mtDNA Variation in Caste Populations of Andhra Pradesh, India

MICHAEL BAMSHAD,[1] ALEXANDER E. FRALEY,[2] MICHAEL H. CRAWFORD,[3] REBECCA L. CANN,[4] BASKARA R. BUSI,[5] J.M. NAIDU,[5] AND LYNN B. JORDE[2]

Abstract Various anthropological analyses have documented extensive regional variation among populations on the subcontinent of India using morphological, protein, blood group, and nuclear DNA polymorphisms. These patterns are the product of complex population structure (genetic drift, gene flow) and a population history noted for numerous branching events. As a result, the interpretation of relationships among caste populations of South India and between Indians and continental populations remains controversial. The Hindu caste system is a general model of genetic differentiation among endogamous populations stratified by social forces (e.g., religion and occupation). The mitochondrial DNA (mtDNA) molecule has unique properties that facilitate the exploration of population structure. We analyzed 36 Hindu men born in Andhra Pradesh who were unrelated matrilineally through at least 3 generations and who represent 4 caste populations: Brahmin (9), Yadava (10), Kapu (7), and Relli (10). Individuals from Africa (36), Asia (36), and Europe (36) were sampled for comparison. A 200-base-pair segment of hypervariable segment 2 (HVS2) of the mtDNA control region was sequenced in all individuals. In the Indian castes 25 distinct haplotypes are identified. Aside from the Cambridge reference sequence, only two haplotypes are shared between caste populations. Middle castes form a highly supported cluster in a neighbor-joining network. Mean nucleotide diversity within each caste is 0.015, 0.012, 0.011, and 0.012 for the Brahmin, Yadava, Kapu, and Relli, respectively. mtDNA variation is highly structured between castes ($G_{ST} = 0.17$; $p < 0.002$). The effects of social structure on mtDNA variation are much greater than those on variation measured by traditional markers. Explanations for this discordance in-

[1] Department of Pediatrics, Health Sciences Center, University of Utah, Salt Lake City, UT 84132.
[2] Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84132.
[3] Department of Anthropology, University of Kansas, Lawrence, KS 66045.
[4] Department of Genetics and Molecular Biology, University of Hawaii at Manoa, Honolulu, HI 96822.
[5] Department of Anthropology, Andhra University, Waltair, Andhra Pradesh, India.

Human Biology, February 1996, v. 68, no. 1, pp. 1–28.

KEY WORDS: mtDNA, ANDHRA PRADESH, CASTES, EVOLUTION, POPULATION STRUCTURE

clude (1) the higher resolving power of mtDNA, (2) sex-dependent gene flow, (3) differences in male and female effective population sizes, and (4) elements of the kinship structure. Thirty distinct haplotypes are found in Africans, 17 in Asians, and 13 in Europeans. Mean nucleotide diversity is 0.019, 0.014, 0.009, and 0.007 for Africans, Indians, Asians, and Europeans, respectively. These populations are highly structured geographically ($G_{ST} = 0.15; p < 0.001$). The caste populations of Andhra Pradesh cluster more often with Africans than with Asians or Europeans. This is suggestive of admixture with African populations.

South Asian evolutionary history is notable for successive migrations of many different peoples from Central and East Asia, Africa, and Europe into the Indian subcontinent (Majumdar 1958). Among these groups the rate of gene flow, level of admixture, effective population size, founder events, and frequency of fission varied extensively (Malhotra and Vasulu 1993; Cavalli-Sforza et al. 1994). As a result, contemporary India—with more than 850 million inhabitants (1991 census)—is an agglomeration of thousands of diverse ethnic groups with different religions, languages, social structures, kinship systems, economic structures, and settlement patterns. Attempts to order this variation led to the construction of numerous systems of population classification based on social, linguistic, religious, morphological, and genetic characteristics (Malhotra 1978; Majumder et al. 1990). To this end, Indians have been the focus of numerous anthropological studies using morphometrics (Majumder et al. 1990; Sirajuddin et al. 1994), dermatoglyphics (Singh 1978; Malhotra et al. 1980), protein polymorphisms (Roychoudhury 1974; Balakrishnan 1978, 1982; Char and Rao 1986; Char et al. 1989), mitochondrial DNA (mtDNA) (Semino et al. 1991; Soodyall and Jenkins 1992; Mountain et al. 1995), and nuclear DNA polymorphisms (Labie et al. 1989; Murthy et al. 1993).

The broadest division of Indians distinguishes tribal from nontribal groups. Although the definition of *tribe* appears to be somewhat arbitrary, it often refers to populations considered aboriginal, inhabiting the Indian peninsula before the immigration of pastoral nomads from western and Central Asia. These aboriginal populations represent the descendants of Paleolithic population expansions into South Asia. The source of these expansions is disputed, although it has been argued that they were of African (Maloney 1974; Chandler 1988) or Australoid origin (Cavalli-Sforza et al. 1994). The size of different tribal groups varies from a few hundred (e.g., the Andamanese) to a few million individuals (e.g., the Gonds), with 400 contemporary tribes (Majumder and Mukherjee 1993) constituting 7.5% of the total Indian population (Sirajuddin et al. 1994).

Most Indians are Hindu (82%), and they constitute the largest nontribal population in India. The most salient feature of Hindu social organization is its former division into approximately 2000 castes (Karve 1968) (the caste

system has been formally abolished by the Indian government). The term *caste* was introduced by the Portuguese in the sixteenth century and is somewhat ambiguous. It is largely defined by two elements: (1) *jathis,* the endogamous community defined by hierarchical, occupational, and geographic limits, and (2) *varna,* broader categories of contemporary castes. The caste structure was introduced into India by immigrating Central Asian populations approximately 3000–4000 years before present and is marked archeologically by the appearance of Painted Grey Ware (Thapar 1980). These immigrants also introduced iron, the domestic horse, the Indo-Aryan languages, the patrilineal kinship system, and a mixed pastoral and agrarian economy. The manifest for the caste system is found in the hymns of the *Rig Veda* in which the extant population was partitioned into four classes or varnas: the Brahmins (priests), the Kshatriyas (warriors), the Vaishyas (traders), and the Sudras. It is thought that Central Asians admixed with the upper varnas (i.e., Brahmins, Kshatriyas, and Vaishyas) and that the displaced Sudras represent aboriginals with African, Southeast Asian, and/or Australian admixture (Majumdar 1958; Balakrishnan 1978). A fifth varna, the Pancham, was added to include the formerly untouchable castes. Each varna was identified by occupation and was associated with a status. Each caste was considered strictly endogamous, whereas lineages within a caste group were exogamous (*gotra*). This social stratification was rationalized and strengthened by ritual and religious philosophy.

