rapid divergence that the hypervariable DNA loci, particularly the STRs, are supposed to undergo brightens this possibility. Towards meeting this goal the present study makes a modest beginning. We consider this study unique in the sense that these local breeding populations represent the most fundamental units of microevolution, and at this level of hierarchy no such studies have yet been attempted using molecular genetic markers.

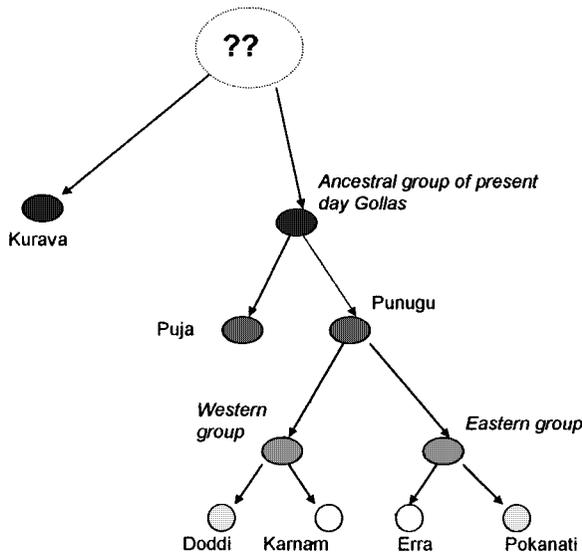
Not only are these populations strictly defined endogamous units, but they are also likely to have originated from a common ancestral source in the recent past, hence their evolutionary history is short. Yet, we have observed considerable variation in the allele frequency of the populations at different loci, and the  $G_{ST}$  values suggest relatively high discrimination (3.1%) among the subgroups, supporting a history of subcaste endogamy. In fact, the average  $G_{ST}$  observed for the Indian populations (both castes and tribes) for traditional markers is only about

1.5%. Based on 6 of the 13 STR loci included in the present study, Deka et al. (1995) observed much lower levels of genetic differentiation among the populations of major races and/ or continents, although a much higher degree of differentiation among American Indian populations and Pacific Islanders. They attributed this to smaller population sizes and greater isolation of the Amerindians. Relatively speaking, our populations represent much lower levels of population hierarchy and a short history of separation and isolation. Our findings, therefore, can be considered as consistent in the above context. The within-group heterogeneity as reflected by average heterozygosities is relatively small in our populations ( $< 0.70$ ) and comparable to those among the Amerindians and Pacific Islanders, in reciprocal relationship with the degree of between-group differentiation as suggested by the  $G_{ST}$ s. Mukherjee et al. (1999) recently reported genetic diversity based on four STR loci among the eight Indian populations, ethnically and geographically quite diverse and with a much higher level of substructure compared to the Gollas. However, only 2 of the 4 loci they included were common among the 13 loci we studied. Yet the average  $G_{ST}$  value is found to be only about 1% higher than that observed for Gollas at the lowest level of caste substructure (4.1% against 3.1%). While this average pattern is truly reflected at one of the two common loci (*TH01*, 6% against 4.9%), a much higher level of differentiation is evident at the *CFR1R* locus in the case of the former (6.2% against 2.4% among the Gollas).

On the other hand, that the position of the Kurava population among the Gollas is disputed, despite their occupational and cultural similarity, is vindicated in the largest genetic distances observed between the Kurava and the other Golla subpopulations. It appears from the clustering pattern and branch lengths emerging from both the NJ and UPGMA trees that the Kuravas and a major ancestral population of the present-day Golla subpopulations might have diverged relatively early so as to overwhelm the sociocultural identity between them. Hence, the genetic pattern based on the 13 STR loci may even be said to suggest a separate origin of the Kurava caste, distinct from the ancestral population of the present-day Gollas.

The seminomadic Golla populations have traditionally subsisted on large herds of sheep and goat, moving seasonally in search of better grazing lands. Such traditional populations are usually guided in such sociocultural behaviors as marriage alliances by the customary laws of the community. Increase in size of such traditional populations often led to dispersal to a wider geographical territory, threatening group solidarity and cohesion as communication became more difficult. Coupled with a violation of certain traditional norms, such as those concerning marriage, by certain members of the community, this dispersal might have led to further subdivisions within the population. Judging from the NJ and UPGMA trees constructed on the basis of pairwise genetic distances, the most likely scenario would have been the one depicted in the schematic diagram (Figure 4). It suggests that there had been a major split among Gollas with Punugu and/ or Puja on the one hand and the remaining groups on the other. One of these

