Genetic Diversity Within a Caste Population of India as Measured by Y-Chromosome Haplogroups and Haplotypes: Subcastes of the Golla of Andhra Pradesh

R. John Mitchell,[1]* B. Mohan Reddy,[2] David Campo,[1] Tania Infantino,[1] Maria Kaps,[1] and Michael H. Crawford[3]

[1]*Department of Genetics and Human Variation, School of Molecular Sciences, La Trobe University, Melbourne, Victoria 3086, Australia*
[2]*Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, India*
[3]*Laboratory of Biological Anthropology, Department of Anthropology, University of Kansas, Lawrence, Kansas 66045*

KEY WORDS genetic structure; Y-chromosome; haplotypes; caste system; Golla; India

ABSTRACT The extent of population subdivision based on 15 Y-chromosome polymorphisms was studied in seven subcastes of the Golla (Karnam, Pokanati, Erra, Doddi, Punugu, Puja, and Kurava), who inhabit the Chittoor district of southern Andhra Pradesh, India. These Golla subcastes are traditionally pastoralists, culturally homogeneous and endogamous. DNA samples from 146 Golla males were scored for seven unique event polymorphisms (UEPs) and eight microsatellites, permitting allocation of each into haplogroups and haplotypes, respectively. Genetic diversity (D) was high (range, 0.9048–0.9921), and most of the genetic variance (>91%) was explained by intrapopulation differences. Median-joining network analysis of microsatellite haplotypes demonstrated an absence of any structure according to subcaste affiliation. Superimposition of UEPs on this phylogeny, however, did create some distinct clusters, indicating congruence between haplotype and haplogroup phylogenies. Our results suggest many male ancestors for the Golla as well as for each of the subcastes. Genetic distances among the seven subcastes, based on autosomal markers (short tandem repeats and human leukocyte antigens) as well as those on the chromosome Y, indicate that the Kurava may not be a true subcaste of the Golla. Although this finding is based on a very small Kurava sample, it is in accordance with ethnohistorical accounts related by community elders. The Punugu was the first to hive off the main Golla group, and the most recently separated subcastes (Karnam, Erra, Doddi, and Pokanati) fissioned from the Puja. This phylogeny receives support from the analysis of autosomal microsatellites as well as HLA loci in the same samples. In particular, there is a significant correlation ($r = 0.8569$; $P = 0.0097$) between Y-chromosome- and autosomal STR-based distances. Am J Phys Anthropol 130:385–393, 2006.

The current population of the Indian subcontinent is the product of numerous migrations from many directions and subsequent admixtures since the Paleolithic. Of crucial importance in Indian prehistory is the invasion of Indo-European speakers in the second millennium BC, who subjugated most of the indigenous populations and imposed the hierarchical caste system that still characterises Indian society today. The effects of the caste system (the rigid division of society into endogamous groups) make the Indian population of great interest to anthropologists and human population geneticists, because those belonging to castes are seen as distinct from the tribal people, and each of these castes is endogamous. This hierarchical caste system is further subdivided along geographic, linguistic, and religious differences, and many castes contain within them many castes that are themselves endogamous. It is probable that most subcastes evolved from a common ancestral group through successive acts of fission, but it is not always clear that this is so. These subcastes are usually characterized by a high degree of isolation, small effective population sizes, and relatively high inbreeding levels. Under such conditions, over time these subcastes would be expected to differentiate genetically from each other. However, though still all-pervasive in India, the caste system probably dates from ~6,000–4,000 ybp (Gadgil et al., 1997), which is a relatively short evolutionary time. Some authors argue, however, that the caste system is more ancient and preceded the Indo-Aryan invasion (Gonem, 1996; Arnaiz-Villena et al., 2001). Because of this marked social structure imposed by subcastes, it is of interest to investigate the extent of genetic differentiation among them, as it may shed light on the key elements of such microevolutionary processes.

One caste with a number of known subcastes, the Golla of Andhra Pradesh, were investigated for the level of genetic microdifferentiation among them (Reddy et al., 2001a,b; Crawford et al., 2001), but the results of the two studies differed greatly, although both reported considerable diversity among the Golla. The relationships re-

vealed after scoring 13 autosomal microsatellites in seven subcastes fitted reasonably well with the ethnohistory of the Golla in Andhra Pradesh as related by the current "elders" to the researchers (Reddy et al., 2001a,b). There are no official or written records describing the history of these groups, however, or even agreement about the number of subcastes within the Golla (Reddy et al., 2001a,b). Results from scoring human leukocyte antigens (HLA) class I and II loci in six of these subcastes (Crawford et al., 2001) generated a different set of relationships from that revealed by these noncoding markers. Both studies employed autosomal loci that are subject to meiotic recombination, and therefore in many ways are not ideal markers for investigating the history of human groups subject to fission. In addition, there is the possible role of selection altering allele frequencies; this is especially so with respect to the HLA system. Loci on the nonrecombining portion of the Y-chromosome (NRPY) present no such problem of loss of mutations through recombination, and they record the history of only the paternal lineages of these subcastes. The NRPY contains both biallelic and unique event polymorphisms (UEPs) as well as microsatellites, characteristics that make it ideal for examining structure in human populations over both the long and short evolutionary period.

The aim of this research was to test the hypothesis that a variety of Y-chromosome polymorphisms can more clearly identify the process(es) of subdivision that generated subcastes of the Golla in Andhra Pradesh. To test this hypothesis, we analyzed a number of UEPs and microsatellites on the Y-chromosome in order to construct a phylogeny from which aspects of both the genetic structure and fissioning process(es) of the Golla subcastes could be inferred. The level of gene diversity in the Golla male gene pool and the extent to which Y haplotypes are shared among subcastes was assessed. We also determined the correspondence between Y-chromosome diversity and that of autosomal markers reported in previous studies.

SUBJECTS AND METHODS

Populations

The Golla are a Telugu-speaking pastoral caste with an estimated population of more than 4,700,000 located in the state of Andhra Pradesh, southeast India (Thurston, 1975; Srinivasulu, 2002). Telugu is a Dravidian language (Ruhlen, 1991), and this means that the Golla can be regarded an ancient Indian population (i.e., Dravidians). People of the Golla tradition are found in other parts of India but are known by different names and most probably have different origins. The Golla of Andhra Pradesh primarily herd sheep and cattle and subsist by selling milk, but some Golla communities own land and practice agriculture. The exact number of subcastes is unknown, but seven subcastes that inhabit the culturally homogeneous Chittoor district could be identified (Doddi, Puja, Erra, Pokanati, Karnam, Punugu, and Kurava) during sample collection. The caste's own oral traditions suggest that these (or at least 6 of the 7) evolved from a common parental stock that has been subject to processes of fission (Reddy et al., 2001a,b). Each subcaste is distinguished by a unique variant of their traditional occupation, as well as by variation in marriage ceremonies and dress patterns. Apart from Thurston (1975) and the data collected in the fieldwork that led to the collection of the samples in the present study (Reddy et al., 2001a,b), there is very little literature available on the ethnography of the Golla.