The internal boundaries of the caste structure were not absolutely rigid, and levels of endogamy varied widely (Malhotra and Vasulu 1993). Mixing of varnas occurred through the processes of hypergamy (an upper-caste man mating with a lower-caste woman) and hypogamy (a lower-caste man mating with a higher-caste woman). Hypergamy led to approved unions, and thus an upper-caste man could marry a woman from a lower caste and remain within his varna. Hypogamy was discouraged by forcing an upper-caste woman to remain in the lower varna of her husband. Thus women had more mobility in the caste structure than men. These mating practices generated many new subdivisions (e.g., subcastes) with varying amounts of social privileges (Tambiah 1973). Admixture was most frequent within the same caste cluster and least frequent between religious groups. The frequency of admixture also varied with differences in education and location. Rates of mixed marriages in multicaste villages were approximately 2/10,000 marriages, whereas in urban areas it was as high as 3/100 marriages. Mixing occurred predominantly in the middle castes, leading to the hypothesis that upper and lower castes were homogeneous across the subcontinent, whereas the middle castes exhibited higher levels of genetic heterogeneity. Balakrishnan (1978) identified such a pattern of genetic variation at the ABO and Rh loci.

The history and structure of caste and tribal populations remain active areas of investigation. By extrapolating from archeological, linguistic, and genetic data, researchers have attempted to (1) determine the ancestral con-

tinental populations of Indian groups (Cavalli-Sforza et al. 1988; Nei and Roychoudhury 1993), (2) differentiate tribal populations from neighboring nontribal groups (Chakraborty and Yee 1973; Balakrishnan 1978; Malhotra 1978; Mukherjee et al. 1979; Char et al. 1989; Roychoudhry 1984, 1992), (3) demonstrate geographic structuring of caste and tribal populations (Chakraborty et al. 1977; Balakrishnan 1982; Roychoudhury 1984, 1992; Sirajuddin et al. 1994), and (4) explore the relationships between tribal and nontribal groups as predicted by ethnographic records (Majumder and Mukherjee 1993). Based on blood group loci, red cell enzymes, and serum proteins, a few broad generalizations have emerged: (1) Genetic variation is much larger within castes and tribes than between populations (Roychoudhury 1982), (2) substantial genetic differences are observed between caste and tribal groups, (3) genetic variation among tribal groups is structured geographically, (4) in tribal groups geographic structuring is more important than linguistic affiliation, (5) within certain geographic regions genetic distances are substantial between caste groups, (6) caste groups belonging to different varnas exhibit considerable genetic distances, and (7) geographically clustered castes are more similar regardless of social hierarchy.

To further investigate the pattern and dynamics of genetic variation in the Hindu caste structure, we have examined mtDNA variation in caste populations of Andhra Pradesh and compared them to continental populations. The four caste populations chosen—Brahmin, Kapu, Yadava, and Relli—demonstrate discrete social, economic, and occupational stratification. Brahmins represent the highest caste and varna. The Kapu are included in the upper middle castes and the Yadava in the lower middle castes. Both groups belong to the Sudra varna. The Relli are found in the lowest varna, the Pancham. Based on historical evidence, we hypothesize that (1) the Brahmin will exhibit closer affinity to Europeans and the Relli will be closer to African and Asian populations; (2) given lower levels of endogamy, genetic variation will be highest in the middle level castes (Kapu and Yadava) and lowest in the upper (Brahmin) and lower castes (Relli); (3) substantial genetic differences will be found between castes; and (4) as a result of admixture, the middle castes will share more mtDNA types than the upper and lower castes.

Certain properties of mtDNA make it an optimal marker to explore caste structure. The mtDNA molecule is inherited as a single locus, yet it is extremely polymorphic and therefore highly informative for discriminating between populations. The mutation rate of human mtDNA is 10 times higher than the mutation rate of noncoding and nondegenerate sites in human nuclear DNA (Brown et al. 1979, 1982; Wilson et al. 1985; Li and Sadler 1991), meaning that local population differentiation will be accelerated compared with nuclear DNA. Because mtDNA is strictly maternally inherited (Giles et al. 1980), the effective population size is about 25% that of the nuclear genome (Birky et al. 1983; Takahata 1991). Thus mtDNA can be more sensitive than nuclear DNA for detecting population structure (Stoneking et al. 1990).

The lack of recombination (Olivo et al. 1983; Merriwether et al. 1991) allows for mtDNA types to be related as matrilines. These characteristics have encouraged the use of mtDNA to study the origin and dispersal of contemporary human populations (Johnson et al. 1983; Cann et al. 1987; Vigilant et al. 1989; Horai and Hayasaka 1990; Schurr et al. 1990; Stoneking et al. 1990; Di Rienzo and Wilson 1991; Ward et al. 1991; Ballinger et al. 1992; Shields et al. 1992; Horai et al. 1993; Lum et al. 1994; Jorde et al. 1995). Excoffier (1990) and Merriwether et al. (1991) suggested that mtDNA variation is not in mutation-drift equilibrium, but it is unclear whether this is a result of selection or of rapid population expansion (Rogers and Jorde 1995).

The 1122-bp control region of the human mitochondrial genome is divided into three domains dependent on base composition and degree of conservation among species. Two hypervariable regions, HVS1 and HVS2, bracket a highly conserved 200-bp G-rich region and contain more than 90% of the polymorphisms reported in the mtDNA molecule (Vigilant et al. 1989). Sixty percent of the polymorphisms in the control region are found in HVS1 and 40% in HVS2. As a consequence, analysis of HVS1 has been the emphasis of many recent investigations of human population history and structure. HVS2 is the most important functional portion of the control region, containing the light and heavy strand promoters and conserved sequence blocks 1–3 (Saccone et al. 1991). The analysis reported here is based on HVS2.

Subjects, Materials, and Methods

Populations. One hundred forty-four individuals from 4 populations distributed on the continents of Europe (36), Africa (36), and Asia (36) and the subcontinent of India (36) were chosen for analysis of the mtDNA control region. The European sample consisted of 28 individuals of mixed English and Northern European ancestry collected from unrelated men in Utah and 8 men of French ancestry from CEPH (Centre d'Étude du Polymorphisme Humain) kindreds. The 36 Africans included 5 Biaka Pygmies from the Central African Republic and 1 Mbuti Pygmy from northern Zaire (Human Genetic Mutant Cell Repository; Camden, NJ), 5 Zulu, 5 Pedi, 5 Tswana, 5 Sekele, 5 Tsonga, and 5 Sotho from southern Africa (samples provided by T. Jenkins). Asians were represented by 6 Malaysians, 5 Japanese (samples provided by J. Kidd), 10 Cambodians (samples collected in California) (Human Genetic Mutant Cell Repository; Camden, NJ), 5 Vietnamese, and 10 Chinese (Jorde et al. 1995).

Indian samples were collected from four castes living in Visakhapatnam, Andhra Pradesh. They included 9 Brahmin, 10 Yadava, 7 Kapu, and 10 Relli. The sampled castes extend from the uppermost to the lowest caste. All four groups speak Telugu and are patrilineal. All donor Indians were inter-

viewed to determine their occupation and knowledge of maternal ancestry and to ensure their inclusion within the appropriate ethnic group.