splinter groups, Puja/Punugu, with population expansion and increasing pressure on the grazing fields for their livestock, might have moved to different directions looking for new and vacant habitats (which they tend to do even today, but only seasonally), settling ultimately in some of those places. At least two such founding populations, eastern and western, might have resulted in the process. The Erra and Pokanati, distributed in the eastern parts of the district, and the Doddi and Karnam, distributed in the western parts, might have been the offshoots of those two regional groups, involving certain processes of fission. For example, most elders of this community think that the Erra Gollas are the descendents of unions involving a Golla woman and a Brahmin man. To substantiate this, they say that traditionally Erra Gollas wore sacred thread and ate only vegetarian food as Brahmins do. It was not uncommon that people involved in such unions were ostracized by their respective communities, forcing them to form a separate endogamous group as was documented in certain other seminomadic groups of western India (Malhotra 1978b; 1979). Similarly, while the Doddi Gollas are supposed to be closely associated with their flock of sheep, even today the Karnams are traditionally given the added responsibility of guarding government and temple treasuries. In addition, certain families within the Karnam caste have been traditionally assigned the job of village *municip* (revenue collector of the village), which was a hereditary job. The person performing this job carried the title “Reddy,” which was otherwise associated with members of another agricultural caste in the state. All those Karnam family members currently carry this title and apparently restrict marriages among such families. Such occupational divergence might have



**Figure 4.** Schematic diagram depicting the probable process of the formation of subgroups among the Gollas.

led to endogamous subdivision within the western group, as was noted earlier in other occupational groups, for example, the marine fishermen of the Karnataka coast (Reddy and Gadgil 1999). It is therefore plausible to surmise that some of these processes might have been responsible for the formation of endogamous regional subunits, despite their contiguous and/or overlapping distributions, as is the case with Erra and Pokanati, and Doddi and Karnam, or even the Puja and Punugu, albeit at different times. Assuming that the genetic distances obtained are linear with the time scale, branch lengths observed in NJ and UPGMA trees suggest a relatively much later differentiation from a common stock of Erra and Pokanati, and Doddi and Karnam when compared to the other three groups—the Puja, Punugu, and Kurava. Further insights into their population structure should await analysis of remaining Golla DNA samples for mitochondrial DNA and Y-chromosome microsatellite markers. For example, analyzing those markers Bamshad et al. (1998) concluded that the genetic stratification of Telugu castes was based on the social mobility of females, not males. It is very well known that even within a caste cluster or within subcastes of India a clearly stratified hierarchy exists. Nevertheless, at this stage it may be said that the insights offered by the results of the application of the Harpending and Ward (1982) regression model conform to the conjectures made above. For example, the alien Kurava on the one hand, and the Erra and Karnam populations with some indications of external gene flow on the other, present themselves above the theoretical regression line, whereas the remaining four groups lie below the line. Furthermore, the Kurava population, with its greater antiquity and the Puja, a small isolated population, lie above and below the line, respectively, as marked outliers, suggesting larger effect of gene flow and genetic drift, respectively.

To conclude, we may say that the 13 STR loci of the populations studied provide useful insights into understanding the population structure and genetic microdifferentiation among substructured local populations with a relatively short evolutionary history. Although a larger number of loci may provide greater confidence in drawing the inferences on the short-term evolution of such populations, there is no doubt that this class of hypervariable loci offers new hope of understanding microvariations at the local level more accurately than have traditional genetic and other biological variables.

*Acknowledgments* This work was done during BMRs tenure as a Fulbright visiting scholar at the Department of Anthropology of the University of Kansas, Lawrence, Kansas. He thanks the Fulbright Foundation for the award of Fellowship. He thanks P. Chengal Reddy, Drs. Venkatramana, Nirmalananda Reddy, and many other friends at S.V. University, Tirupati, in addition to the lab technicians and Medical Officers at different field camps, whose cooperation was vital for the successful collection of blood samples. We are also thankful to Professors A. Ramesh and Gopinath at Madras University, who provided lab facilities and expertise for extraction of buffy coats from the blood samples. Last, but not the least, we are grateful to the subjects of our study and the Golla leaders for

their cooperation during the fieldwork. We thank Joe McComb for technical help in the extraction of DNA, Dario Demarchi for help in *R*-matrix analysis of the STR data, and my student Banrida Theresa Langstieh for help in graphics. The comments of two anonymous reviewers on the initial version of this manuscript were helpful in improved presentation of the results. We thank them.