The Golla subcastes practice consanguineous marriage and village endogamy. Reddy et al. (2001a,b) estimated inbreeding coefficients (based on the consanguinity of the parents of the study subjects as $F = 0.0272$ for all Golla, but values ranged from 0.0175 among the Karnam to 0.0354 in the Pokanati subcaste). These F-values are of a similar magnitude to values reported for other populations of coastal and rural Andhra Pradesh (Dronamraju, 1964; Reid, 1973).

Samples

DNA was extracted from 5- or 10-ml intravenous blood samples collected from 146 unrelated males who gave their informed consent. These subjects belonged to seven subcastes of the Golla caste resident in 30 villages of the Chittoor district in Southern Andhra Pradesh (Fig. 1). The sample distribution was: Doddi ($n = 24$), Erra ($n = 23$), Karnam ($n = 22$), Pokanati ($n = 29$), Puja ($n = 20$), Punugu ($n = 21$) and Kurava ($n = 7$).

DNA polymorphisms

DNA samples were extracted from blood using a standard method (Sambrook et al., 1989), and were typed for seven previously described single UEPs: YAP (Hammer, 1994), SRY10831a, SRY10831b (Whitfield et al., 1995), 92R7 (Mathias et al., 1994), 12f2 (Casanova et al., 1985), M9 (Underhill et al., 1997), and M122 (Su et al., 1999). We followed the terminology recommended by the Y Chromosome Consortium (2002) for naming non-recombining portion of the Y chromosome (NRY) lineages determined after scoring these UEP markers. In the present study, eight microsatellites were genotyped: DYS19, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439 (Kayser et al., 1997; Ayub et al., 2000). Scoring of these microsatellites generated eight-locus Y haplotypes.

Data analysis

Unbiased estimates of locus and haplotype diversity were calculated according to Nei (1987), using the program ARLEQUIN (Schneider et al., 2000). Genealogical relationships among Y haplotypes were reconstructed by means of a median-joining (M-J) network analysis using NETWORK 4.0 (Rohl, 2001). The estimated age of the most recent common ancestor (TMRCA) in the phylogeny was measured as age in mutations (rho statistic). The mutational age was then converted into years by multiplication with the mutation rate (Saillard et al., 2000). We used the effective mutation rate recently determined by Zhivotovsky et al. (2004) of 6.9×10^{-4} per 25 years, which was based on analysis of 10 Y-short tandem repeats (STRs) (including six used in this study).

Genetic relationships among the seven populations, based on the eight-loci Y-microsatellite haplotypes, were further explored by analysis of molecular variance (AMOVA), as implemented in ARLEQUIN (Schneider et al., 2000). The genetic structure was analyzed with consideration for the molecular differences between individual haplotypes, in addition to differences in haplotype frequencies, resulting in estimates of Φ_{ST} (an F_{ST} analogue). Significance levels of genetic variance components were estimated by the use of 10,000 permutations.

POPULATIONS 1.2.28 (http://www.pge.cnrs-gif.fr/bioinfo/populations/index.php) was applied to compute

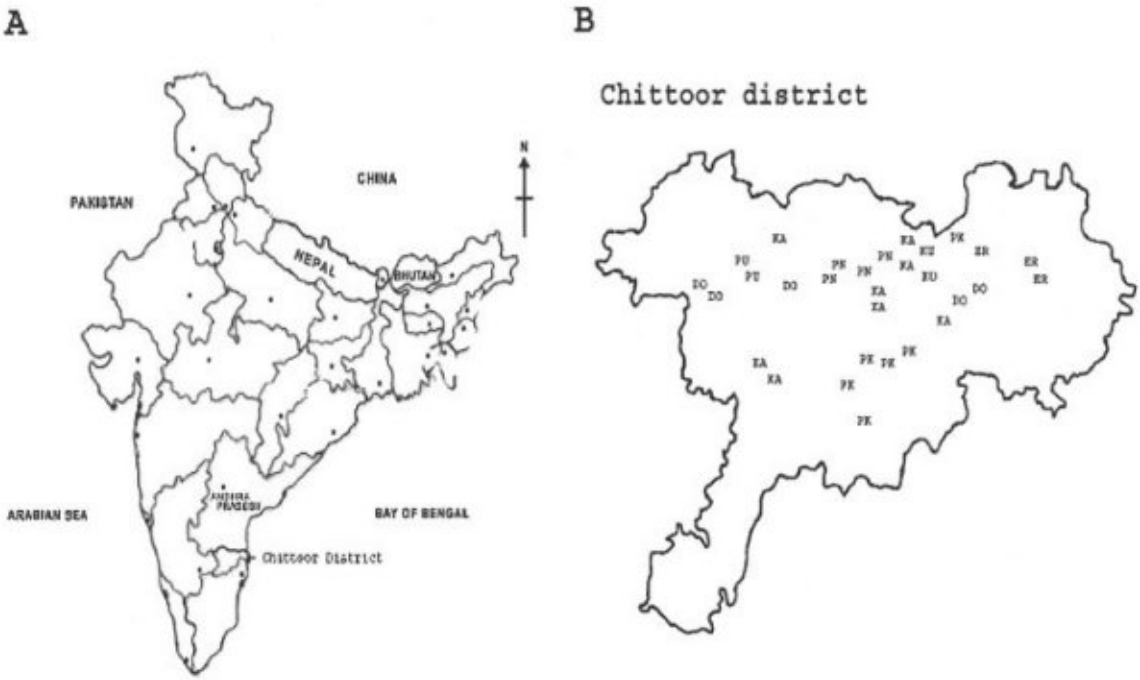

Fig. 1. **A:** Map showing location of study area in Andhra Pradesh, India. **B:** Chittoor district map shows distribution of 30 villages where Golla samples were collected: Doddi (DO), Erra (ER), Karnam (KA), Pokanati (PK), Puja (PU), Punugu (PN), and Kurava (KU).