Extraction, Amplification, and Sequencing. CEPH, Asian, and African Pygmy mtDNAs were prepared from lymphoblast cell lines. DNAs from Northern Europeans were prepared from whole blood. All DNA extractions were performed according to the method described by Bell et al. (1981). DNAs from the remaining Africans were prepared from whole blood, and DNA extractions were performed according to the method described by Sykes (1983). DNA concentrations were standardized to 200 ng/ml. Indian whole blood was collected and centrifuged, and the buffy coat was smeared on blotter paper and dried. DNA from blotter paper was extracted by placing approximately 2.0 cm^2 of stained paper in 500 μl of PCR buffer/detergent solution [10 m*M* Tris (pH 8.0), 50 m*M* KCl, 2.5 m*M* $MgCl_2$, 0.1 mg/ml gelatin, 0.45% NP40, 0.45% Tween 20] and then incubating the mixture at 56°C for 24 hr. After removal of the paper, 0.6 μl of proteinase K (10 mg/ml) was added, and the mixture was incubated at 26°C for 2 hr. DNA was extracted with equal volumes of phenol/chloroform (1:1 wt/vol) and a 1-butanol wash. DNA was collected by centrifugal filtration with an ultrafree-MC 30,000 NMWL filter unit (Millipore) per the manufacturer's instructions; it was stored in TE.

The mtDNA region corresponding to hypervariable sequence 2 (HVS2) from Vigilant et al. (1989) was amplified by the polymerase chain reaction (PCR) (Saiki et al. 1985; Kocher et al. 1989) in 1 × buffer [10 m*M* Tris (pH 8.3), 50 m*M* KCl, 1.5 m*M* $MgCl_2$] using 5 μl of template (12.5% of recovered product from each Indian), 250 μM dNTPs, 50 pmol of each primer, and 1 u *Taq* DNA polymerase in a total reaction volume of 50 μl. Reactions were completed in a Perkin-Elmer 9600 with the following conditions: denaturation at 94°C for 2 min for 1 cycle; denaturation at 94°C for 15 s, annealing at 53°C for 15 s, extension at 72°C for 15 s for 35 cycles; and extension at 72°C for 5 min for 1 cycle. The Indian samples required a second amplification using 5 μl of the first reaction as a template for the second PCR. DNA was amplified with the following primer sets: UL29 (5′-TGTAAAACGACGGCCAGT*GGTCTATCACCCTATTAACCAC-3′) and BRH408 (5′-BIOTIN-CAGGAAACAGCTATGACC*CTGTTAAAAGTGCATACCGCCA-3′); and BUL29 (5′-BIOTIN-TGTAAAACGACGGCCAGT*GGTCTATCACCCTATTAACCAC-3′) and RH408 (5′-CAGGAAACAGCTATGACC*CTGTTAAAAGTGCATACCGCCA-3′). The numbers of the primer designations identify the 3′ ends according to the Cambridge reference sequence (CRS) (Anderson et al. 1981), and L and H denote the light and heavy strands of the mtDNA molecule. To the conventional L29 and H408 primers, a biotinylated oligonucleotide leader (21M13 and M13RP1, respectively) was added at the asterisk (Arnold and Hodgson 1991).

Amplified products were separated and prepared for sequencing using Dynabeads® M-280 Streptavidin beads (Dynal) and a Magnetic Particle Concentrator (MPC®) according to the manufacturer's instructions. The nonbiotinylated primer of each set was radiolabeled using γ^{32}P-ATP and polynucleotide kinase. After extensive sequencing of the HVS2 region in 20 individuals (data not shown), a 200-bp region of the HVS2 region corresponding to bases 80–280 of the CRS was found to contain 90% of the polymorphisms in the region and thus was examined in all individuals. Sequencing was completed using an AmpliTaq® cycle sequencing kit (Perkin-Elmer) by denaturing at 95°C for 1 min and extending at 60°C for 1 min for 20 cycles. Samples were electrophoresed through 4–8% polyacrylamide gels prepared with Sequenase reagents (National Diagnostics) and affixed to a glass plate. Gels were fixed in 5% glacial acetic acid/5% methanol for 20 min, rinsed with deionized water, and baked at 65°C for 60 min. Sequencing ladders were visualized by autoradiography.

Sequence Analysis. mtDNA variation within populations was quantified by two measurements. The first is the gene diversity value h, calculated as

$$h = \left(1 - \sum x_i^2\right) n/(n - 1), \tag{1}$$

where x_i is the frequency of a type and n is the number of individuals in the sample (Nei 1987). The probability of identity

$$\text{PI} = \sum x_i^2 \tag{2}$$

is the probability of randomly picking two individuals from a population with the same haplotype. As the number of sampled nucleotides increases, h approaches 1 and thus provides little information. A more appropriate measure is the average number of nucleotide differences per site, or nucleotide diversity, which is estimated by

$$\pi = [n/(n - 1)] \sum x_i x_j \pi_{ij}, \tag{3}$$

where n is the number of DNA sequences examined, x_i and x_j are the population frequencies of the ith and jth type of DNA sequences, and π_{ij} is the proportion of nucleotides that differ between the ith and jth types of DNA sequence. The variance of this measure is described by Eq. (10.7) of Nei (1987). This statistic approximates mean pairwise sequence divergence, corrected for the number of nucleotides sampled.

Individual genetic distance measurements were calculated by the program DNADIST in the computer package PHYLIP (Felsenstein 1993) using Kimura's two-parameter model (Kimura 1980). The transition:transversion ratio was set at 30:1 and was derived empirically. Average within- and between-

population genetic distance measurements were calculated according to Eqs. (8.21), (8.24), and (8.25) of Nei (1987). One thousand replicates of each data set were generated for bootstrap analysis with the SEQBOOT program in PHYLIP.

The proportion of total mtDNA variation attributable to between-population differences was measured using Nei's coefficient of differentiation (G_{ST}) and estimated using Eqs. (8.26) and (8.27) of Nei (1987). Significance levels of G_{ST} were estimated using a permutation procedure in which each individual was randomly assigned to one of the subpopulations, maintaining the same sample size for each subpopulation. Replication of this procedure 1000 times provided a distribution of expected G_{ST} values from which a significance level was obtained (Stoneking et al. 1990). G_{ST} is a convenient measure of population differentiation because it reaches equilibrium quickly (Crow and Aoki 1984; Rogers and Harpending 1986) and is relatively insensitive to the number of population subdivisions (Takahata and Palumbi 1985).

Unrooted neighbor-joining trees (Saitou and Nei 1987) relating populations were constructed from the genetic distance matrices using the NEIGHBOR program in PHYLIP after randomizing the input order of populations. Unique Indian caste haplotypes were also connected by an unrooted neighbor-joining network. Bootstrapped majority-rule consensus trees were created using the programs CONSENSE and DRAWGRAM in PHYLIP. These trees provide an estimate of the robustness of each branch point of a network or tree, and only branch points found in greater than 50% of trees were considered meaningful.