*Received 17 December 1999; revision received 12 September 2000.*

## Literature Cited

- Bamshad, M.J., W.S. Watkins, M.E. Dixon et al. 1998. Female gene flow stratifies Hindu castes. *Nature* 395:651–652.
- Basu, A. 1969. The Pahira: A population genetical study. *Am. J. Phys. Anthropol.* 31:399–416.
- Deka, R., M.D. Shriver, L.M. Yu et al. 1995. Intra and Inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* 16:1659–1664.
- Deka, R., M.D. Shriver, L.M. Yu et al. 1999. Genetic variation at twenty-three microsatellite loci in sixteen world-wide populations. *J. Genet.* 78:99–121.
- Eller, E. 1999. Population substructure and isolation by distance in three continental regions. *Am. J. Phys. Anthropol.* 108:147–159.
- Harpending, H.C., and R. Ward. 1982. Chemical systematics and human evolution. In *Biochemical Aspects of Evolutionary Biology*, M. Nitecki, ed. Chicago, IL: University of Chicago Press, 213–256.
- Jorde, L.B., A.R. Rogers, M. Bamshad, et al. 1997. Microsatellite diversity and demographic history of modern humans. *Proc. Natl. Acad. Sci., USA.* 94:3100–3103.
- Karve, I., and K.C. Malhotra. 1968. Biological comparison of eight endogamous groups of the same rank. *Current Anthropol.* 9:109–124.
- Li, C.C. 1976. *First Course in Population Genetics*. Pacific Grove, CA: Boxwood.
- Malhotra, K.C. 1978a. Microevolutionary dynamics among the Govdas of Goa. In *Evolutionary Models and Studies in Human Diversity*, R.J. Meier et al., eds. The Hague, Netherlands: Mouton, 315–323.
- Malhotra, K.C. 1978b. Founder effect, gene drift and natural selection among four nomadic mendelian isolates. In *Evolutionary Models and Studies in Human Diversity*, R.J. Meier et al., eds. The Hague, Netherlands: Mouton, 279–314.
- Malhotra, K.C. 1979. Excommunication as a process leading to the formation of new groups. *The Eastern Anthropologist* 32:49–53.
- Malhotra, K.C., R. Chakraborty, and A. Chakraborty. 1978. Gene differentiation among the Dhangar caste-cluster of Maharashtra, India. *Human Heredity* 28:26–36.
- Mukherjee, N., P.P. Majumder, B. Roy et al. 1999. Variation at 4 short tandem repeat loci in eight population groups of India. *Hum. Biol.* 71:439–446.
- Murty, J.S., B. Muralidhar, J.D. Goud et al. 1993. Hierarchical gene diversity and genetic structure of tribal population of Andhra Pradesh, India. *Am. J. Phys. Anthropol.* 90:169–183.
- Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19:153–170.
- Nei, M., and N. Takezaki. 1996. The root of phylogenetic tree in human populations. *Mol. Biol. Evol.* 13:170–177.
- Ramana, G.V., J.M. Naidu, and J.S. Murty. 1996. Genetic structure of the Manne Dora, a tribal population of Andhra Pradesh. *Hum. Biol.* 68:63–74.

- Reddy, B.M. 1982. *Population biology of the marine fishermen of Puri, Orissa, India*. Unpublished Ph.D. dissertation, University of Calcutta, India.
- Reddy, B.M., V.P. Chopra, and D.P. Mukherjee. 1987. Biological affinities between the migrant groups of fishermen of Puri coast, India. *Am. J. Phys. Anthropol.* 74:407–416.
- Reddy, B.M., V.P. Chopra, A. Rodewaldt et al. 1989. Genetic differentiation among some endogamous groups of fishermen of the eastern coast, India. *Ann. Hum. Biol.* 16: 321–333.
- Reddy, B.M., V.P. Chopra, A. Rodewaldt et al. 1995. Genetic affinities between migrant and parental populations of fishermen on the east coast of India. *Am. J. Hum. Biol.* 7:51–63.
- Reddy, B.M., and M. Gadgil. 1999. Dynamics of endogamy and resource utilization patterns among the marine fishermen castes of the Karnataka coast, India. (Manuscript.)
- Saitou, N., and M. Nei. 1987. The neighbor joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.
- 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.
- Shriver, M.D., L. Jin, R.E. Ferrell et al. 1997. Microsatellite data support an early population expansion in Africa. *Genome Research* 7:586–591.
- Sirajuddin, S.M., R. Duggirala, and M.H. Crawford. 1994. Population structure of the Chenchu and other south Indian tribal groups: Relationship between genetic, anthropometric, dermatoglyphic, geographic and linguistic distances. *Hum. Biol.* 66:865–884.
- Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.
- Thurston, E. 1909. *Castes and Tribes of Southern India*. Vol. 2. C-J. Madras: Government Press, 284–295.
- Weber, J.L., and C. Wong. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 8:1123–1128.