TABLE 1. Y-chromosome haplogroup frequencies in seven Golla subcastes

Group	n	P-92R7	J-12f2	K-M9	Y (xD, E, J, K)	R1A-SRY10831a
Doddi	24	16.67	8.33	45.83	29.17	0.0
Erra	23	21.74	4.35	21.74	47.83	4.35
Karnam	22	9.09	18.18	31.82	36.36	4.55
Pokanati	29	17.24	10.34	17.24	51.72	3.45
Puja	20	20.0	15.0	15.0	40.0	10.0
Punugu	21	4.76	14.29	23.81	19.05	38.1
Kurava	7	0.0	14.29	28.57	57.14	0.0
Total	146	21.0	17.0	38.0	57.0	13.0

TABLE 2. Y-chromosome microsatellite haplotype diversity in seven Golla subcastes

Population	n[1]	Hap[2]	Shared[3]	D[4]	M[5]	Het[6]
Doddi	24	17	5 (29.42%)	0.9638	0.5883	0.678
Erra	23	21	6 (31.72%)	0.9921	0.5998	0.688
Karnam	22	16	5 (31.25%)	0.9567	0.619	0.697
Pokanati	29	25	5 (20%)	0.9901	0.5785	0.683
Puja	20	14	2 (15%)	0.9579	0.5572	0.643
Punugu	21	12	5 (41.67%)	0.9381	0.5761	0.662
Kurava	7	5	1 (20%)	0.9048	0.5535	0.68
Total	146	93		0.9878	0.6117	0.6759

[1] Number of individuals.
[2] Number of different haplotypes.
[3] Number of haplotypes shared with any other population.
[4] Haplotype diversity.
[5] Average diversity of eight Y-microsatellites.
[6] Average heterozygosity of 13 autosomal microsatellites (Reddy et al., 2001a).

the genetic distance D_A (Nei et al., 1983), using the haplogroup and microsatellite haplotype frequencies in each subcaste, respectively. Takezaki and Nei (1996) calculated the probability of obtaining the correct tree topology for many different distances, and found that in both the infinite-allele model (IAM) and the stepwise mutation model (SMM), the D_A measure generally was more appropriate than other distance measures for clarifying the phylogeny of closely related populations. A Mantel test for calculating the correlation between matrix distances was performed, using the program ZT 1.0 (http://www.psb.ugent.be/~erbon/mantel/) with 100,000 permutations. Nonmetric multidimensional scaling (NM-MDS) analysis (Kruskal, 1964) was performed using STATISTICA (Statsoft) in order to visualize genetic distances among subcastes. To evaluate how closely a particular configuration in the NM-MDS plot reproduces the observed distance matrix, the "stress" value is given. The smaller the stress value, the better the fit.

RESULTS

The seven UEPs scored in the Golla potentially define seven NRY lineages or haplogroups, but only five were found (Table 1). The most common haplogroup in the total Golla sample was Y(xD, E, J, K) (39%), having the ancestral state for each of the seven UEPs. Then followed haplogroups K-M9 (26%), P-92R7 (14.4%), J-12f2 (11.6%), and R1a-SRY10831a (8.9%). All five haplogroups were observed in each of the seven subcastes, except for haplogroups P-92R7 and R1a-SRY10831a that were absent in the Kurava, and R1a, also absent in the Doddi. Overall haplogroup diversity was 0.7427, and ranged from 0.6666 in the Kurava to 0.7761 in the Punugu.

Ninety-three different microsatellite haplotypes were observed among 146 individuals (Table 2). Sixty-four haplotypes (accounting for 68.81% of unique haplotypes) were found in only one individual, 20 (21.50%) in two individuals, 3 (3.22%) in three individuals, 4 (4.30%) in four individuals, 1 (1.07%) in five individuals and 1 (haplotype 24, 1.07%) in 12 individuals. The most common