Results

Nucleotide Sequences. In comparisons between continental populations, we found that a 200-bp region of HVS2 spanning base pairs 80–280 of the CRS contains most of the polymorphic sites. Thirty polymorphic sites were identified. Two hundred fifty-five transitions and 9 transversions were distinguished, producing 69 unique haplotypes in 144 individuals (Table 1). Eighty percent of the polymorphisms are found between bp 141 and bp 160 and between bp 181 and bp 200. Seven transversions were found in Africans, one in an Indian, and one in a European. The transition:transversion ratio is 28.3:1 and is consistent with estimates obtained by other investigators sequencing different areas (e.g., HVS1) of the control region (Horai and Hayasaka 1990; Lum et al. 1994).

According to the CRS, bp 263 of the light mtDNA sequence is occupied by adenine (A). In this analysis 130 individuals, including all the Europeans, have guanine (G) at this position. In a larger data set containing 278 individuals sequenced in our lab (Jorde et al. 1995), only 29 individuals (27 Africans and 2 Indians) have adenine at bp 263. Not one of 90 Europeans had an adenine at bp 263. In addition, guanine is conserved at this position in com-

Table 1. Sequence Variation and Lineage Groups

	1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2[a]	
	9 9 3 4 4 5 5 5 8 8 8 8 9 9 9 9 0 0 0 1 1 1 1 2 3 4 4 4 5 6	
	3 5 9 3 6 0 1 2 2 5 6 9 4 5 8 9 0 4 7 2 4 5 7 8 6 1 6 7 6 3	
Lineage		*Identity*[c]
1[b]	A A T G T C C T C G C A C T C T A T G T A A T G T A T G C G	EUR, AF, AS, IN
2	G . . . . . C . . . G . . . . . C A . . . . . C . . A . A	AF
3	G . . . . . C . . . G . . . . . C A . . . . . C . . A . .	AF
4	. . . . C . . C . . . . . C . . . . A . A	AF
5	. . . . C . . C . . . G . C T . . . A . . . . A . A	AF
6	. . . . . C . . . C . . . . . C A . . . . . C . . A . .	AF
7	. . . . C T . . . . . C . . . . A . A	AF
8	. . . . C . . C . . . . A . A	AF
9	. . . . C . . C . . . . . C T . . . . A . A	AF
10	. . . . C . . C . . . G . C T . . . . A . A	AF
11	. . . . . C T . A C . C . . . . C . . A . .	AF
12	G . . . . . A . G . . . . C . . A . .	AF
13	G C . . . . . A . G . . . . C . . A . .	AF
14	. . . . C . . C . . . . . C . . . . G . A . A	AF
15	. . . . C . . C . . . . . C . . . . A	AF
16	. . . . . C T A . . . C . . . . A . .	AF
17	. . . . C . . C . . . . A	AF
18	. . . . . T C T . A C . . . . A . .	AF
19	. . . . . T . C . . . . . C	AF
20	. . . . C . . C	AF
21	. . . . C . . C . . . . . C . . G	AF
22	. . . . . T . . . A . G	AF
23	. . . . C . . C . . . . . C T	AF
24	. . . . . T . . . . . G	AF
25	. . . . . T . C . . . G . . . . G	AF
26	. . . . A	IN (R)
27	. . . . . T . . . . A	IN (Y)
28	. . . . C . . .	IN (B)
29	. . . . C A	IN (K)
30	. . . . . C . . . . A	IN (K)
31	. . . . . C . . . . . C . . . . . G . . . . . . G .	IN (R)
32	. . C . C . . C T . . . . C	IN (B)
33	. . . A . . . . C	IN (R)
34	. . . . C . . C T . . . . C	IN (R)
35	. . . . . C . . . . C	IN (B)
36	. . . . . T C . . . . . A . . G	IN (B)
37	. . . . . C . . . . . G . A	IN (Y)
38	. . . A . . . . T	IN (Y)
39	. . . . . T	IN (Y)
40	. . . . . C . . . . . C . . . . . G	IN (R)
41	. . . . . G	IN (Y)
42	. . . A . . . C	IN (Y)
43	. . . . C . . . . . C	AS

Table 1. Continued

Lineage	1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2[a] 9 9 3 4 4 5 5 5 8 8 8 8 9 9 9 9 0 0 0 1 1 1 1 2 3 4 4 4 4 5 6 3 5 9 3 6 0 1 2 2 5 6 9 4 5 8 9 0 4 7 2 4 5 7 8 6 1 6 7 6 3	Identity[c]
44	. . . . . C . . . . . C	AS
45	. . . . C . . C . . . . . C	AS
46	. . . . . C	AS
47	. . . . . T	AS
48	. . . AC	AS
49	. . . . CT . . . . . C	AS
50	. . . . . GTC . . . . A	AS
51	. . . . C . . . . . C	AS
52	. . . . . T . . . . . C	AS
53	. . . . . A . . . . . A	EUR
54	. . . . C . . . . . A	EUR
55	. . . . . A	EUR
56	. . . . . C . . . . A	EUR
57	. . . . . A	EUR
58	. . . . . A . G . . . . . C	EUR
59	. . . . . AG	EUR
60	. . . . . C	EUR
61	. . . . C . . C . . . . . C	AF, IN (K)
62	. . . . . C . . . . . C	AF, IN (B)
63	. . . . C . . . . . C	AF, AS
64	. . . . . C	AF, AS
65	. . . . . TC	IN (B), AS
66	. . . . . C	IN (R,Y), EUR
67	. . . . C	IN (B,K), AS, EUR
68	. . . . . T	IN (Y), AS, EUR
69	. . C . . . . C	IN (Y), AS, EUR, AF

a. Position refers to Cambridge reference sequence; note position 263 has been corrected.
b. Identical to the Cambridge reference sequence.
c. EUR, European; AF, African; AS, Asian; IN, Indian; B, Brahmin; K, Kapu, Y, Yadava; R, Relli.

mon and pygmy chimpanzees (Kocher and Wilson 1991). Therefore a substitution at bp 263 was scored only if it was G to A, not A to G.