TABLE 3. Y-chromosome haplogroups and haplotypes observed in Golla by subcaste

ht	STR haplotype	Haplogroup	Doddi	Erra	Karnam	Pokanati	Puja	Punugu	Kurava	n
h1	15, 24, 10, 11, 15, 8, 10, 11	Y(xD, E, J, K)	1	0	0	0	0	0	0	1
h2	16, 23, 10, 11, 12, 8, 10, 11	Y(xD, E, J, K)	1	0	0	0	0	0	0	1
h3	15, 22, 10, 11, 12, 8, 6, 11	Y(xD, E, J, K)	1	0	0	0	0	0	0	1
h4	15, 22, 10, 11, 12, 7, 9, 11	Y(xD, E, J, K)	1	1	0	0	0	0	0	2
h5	15, 22, 10, 11, 12, 8, 9, 13	Y(xD, E, J, K)	1	1	0	0	0	0	0	2
h6	16, 22, 10, 11, 13, 9, 10, 13	Y(xD, E, J, K)	2	0	0	0	0	0	0	2
h7	15, 24, 10, 10, 14, 9, 11, 10	P-92R7	1	0	0	0	0	0	0	1
h8	16, 25, 11, 11, 13, 8, 11, 10	P-92R7	1	0	0	0	0	0	0	1
h9	16, 25, 10, 11, 13, 9, 11, 11	P-92R7	2	0	0	0	0	0	0	2
h10	14, 24, 10, 13, 13, 8, 11, 11	J-12f2	1	0	0	0	0	0	0	1
h11	14, 23, 10, 11, 14, 8, 9, 11	J-12f2	1	0	0	0	0	0	0	1
h12	14, 23, 10, 17, 11, 8, 11, 12	K-M9	1	0	0	0	0	0	0	1
h13	14, 23, 10, 14, 11, 8, 11, 12	K-M9	2	0	0	1	0	0	0	3
h14	14, 22, 10, 14, 11, 9, 10, 11	K-M9	4	0	0	0	0	0	0	4
h15	14, 22, 10, 14, 11, 9, 10, 12	K-M9	2	1	0	0	0	2	0	5
h16	14, 23, 10, 16, 11, 8, 11, 12	K-M9	1	0	0	0	0	0	0	1
h17	14, 22, 10, 15, 11, 9, 10, 12	K-M9	1	0	1	0	0	0	0	2
h18	14, 23, 10, 10, 14, 10, 11, 10	R1a-SRY10831a	0	1	1	0	0	0	0	2
h19	15, 23, 10, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	2	0	0	0	0	0	2
h20	17, 24, 10, 11, 12, 8, 10, 13	Y(xD, E, J, K)	0	2	0	0	0	0	0	2
h21	15, 23, 9, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	1	0	0	0	0	0	1
h22	15, 23, 11, 11, 13, 8, 10, 12	Y(xD, E, J, K)	0	1	0	0	0	0	0	1
h23	17, 23, 10, 11, 12, 8, 10, 13	Y(xD, E, J, K)	0	1	0	0	0	0	0	1
h24	15, 22, 10, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	1	3	2	3	1	2	12
h25	15, 21, 10, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	1	0	0	0	0	0	1
h26	15, 26, 10, 12, 14, 8, 11, 10	P-92R7	0	1	0	0	0	0	0	1
h27	15, 24, 10, 11, 13, 8, 10, 11	P-92R7	0	1	0	0	0	0	0	1
h28	15, 25, 10, 12, 13, 8, 11, 10	P-92R7	0	1	0	0	0	0	0	1
h29	15, 24, 10, 13, 13, 8, 11, 10	P-92R7	0	1	0	0	0	0	0	1
h30	17, 24, 10, 11, 13, 8, 11, 11	P-92R7	0	1	0	0	0	0	0	1
h31	16, 21, 10, 11, 12, 8, 9, 11	J-12f2	0	1	0	0	0	0	0	1
h32	14, 23, 10, 13, 11, 9, 10, 12	K-M9	0	1	0	0	0	0	0	1
h33	14, 24, 10, 14, 11, 9, 10, 12	K-M9	0	1	0	0	0	0	0	1
h34	15, 22, 10, 12, 12, 10, 10, 13	K-M9	0	1	0	0	0	0	0	1
h35	14, 22, 10, 14, 11, 9, 11, 11	K-M9	0	1	0	0	0	2	0	3
h36	16, 22, 11, 11, 13, 9, 10, 12	Y(xD, E, J, K)	0	0	1	0	0	0	0	1
h37	15, 22, 10, 11, 12, 8, 9, 12	Y(xD, E, J, K)	0	0	1	0	0	0	0	1
h38	16, 22, 10, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	0	1	0	0	0	0	1
h39	16, 22, 9, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	0	1	0	0	0	0	1
h40	15, 21, 10, 11, 13, 10, 10, 12	Y(xD, E, J, K)	0	0	1	0	0	0	0	1
h41	16, 26, 10, 11, 12, 8, 11, 10	P-92R7	0	0	1	0	0	0	0	1
h42	15, 24, 10, 11, 13, 8, 11, 10	P-92R7	0	0	1	0	1	0	0	2
h43	14, 23, 10, 11, 12, 8, 9, 13	J-12f2	0	0	4	0	0	0	0	4
h44	14, 25, 11, 14, 14, 8, 10, 12	K-M9	0	0	2	0	0	0	0	2
h45	14, 25, 11, 14, 14, 8, 10, 13	K-M9	0	0	1	1	0	0	0	2
h46	15, 24, 11, 14, 14, 8, 10, 13	K-M9	0	0	1	0	0	0	0	1
h47	14, 22, 10, 14, 11, 9, 10, 14	K-M9	0	0	1	0	0	0	0	1
h48	14, 23, 10, 14, 11, 9, 10, 12	K-M9	0	0	1	0	0	0	0	1
h49	14, 25, 10, 14, 13, 10, 11, 11	R1a-SRY10831a	0	0	0	1	0	0	0	1
h50	15, 22, 11, 11, 12, 8, 10, 12	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h51	15, 22, 10, 14, 13, 9, 10, 13	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h52	15, 22, 10, 15, 12, 8, 9, 12	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h53	15, 22, 10, 16, 12, 8, 9, 11	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h54	15, 22, 10, 15, 12, 8, 9, 11	Y(xD, E, J, K)	0	0	0	2	0	0	0	2
h55	15, 22, 10, 15, 12, 9, 10, 12	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h56	15, 22, 10, 14, 12, 8, 9, 11	Y(xD, E, J, K)	0	0	0	2	0	0	0	2
h57	15, 21, 10, 14, 12, 8, 8, 11	Y(xD, E, J, K)	0	0	0	2	0	0	0	2
h58	15, 22, 10, 14, 12, 8, 9, 12	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h59	16, 23, 10, 11, 12, 8, 9, 11	Y(xD, E, J, K)	0	0	0	1	0	0	0	1
h60	15, 24, 10, 11, 13, 8, 10, 10	P-92R7	0	0	0	1	0	0	0	1
h61	16, 25, 11, 11, 13, 8, 11, 11	P-92R7	0	0	0	1	0	0	0	1
h62	15, 25, 10, 15, 13, 8, 11, 10	P-92R7	0	0	0	1	0	0	0	1
h63	15, 23, 10, 13, 13, 8, 11, 10	P-92R7	0	0	0	1	0	0	0	1
h64	16, 25, 10, 14, 13, 8, 11, 10	P-92R7	0	0	0	1	0	0	0	1
h65	17, 24, 10, 11, 12, 9, 9, 13	J-12f2	0	0	0	1	0	0	0	1
h66	14, 23, 10, 11, 13, 8, 9, 11	J-12f2	0	0	0	1	0	3	0	4
h67	14, 24, 10, 15, 12, 8, 9, 13	J-12f2	0	0	0	1	0	0	0	1
h68	14, 22, 10, 15, 11, 9, 10, 13	K-M9	0	0	0	1	0	1	0	2
h69	14, 22, 11, 15, 11, 9, 10, 11	K-M9	0	0	0	1	0	0	0	1
h70	14, 22, 10, 14, 11, 8, 11, 12	K-M9	0	0	0	1	0	0	0	1

(*continued*)

TABLE 3. (Continued)

ht	STR haplotype	Haplogroup	Doddi	Erra	Karnam	Pokanati	Puja	Punugu	Kurava	n
h71	14, 23, 10, 10, 14, 10, 11, 11	R1a-SRY10831a	0	0	0	0	2	0	0	2
h72	15, 22, 10, 11, 13, 9, 10, 12	Y(xD, E, J, K)	0	0	0	0	2	0	0	2
h73	15, 21, 10, 11, 12, 9, 8, 12	Y(xD, E, J, K)	0	0	0	0	1	0	0	1
h74	15, 22, 10, 11, 13, 8, 9, 11	Y(xD, E, J, K)	0	0	0	0	1	0	0	1
h75	15, 21, 10, 11, 12, 8, 8, 12	Y(xD, E, J, K)	0	0	0	0	1	0	0	1
h76	15, 23, 10, 11, 13, 8, 11, 10	P-92R7	0	0	0	0	3	0	0	3
h77	15, 24, 10, 11, 12, 9, 9, 11	J-12f2	0	0	0	0	1	0	0	1
h78	15, 25, 11, 11, 12, 9, 9, 11	J-12f2	0	0	0	0	1	0	0	1
h79	14, 23, 10, 11, 13, 9, 9, 11	J-12f2	0	0	0	0	1	0	0	1
h80	15, 23, 10, 13, 11, 9, 11, 13	K-M9	0	0	0	0	1	0	0	1
h81	15, 22, 11, 14, 12, 9, 10, 13	K-M9	0	0	0	0	1	0	0	1
h82	14, 22, 10, 14, 11, 9, 11, 13	K-M9	0	0	0	0	1	0	0	1
h83	14, 25, 10, 13, 13, 9, 11, 11	R1a-SRY10831a	0	0	0	0	0	1	0	1
h84	14, 22, 10, 10, 14, 10, 10, 11	R1a-SRY10831a	0	0	0	0	0	1	0	1
h85	14, 24, 10, 10, 14, 11, 11, 12	R1a-SRY10831a	0	0	0	0	0	2	0	2
h86	14, 24, 10, 10, 14, 11, 11, 11	R1a-SRY10831a	0	0	0	0	0	4	0	4
h87	15, 23, 10, 11, 13, 9, 10, 12	Y(xD, E, J, K)	0	0	0	0	0	2	0	2
h88	15, 22, 10, 11, 12, 8, 9, 10	Y(xD, E, J, K)	0	0	0	0	0	1	0	1
h89	15, 25, 10, 11, 13, 8, 11, 10	P-92R7	0	0	0	0	0	1	0	1
h90	15, 22, 11, 11, 13, 9, 10, 11	Y(xD, E, J, K)	0	0	0	0	0	0	1	1
h91	13, 25, 10, 11, 13, 8, 10, 12	Y(xD, E, J, K)	0	0	0	0	0	0	1	1
h92	16, 25, 10, 11, 13, 9, 9, 12	J-12f2	0	0	0	0	0	0	1	1
h93	14, 22, 10, 15, 11, 9, 10, 11	K-M9	0	0	0	0	0	0	2	2
Total			24	23	22	29	20	21	7	146