Within the 200-bp region the distribution of polymorphisms is bimodal in the Indians and continental populations. Figure 1 displays the relationship between the absolute sequence position (x axis), the number of variable sites within each 10-bp region (y axis), and the total number of differences from the consensus sequence in the 10-bp region summed across individuals (z axis) for the total sample. Figures 2–5 illustrate the differences in regional nucleotide variability between populations. Africans (Figure 2) exhibit the

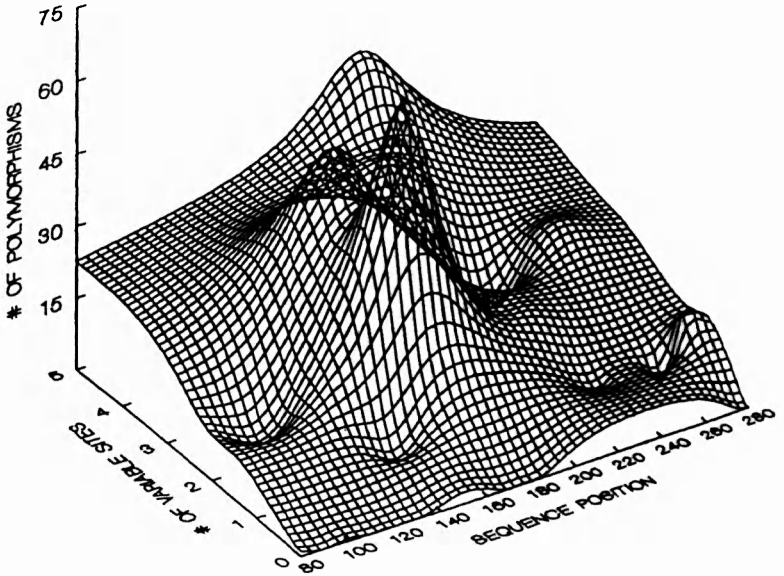

Figure 1. Three-dimensional plot of sequence position (x axis), the number of variable sites within each 10-bp region (y axis), and the total number of differences from the consensus sequence in the 10-bp region summed across individuals in all populations (z axis).

most variable regions and the highest frequency of polymorphisms within a region. Indians (Figure 3) appear intermediate between Africans and Asians (Figure 4). Europeans (Figure 5) reveal a similar number of variable regions, but fewer individuals are polymorphic. Obviously, there is a positive correlation between the number of variable sites in a region and the total number of polymorphisms in that region. Compared with a histogram depicting sequence position versus number of polymorphisms, this additional information is useful. The three-dimensional plots illustrate how genetic variation within a population is distributed across a DNA sequence.

The number of distinct lineages and the gene diversity (h) within each continental population are listed in Table 2. The number of mtDNA lineages is highest in the African population (30) and lowest in the Europeans (13). Indian and Asian populations contain 25 and 17 lineages, respectively. Thirty-seven individuals (26%) were identical to the CRS. Only one haplotype was shared between all continental populations, and it was just one mutation-step away from the CRS. Six additional haplotypes were shared between continental populations. Sixty-one haplotypes (88%) are specific to a continental population. Within Africans only two haplotypes are shared between subpopulations. Subpopulations of Asians share five haplotypes. Aside from the

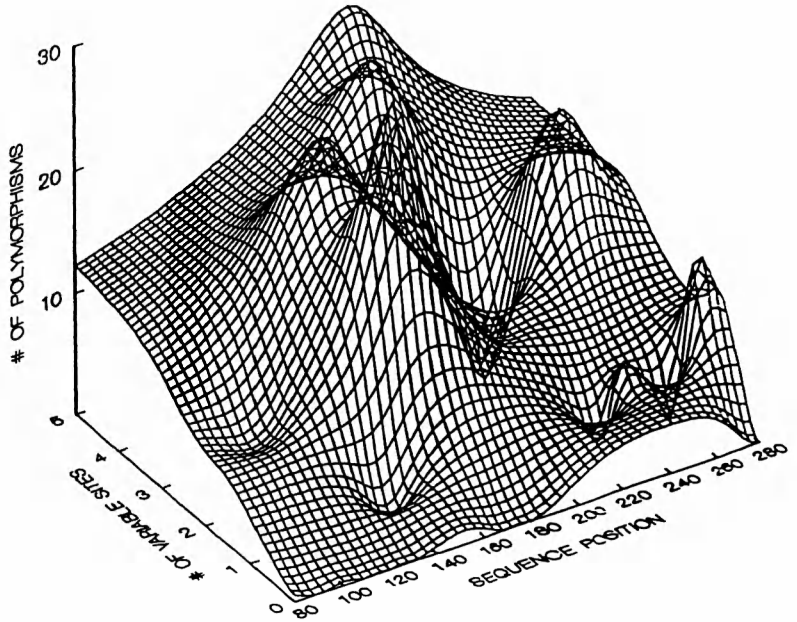

Figure 2. Three-dimensional plot of sequence position (x axis), the number of variable sites within each 10-bp region (y axis), and the total number of differences from the consensus sequence in the 10-bp region summed across all Africans (z axis).

CRS only two Indian haplotypes are shared between castes (Table 2). Furthermore, these haplotypes are found only between adjacent castes in the Hindu caste hierarchy (e.g., Relli and Yadava or Kapu and Brahmin) and never between the highest and lowest caste groups.

Gene diversity (h) within continental populations ranges from a minimum of 0.74 in Europeans to a maximum of 0.99 in Africans. The diversity of Indians is second highest at 0.98 (Table 2). Of the Indian caste populations (Table 3) the highest level of diversity is found in the Yadava (0.98) and the lowest level is found in the Kapu (0.90). Estimates for the Brahmin (0.92) and the Relli (0.93) are intermediate. The average nucleotide diversity within each continental population is listed in Table 4. The average nucleotide diversity is significantly higher in Africans (0.027) and Indians (0.014) compared with Europeans and Asians. Nucleotide diversity does not differ significantly between Europeans and Asians. The Brahmin caste shows the highest average nucleotide diversity, yet no caste group demonstrates significantly greater nucleotide variation than another caste (Table 5).

Genetic distances between each pair of continental populations are given in Table 6. Africans are consistently the most distant from any other

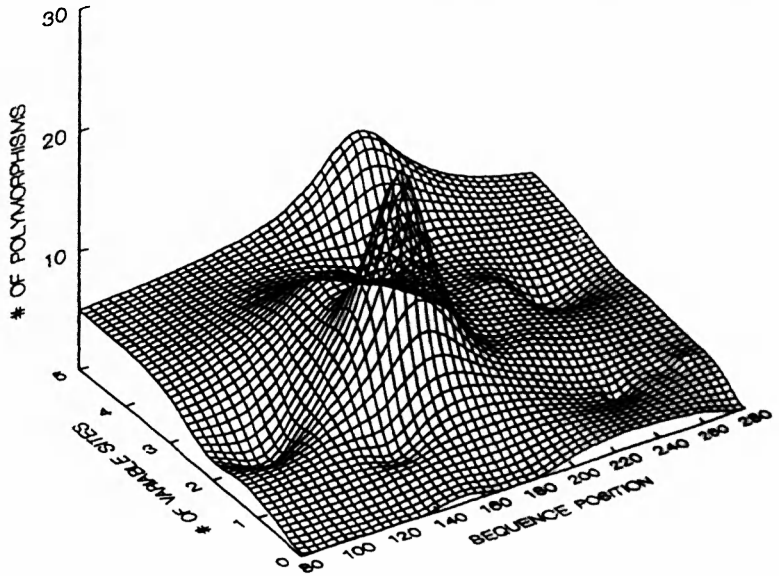

Figure 3. Three-dimensional plot of sequence position (*x* axis), the number of variable sites within each 10-bp region (*y* axis), and the total number of differences from the consensus sequence in the 10-bp region summed across all Indians (*z* axis).

continental population. Of the three non-African populations, the Indians are closest to the Africans. The Indians are nearly equidistant from Europeans and Asians. The smallest genetic distance is between Asians and Europeans. Table 7 lists the genetic distance between each pair of caste populations. The Brahmin and the Yadava appear to be farther apart than any other two pairs of castes. The smallest genetic distance is between the Kapu and the Relli. All these genetic distances are similar to each other.