allele at each locus is: 15, 22, 10, 11, 12, 8, 9, and 11 (locus order, DYS19, 390, 391, 392, 393, 437, 438, and 439), which generates haplotype 24, by far the most frequent and the most widespread.

There is a low level of sharing of haplotypes among the seven subcastes (Table 2). Eighty-three or 89.24% of the different haplotypes were found in only one subpopulation, 8 (8.60%) were shared between two populations, 1 (1.07%) was shared among three populations, and only 1 (haplotype 24, 1.07%) was found in six populations, with no haplotype present in all seven populations. All pairs of individuals with an identical haplotype belonged to the same haplogroup, suggesting very low levels of homoplasy across haplogroups, which was confirmed in the phylogenetic analysis. Haplotype diversity in the total sample was 0.9878, and it was similar across all seven groups, ranging from 0.9048 in the Kurava to 0.9921 in the Erra. Both a relatively high diversity (0.67590 and small range of variation (0.6430–0.6970) among subcastes were also observed in the autosomal microsatellites (Table 3).

The M-J network does not assign microsatellite haplotypes into any clear population grouping(s), suggesting that each subcaste has many different founder Y-haplotypes (see Figure 2). Such a conclusion is also supported by the networks constructed for each of the seven subcastes in which neither grouping nor an identifiable root haplotype can be identified (data not shown). Superposition of UEP markers on the overall microsatellite haplotype network, however, shows a strong clustering according to haplogroup. The only haplogroup not showing a clear pattern is J-12f2.

Given that haplotype 24 1) is in the center of the network, 2) is the most frequent haplotype, 3) is present in the highest number of populations (except the Doddi), 4) has the highest number of mutational links with other haplotypes, and 5) is constituted by the most frequent allele at each loci, it is probably the most recent common ancestor of all the sample. This view is further supported by the fact that even though there is a very low level of haplotype-sharing among the seven populations, this particular haplotype is present in 6 of 7 subcastes (absent in the Doddi). The TMRCA was estimated as 34,370 years before present (±5,390 ybp).

The contribution of genetic variance within Golla populations, regarding molecular differences between haplotypes and differences in haplotype frequencies, was 91.14%, a very high value, compared with the average amount of intrapopulation genetic variance (77%) among 20 world populations reported by Kayser et al. (2001). The portion of total genetic variance explained by differences between subcastes was 8.86% ($P < 0.0001$). If haplogroup frequencies are used, a similar between-populations value is obtained (5.05%, $P < 0.0001$). The Φ_{ST} value is almost three times higher than the G_{ST} value calculated from the analysis of 13 autosomal tri- and tetranucleotide microsatellites (Reddy et al., 2001a), but given the lower effective number of Y-chromosomes compared with autosomes (ratio of 1:4), this level of difference is expected.

The two-dimensional plot from MDS, using pairwise D_A values based on haplogroup and microsatellite haplotype frequencies, is shown in Figure 3. It is noteworthy that both distance matrices are highly correlated ($r = 0.7334$, $P = 0.0150$), and the plots form similar clusters of subcastes, even though they are in a slightly different relationship to each other. A large cluster comprises the Erra, Puja, Pokanati, and Karnam. The Punugu, Doddi, and Kurava are separated from each other and also from the large cluster. The Kurava are clearly the most distinct of all seven subcastes in either plot, with the Doddi and Punugu much closer to the other four subcastes than either is to the Kurava. The different positions of the Kurava, Doddi, and Punugu relative to the main cluster are explained to a great extent by the frequencies of haplogroups P-92R7 and R1a-SRY10831a. The Kurava lack both these haplogroups, the Doddi have no R1a-SRY10831a, and the Punugu have a low frequency of P-92R7 and the highest incidence of R1a-SRY10831a.

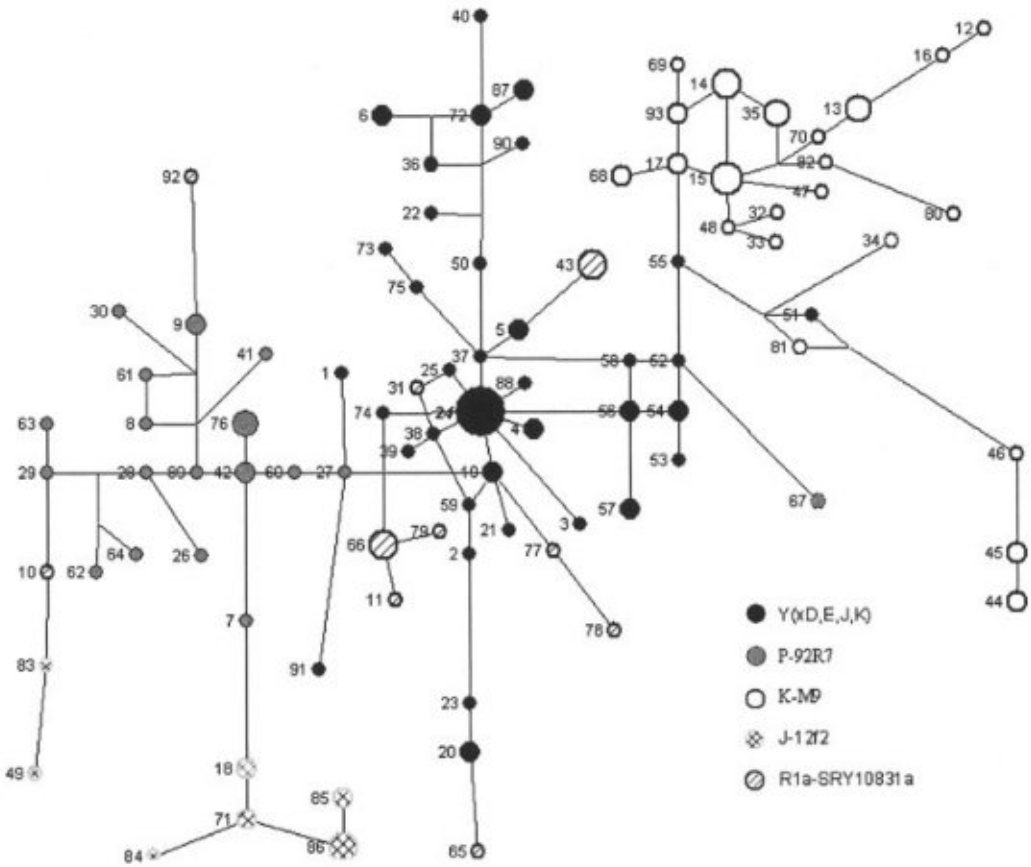

Fig. 2. Median-joining network of Y-microsatellite haplotypes in Golla sample.

DISCUSSION

The haplogroup Y(xD, E, J, K) is a Y-chromosome lineage not defined by the presence of any derived character state, and therefore represents an interior node of the NRY haplogroup tree (Y Chromosome Consortium, 2002) and is potentially paraphyletic. In an investigation of caste origins in Andhra Pradesh, Bamshad et al. (2001) found that the haplogroup Y(xD, E, J, K) split into three different lineages after scoring for three additional UEPs ($RPS4Y_{711}$, $DYS188_{792}$, and $DYS221_{136}$). Ninety percent of them were found to be haplogroup F-P14, defined by the presence of a C→T mutation at $DYS188_{792}$. Examination of the distribution of haplogroup Y(xD, E, J, K) in the phylogenetic network (Fig. 2) reveals that the majority form a fairly tight cluster, and therefore it is probable that the majority of Golla individuals bearing Y(xD, E, J, K) belong to F-P14.

The five haplogroups present in the Golla are among the six haplogroups found in India and Pakistan by Underhill et al. (2001). This finding suggests that our Golla sample is similar in composition to other Indian groups, at least with respect to the male gene pool. Y(xD, E, J, K), J-12f2, and R-1a SRY10891 are very common in Europe, and are present at high frequency in the Middle East and India, while K-M9 and P-92R7 are very common in India and Central Asia.

The high gene diversity of the subcastes suggests that each has many different founder Y-haplotypes, a view strongly supported by the absence of population-grouping branches in the general network and the lack of a noticeable root haplotype in the subcaste networks. In addition, the low level of haplotype-sharing and the significant genetic differentiation among the seven populations reflect their endogamous nature and indicate minimal gene flow.

Superposition of the UEP on the microsatellite haplotype network showed marked grouping according to haplogroups. A similar congruence between different types of Y markers was reported by Forster et al. (2000) when they empirically validated the accuracy of a global Y-chromosomal tree based on Y microsatellites by using slow evolving Y-UEPs. Further, the similarity between the eight Y-microsatellite-based MDS and the haplogroup-based MDS gives weight to the results, given that the two marker types involve very different timescales.

The TMRCA estimate of present-day Golla Y-chromosomes (34,370 ybp) clearly predates the formation of the caste system (6,000–4,000 ybp; Gadgil et al., 1997), indicating that the source of what was to become the Golla gene pool was well-formed before either the Dravidian or Indo-European migrations to the subcontinent. However, this date does not exclude the possibility that some Y-haplogroups in the Golla are the product of admixture with more recent migrants to Andhra Pradesh. It is important to emphasize that the TMRCA of Y-chromosomes in a population can antedate its establishment as a distinct cultural entity. Clearly, some, if not the majority, of Y-chromosomes in present-day Golla experienced a long evolutionary period prior to the establishment of this particular caste group.