The G_{ST} value for the three continental populations and the Indian caste groups is 0.15 ($p < 0.001$). This means that 15% of the total variance of this mtDNA region occurs between these four populations and that there is significant geographic structuring among continental populations with respect to mtDNA. This is consistent with previous studies using mtDNA (Jorde et al. 1995), nuclear DNA (Wright 1978; Jorde 1980; Nei and Roychoudhury 1982; Bowcock et al. 1987, 1991; Nei and Livshits 1990; Jorde et al. 1995; Nei et al. 1993), and craniometric data (Relethford and Harpending 1994). The higher estimate of G_{ST} from mtDNA compared with nuclear DNA and morphometric data is partly due to differences in mtDNA and nuclear DNA effective population sizes and may also reflect differences in male and female dispersal patterns, sampling of populations that are not in migration-drift equilibrium (Rogers and Jorde 1995), or selection.

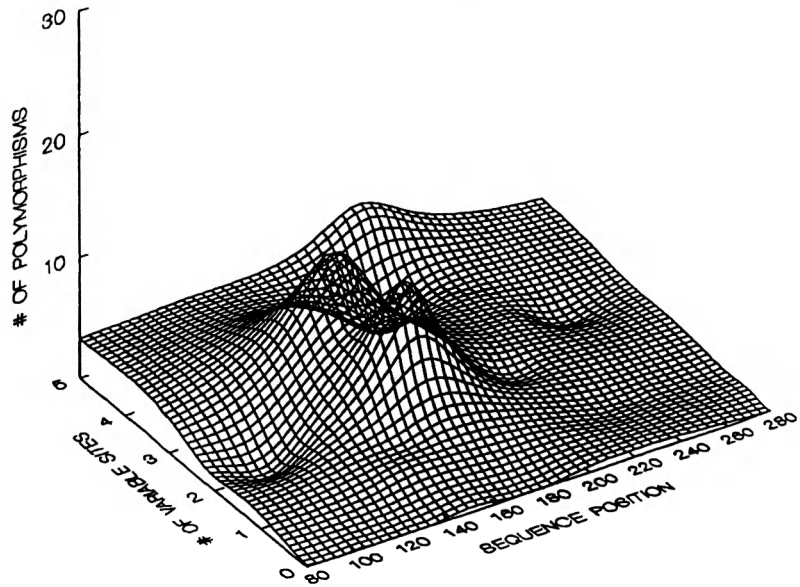

Figure 4. Three-dimensional plot of sequence position (x axis), the number of variable sites within each 10-bp region (y axis), and the total number of differences from the consensus sequence in the 10-bp region summed across all Asians (z axis).

The G_{ST} value for the caste populations is 0.17 ($p < 0.002$). This means that 17% of the total variance of this mtDNA region occurs between caste groups, indicating that there is significant structuring of mtDNA variation based on social organization. This value is much higher than previous estimates of genetic differentiation based on protein polymorphisms among tribal groups ($G_{ST} = 0.02$) (Murthy et al. 1993; Sirajuddin et al. 1994) and between caste populations ($G_{ST} = 0.013$) (Char et al. 1989) of Andhra Pradesh.

Figure 6 depicts an unrooted neighbor-joining tree and illustrates the relationships between the three continental populations and the Indians. The African-Indian cluster is supported by 88% of 1000 bootstrapped trees. The African branch length is much longer than any other branch, reflecting significantly higher diversity.

The unrooted neighbor-joining network linking caste populations is depicted in Figure 7. It should be emphasized that these trees are not phylogenies indicating ancestor-descendant relationships between populations; they are heuristic tools that display similarity between populations. This similarity may be due to admixture or common ancestry. The Yadava-Kapu cluster is supported by 66% of 1000 bootstrapped trees.

Figure 8 is a majority-rule consensus tree of unique Indian haplotypes. Only two branches clustering different caste populations are supported by

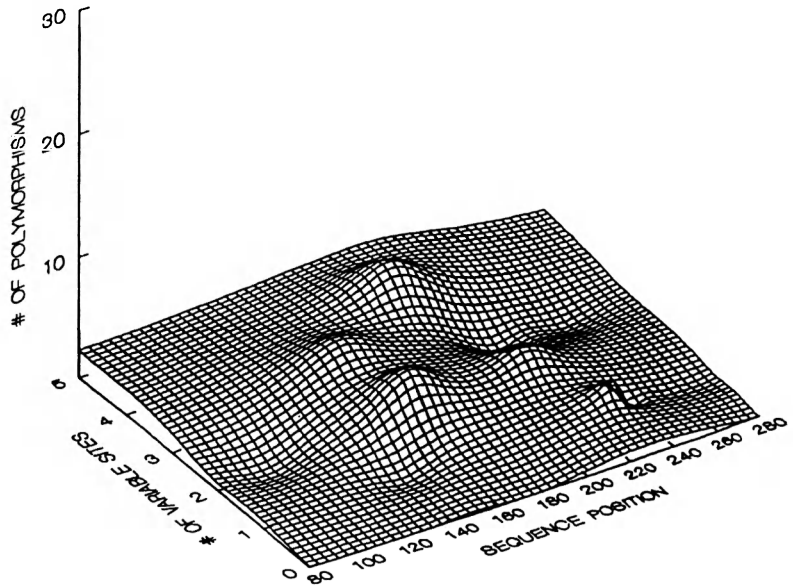

Figure 5. Three-dimensional plot of sequence position (x axis), the number of variable sites within each 10-bp region (y axis), and the total number of differences from the consensus sequence in the 10-bp region summed across all Europeans (z axis).

more than 50% of bootstrapped trees. The three haplotypes on these two branches are separated by two mutation-steps.

Discussion

We have analyzed a 200-bp segment of HVS2 in the mtDNA control region bracketed by bases 80–280 of the CRS to examine the population

Table 2. Variation in 200 Nucleotides in HVS2 of the mtDNA Control Region in Human Populations

Population	*Individuals*	*Lineages*	*Diversity*
African	36	30	0.99
East Indian	36	25	0.98
Asian	36	17	0.84
European	36	13	0.74
Total	144	69	0.92

Diversity is $h = (1 - \Sigma x_i^2)n/(n - 1)$, where x_i is the frequency of the type and n is the number of individuals (Nei 1987).