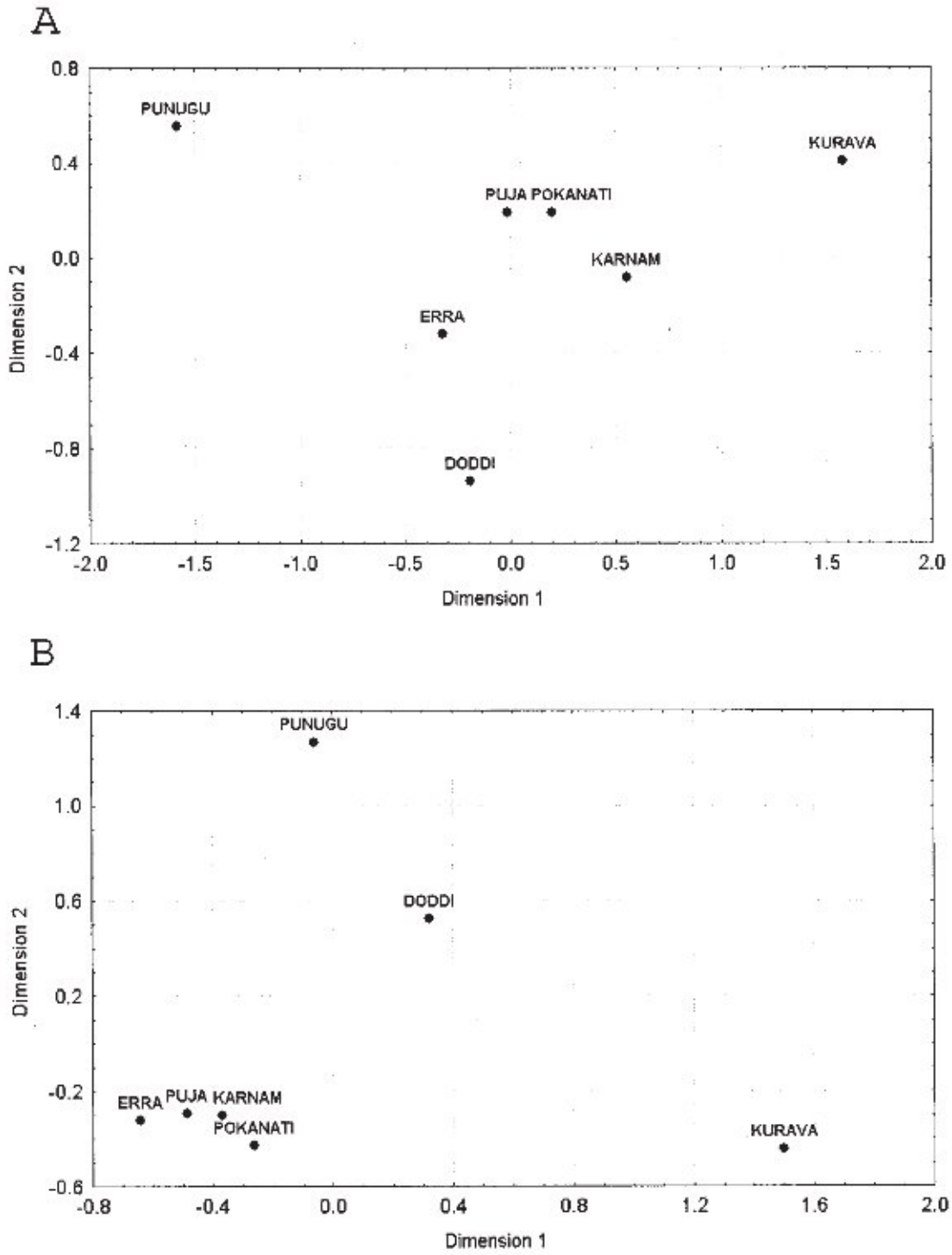

Fig. 3. Multidimensional scaling using D_A pairwise distances based on (**A**) Y-haplogroup frequencies (stress = 0.0000049) and (**B**) Y-microsatellite haplotypes (stress = 0.0000028).

The oral tradition of the Golla that claims the Kurava are not true members of the Golla (despite their occupational and cultural similarities) receives support from the present study of Y-chromosome variation. The Kurava are by far the most distinct of all seven subpopulations investigated for both Y-haplogroups and haplotypes. This distinction of the Kurava from the other subcastes is also evident from the analysis of 13 autosomal microsatellites and, to a lesser extent, from HLA class I and II allele distributions. These cumulative findings, based on both uniparental and biparental genetic markers, although based on an admittedly inadequate sample size in each case, suggest that the Kurava have a distinct origin from that of the Gollas of Andhra Pradesh, or that this group split from the ancestral Golla stock very early (Fig. 4).

The results of autosomal microsatellite scoring suggest that the recently evolved subcastes, Doddi, Karnam, Erra and Pokanati, arose by fissioning from the Punugu rather than the Puja ancestral group (Reddy et al., 2001a; Crawford et al., 2001). However, Y-chromosome data (both haplogroups and haplotypes) clearly indicate that the Puja, and not the Punugu, are closest genetically (very close) to 3 of the 4 recent evolved subcastes (Erra, Karnam, and Pokanati). In fact, the Punugu are distant from all four subcastes, suggesting that it was this group, rather than the Puja, that fissioned from the rest of the Golla in southern Andhra Pradesh quite early (Fig. 3). Analysis of autosomal microsatellite data gives an equivocal finding on the matter of which group, the Puja or Punngu, first split off from the main Golla cluster (Reddy et al., 2001a). Two types of genetic distance measure were used in analyzing the autosomal microsatellites, D_A and D_{SW}, and the trees generated some minor differences in the clustering of subcastes, even though the correspondence between the two matrices was very high. This is one of the reasons why, in some cases, it is better to use an MDS plot instead of a tree, given that the MDS does not create obligatory bifurcations. Whereas D_{SW} values indicated that the Puja split first from the other Golla, D_A values demonstrated that the Punugu were the first to split from the main cluster. That the Puja are better candidates than the Punugu as common ancestors of the Pokanati, Doddi, Karnam, and Erra subcastes also receives support from analysis of the HLA data of Crawford et al. (2001). The Doddi, Puja, and Pokanati form a very tight cluster for the HLA class II DRB1 locus, and sit on the same branch of the tree (the Erra were not included in this study); the Punugu and Kurava lie on a separate branch of the tree. The probable sequence of fission events that resulted in the present-day subcastes of the Golla of Andhra Pradesh is shown in Figure 4. This scheme is based on the cumulative evidence from three sets of genetic markers; Y-chromosomal, HLA class I and II, and autosomal microsatellites. It differs from previous representations of the historical process of fissioning among the Golla (Reddy et al., 2001a; Crawford et al., 2001) in its identification of the Puja as the ancestors of the Karnam, Pokanati, Doddi, and Erra. The present study also differs from that of Reddy et al. (2001a) in detecting no evidence from Y-chromosomal markers of differentiation between the western (Doddi and Karnam) and eastern (Erra and Pokanati) subcastes.

A series of Mantel tests was performed on the D_A distance matrices derived from Y-haplogroup, Y-STR, autosomal STR, and HLA allele frequencies, respectively. These tests were limited to the six subcastes (minus Erra) for which HLA data were available. There is a very high and significant correlation ($r = 0.8569$, $P = 0.0097$) between Y-chromosome STRs and autosomal STRs, and also between Y-haplogroups and Y-STRs ($r = 0.7317$, $P = 0.0389$). No correlation was found between the HLA and any other matrix, suggesting that other evolutionary processes such as admixture and selection maintain HLA variability. The strong concordance between uniparental and biparental loci strongly supports the hypothesis of Golla phylogeny described above, and also confirms the power of STRs on the Y chromosome and autosomes for the reconstruction of the evolutionary history of closely related human groups.

According to Thurston (1975) and the oral traditions reported by the elders of the Golla community, the Erra and Karnam subcastes were expected to show evidence of admixture with non-Golla groups. This is because the Erra are said to be the result of exogamous unions between Golla women and Brahmin men, and the Karnam are supposed to have admixed with an agricultural caste group called Reddy (which has certain occupational similarities to the Golla). Y-chromosome-specific markers provide an excellent means of investigating the extent of male admixture. The data from both Y-haplogroups and haplotypes provide no evidence of a distinct male gene pool in the Erra; in fact, they cluster tightly with the Karnam, Pokanati, and Puja (Fig. 3). This does not mean that the Erra did not arise from Erra women mating with some Brahmin men, only that if they did, these males did not differ in Y-chromosome composition from males of the Golla caste. Similarly, the present Y-chromosome data give no support for the presence of substantial alien male-specific genes in the Karnam. The Karnam, like the Erra, plot close to three other subcastes (Puja, Pokanati, and Erra). The Karnam are also not distinguished from most of the other Golla on the basis of 13 autosomal microsatellite markers (Reddy et al., 2001a). However, the Karnam are very different from the other Golla on the basis of HLA class II DRB1 alleles (Crawford et al., 2001), clustering more with north Asian populations rather than south Asians. Crawford et al. (2001) suggested that the Karnam had experienced the greatest gene flow of all the subcastes, and most probably from invaders in historical times such as the Moguls.

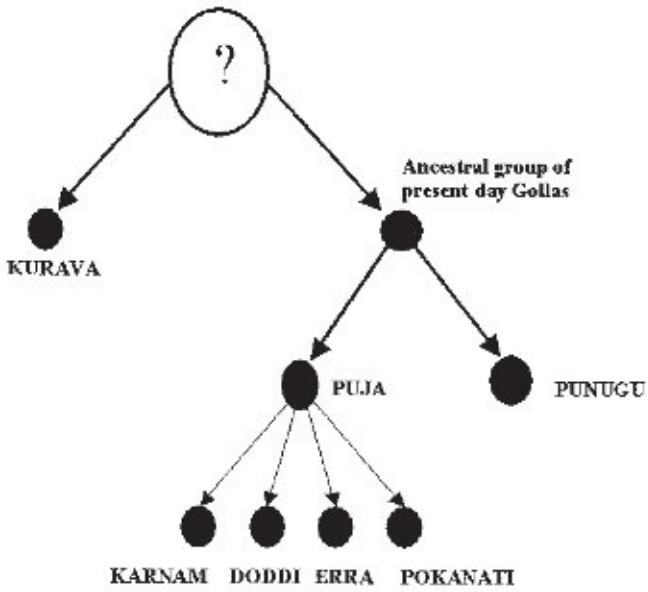

Fig. 4. Schematic diagram depicting probable sequence in formation of Golla subcastes.

CONCLUSIONS

A reasonable conclusion from the cumulative evidence of Y-chromosome UEPs and microsatellites, autosomal microsatellites, and HLA alleles is the following phylogeny of the seven Golla subpopulations of Andhra Pradesh. Firstly, the Kurava may be a population wholly distinct from the Golla subcastes, or they split from the main group very early in the caste's evolution. This finding corroborates the oral history of the Golla elders as

reported in Reddy et al. (2001a,b). Secondly, the Puja are the subcaste that is closest genetically to the other Golla, and they are, therefore, the probable ancestor of these more recently evolved groups. The Punugu, on the other hand, are much more distant genetically from the others, indicating that this subcaste split off before the Karnam, Erra, Pokanati, and Doddi were formed by fissioning of the Puja. In addition, the Y-chromosome data provide no support for the idea that some of the subcastes may have experienced male-specific gene flow from outside caste groups, such as the Brahmin.

ACKNOWLEDGMENTS

We acknowledge the help of colleagues and friends, and in particular P. Chengal Reddy and P. Venkatramana, at Sri Venkateswara University, Timpati, for assistance with fieldwork. The cooperation and hospitality of the members of the Golla community are gratefully appreciated. D.C. is the grateful recipient of a La Trobe University Postgraduate Research Scholarship. We also thank Kristin Melvin for statistical assistance.

LITERATURE CITED

Arnaiz-Villena A, Karin M, Bendikuze N, Gomez-Casado E, Moscoso J, Silvera C, Oguz FS, Sarper Diler A, De Pacho A, Allende L, Guillen J, Martinez Laso J. 2001. Alleles and haplotypes in the Turkish population: relatedness to Kurds, Armenians, and other Mediterraneans. Tissue Antigens 57:308–317.

Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C. 2000. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res 15;28:8.

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB. 2001. Genetic evidence on the origins of Indian caste populations. Genome Res 11:994–1004.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G, Siniscalco M. 1985. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230:1403–1406.

Crawford MH, Reddy BM, Martinez-Laso J, Mack SJ, Erlich H. 2001. Genetic variation among the Golla pastoral caste subdivisions of Andhra Pradesh, India, according to the HLA system. Hum Immunol 62:1031–1041.

Dronamraju KR. 1964. Mating systems of the Andhra Pradesh people. Cold Spring Harbor Symp Quant Biol 29:81–4.

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. A short tandem repeat-based phylogeny for the human Y chromosome. 2000. Am J Hum Genet 67:182–196.

Gadgil M, Joshi N, Shambu U, Manoharan S, Patil S. 1997. Peopling of India. In: Balasubrumanian D, Appaji Rao N, editors. The Indian human heritage. Hyderabad, India: Universities Press. p 100–129.

Gonem A. 1996. The encyclopedia of the peoples of the world. Jerusalem: Jerusalem Publishing House.

Hammer MF. 1994. A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761.

Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L, et al. 1997. Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133, 141–149.

Kayser M, Krawczac M, Excoffier L, Dieltjes P, Corach D, Pascalli V, Gehrig C, Bernini L, Jespersen J, Bakker E, Roewer L, De Knijff L. 2001. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018.

Kruskal J. 1964. Nonmetric multidimensional scalling: a numerical method. Psychometrika 29:28–42.

Mathias N, Bayés M, Tyler-Smith C. 1994. Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nei M, Tayima F, Tayeno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170.

Reddy BM, Sun G, Rodriguez-Luis J, Crawford MH, Heman NS, Deka R. 2001a. Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. Hum Biol 73:175–190.

Reddy BM, Pfeiffer A, Crawford MH, Langstieh BT. 2001b. Population substructure and patterns of quantitative variation among the Gollas of southern Andhra Pradesh, India. Hum Biol 73:291–306.

Reid RM. 1973. Inbreeding in human populations. In: Crawford MH, Workman PL, editors. Methods and theories of anthropological genetics. Albuquerque: University of New Mexico Press. p 83–116.

Rohl A. 2001. NETWORK 4.1. http://www.fluxus-engineering.com

Ruhlen M. 1991. A guide to the world's languages. Volume 1: classification. Stanford: Stanford University Press.

Saillard J, Forster P, Lynnerup N, Bandelt H, Norby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726.

Sambrook J, Fritsch E, Maniatis T. 1989. Molecular cloning. A laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Schneider S, Roessli D, Excoffier L. 2000. ARLEQUIN version 2000: software for population genetic data analysis. Geneva: Genetics and Biometry Laboratory, University of Geneva.

Srinivasulu K. 2002. Caste, class and social articulation in Andhra Pradesh: mapping differential regional trajectories. Background report for the Livelihood Options Project. Overseas Development Institute, London.

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L. 1999. Y Chromosome evidence for a northward migration of modern humans in East Asia during the last ice age. Am J Hum Genet 65:1718–1724.

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

Thurston E. 1975. Castes and tribes of southern India. Delhi: Cosmo.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005.

Underhill P, Passarino G, Lin A, Shen P, Mirazo M, Lahr M, Foley R, Oefner P, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62.

Whitfield LS, Sulston JE, Goodfellow PN. 1995. Sequence variation of the human Y chromosome. Nature 378:379–380.

Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348.

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50–61.