Table 3. Variation in 200 Nucleotides in HVS2 of the mtDNA Control Region in Indian Caste Populations

Population	*Individuals*	*Lineages*	*Diversity*
Brahmin	9	7	0.92
Kapu	7	5	0.90
Yadava	10	9	0.98
Relli	10	8	0.93

Diversity is $h = (1 - \Sigma x_i^2)n/(n - 1)$, where x_i is the frequency of the type and n is the number of individuals (Nei 1987).

Table 4. Mean Nucleotide Diversity (±SE) of 200 bp in HVS2 of the mtDNA Control Region of Human Continental Populations

Group	*Nucleotide Diversity*
African	0.027 (0.0019)
Indian	0.014 (0.0012)
Asian	0.009 (0.0011)
European	0.007 (0.0014)

Table 5. Mean Nucleotide Diversity (±SE) of 200 bp in HVS2 of the mtDNA Control Region of East Indian Caste Populations

Group	*Nucleotide Diversity*
Brahmin	0.015 (0.0028)
Kapu	0.011 (0.0020)
Yadava	0.012 (0.0019)
Relli	0.012 (0.0020)

structure of Indian caste populations of Andhra Pradesh. HVS2 appears to be highly informative, containing 1 polymorphic site per 6 bp on average. Although HVS2 contains elements critical to mitochondrial transcription and translation, only conserved sequence block 1 (CSB1) is found in this 200-bp segment. Three transition mutations were identified in CSB1 from 30 poly-

Table 6. Genetic Distances between Continental Populations

Group	*African*	*Indian*	*Asian*
African			
Indian	0.00440		
Asian	0.00584	0.00085	
European	0.00658	0.00115	0.00046

Table 7. Genetic Distances between East Indian Caste Populations

Group	*Brahmin*	*Kapu*	*Yadava*
Brahmin			
Kapu	0.00315		
Yadava	0.00497	0.00320	
Relli	0.00216	0.00243	0.00288

morphic sites. Although this region was able to discriminate many unique lineages in Africans and Indians, 50% of the Europeans and 40% of the Asians were identical to the CRS.

The distribution of genetic variation in each continental population across 200 bp in HVS2 is summarized by Figures 1–5. Although different populations may have similar levels of genetic diversity, this diversity is an aggregate of polymorphisms at various nucleotide sites. Rather than presenting just the frequency of variable sites in a genomic region without reference to the total number of polymorphisms in a population, these plots display all the descriptive information about the distribution of genetic variation. That is, a plot illustrates how many individuals in a population are polymorphic at each variable site. Furthermore, these illustrations demonstrate those differences effectively, enhancing the recognition of patterns that vary among pop-

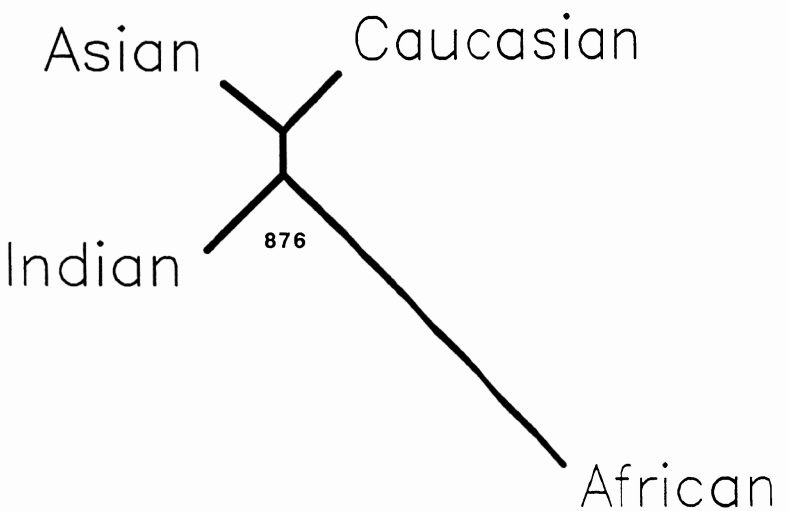

Figure 6. Neighbor-joining network linking continental populations. Note that the Indian-African cluster is supported by almost 90% of 1000 bootstrapped trees.

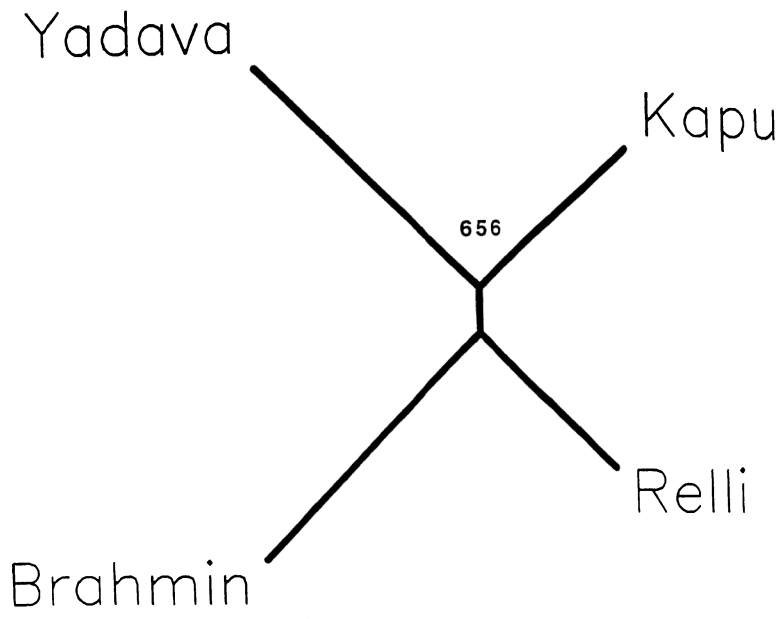

Figure 7. Neighbor-joining network linking East Indian caste populations. Note that the Yadava-Kapu cluster is highly supported.

ulations. These plots are analogous to topographic maps, containing plateaus and peaks of nucleotide variation that vary in size, height, and inclination. In fact, each continental population can be distinguished pictorially. Analyzing more sequence data or individuals could increase the resolution of a map to a point where more closely related populations could be distinguished.

As expected, the African plot has the most peaks (most variable sites) as well as the highest peaks (highest total number of polymorphisms at each variable site), whereas the European plot is nearly flat. The African plot also contains a unique peak between bp 230 and bp 250. Three of the four variable sites between bp 230 and bp 250 are African specific; one is shared by an Indian. The Indian plot combines features of the African and Asian plots. For example, the peak between bp 140 and bp 160 is found in Africans, Asians, and Indians, whereas the depression between bp 240 and bp 260 is seen only in the Asians and Indians. Apparently, any European contribution to the Indian plot is hidden by African and Asian contributions.

It has been argued that the amount of genetic diversity in Indian populations is comparable to that existing in the major "races of man" (Majumder and Mukherjee 1993). The analyses from which this conclusion is drawn have been limited to the examination of a small number of red blood cell enzymes, serum protein polymorphisms, and blood group polymorphisms. Because so

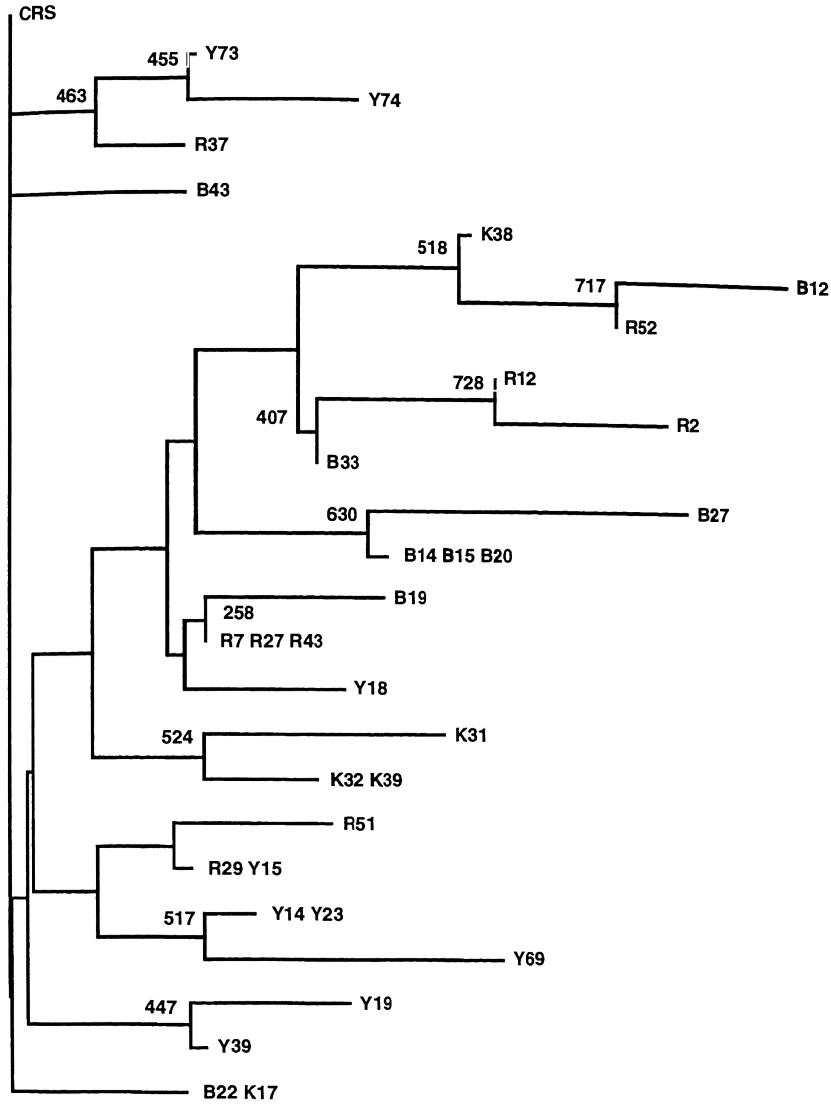

Figure 8. Neighbor-joining tree of unique East Indian haplotypes. A branch linking caste groups not immediately above or below in the caste hierarchy is supported only once.

few loci have been studied, inferences made from these data must be considered preliminary and should be viewed cautiously because gene frequency changes are subject to large stochastic errors (Nei 1987). Likewise, the nuclear DNA markers that have been studied in a few Indian populations are located

in genomic regions near genes under relatively intense selection (e.g., β-globin). No analysis of Indian populations using a large number of highly varied and selectively neutral markers distributed throughout the genome has been performed.

Our analysis examines genetic variation in mtDNA in caste populations of South India. We suggest that mtDNA diversity in Indian caste populations is intermediate between African and other continental populations. Although the sample sizes are relatively small, this result is substantiated by the facts that (1) as in other investigations, African populations have the greatest mtDNA diversity and Europeans the least and (2) the amount of differentiation among continental populations, as estimated by G_{ST}, is similar to estimates made by other investigators using nuclear DNA data sets [e.g., Jorde et al. (1995)]. The G_{ST} estimates of Merriwether et al. (1991) and Stoneking et al. (1990) are higher than our estimated G_{ST} but were calculated according to the method of Takahata and Palumbi (1985). This method recently has been demonstrated to inflate estimates of G_{ST} (Harpending et al. 1995). In addition, this may reflect in part the fact that our estimate is based on the more rapidly mutating control region, whereas previous estimates are based on noncontrol region DNA. Our findings are consistent with the conclusions of other studies of continental populations using mtDNA (Cann et al. 1987; Vigilant et al. 1989, 1991; Horai et al. 1993; Jorde et al. 1995), nuclear DNA polymorphisms (Bowcock et al. 1991, 1994; Jorde et al. 1995), and craniometric data (Relethford and Harpending 1994).

It is not clear why there is more mtDNA variation in the Indian castes than in European and Asian populations. It is known that within-group variation is partly a function of long-term effective population size (Stoneking 1993). In fact, significantly greater African mtDNA variation has been postulated to result from a larger effective population in Pleistocene Africa (Relethford and Harpending 1994). Unfortunately, little information exists on long-term effective population sizes of Indian castes. Population expansions increase genetic variation because the number of new mutations retained is proportionately higher than in a population at equilibrium. This feature generates waves in the distribution of pairwise sequence differences within a population and can be used to estimate ancient demographic parameters (Rogers and Harpending 1992; Rogers 1994; Rogers and Jorde 1995). Many human populations have undergone recent rapid expansion (Weiss 1989), and indeed waves are found in the mismatch distributions of the continental populations and the Indians (data not shown). But there is no evidence that Indian castes have expanded earlier or more rapidly than European or Asian populations. Last, unequal admixture of caste populations with different continental populations could result in population heterogeneity and an intermediate level of diversity. This may be the most reasonable explanation.

The patterns of genetic variation among Indian populations are complex. Social, linguistic, and geographic barriers to gene flow, leading to small

effective populations that differentiate by genetic drift, can explain differences between some tribal groups, caste populations, and geographically separated caste sects (Murthy et al. 1993; Roychoudhury 1977). Numerous studies have indicated gene flow and subsequent admixture of caste and tribal groups with European populations of West Asia and Asian populations of Central and East Asia (Balakrishnan 1982; Roychoudhury 1977; Walter et al. 1981; Walter 1986) and admixture between tribal and nontribal groups (Chakraborty et al. 1986). Little evidence of admixture with African and/or Australian aboriginal populations has been found (Roychoudury 1984). Some general features of Indian population structure are (1) firm evidence of admixture with European and Asian populations, (2) substantial geographic structuring of some tribal and caste populations (Balakrishnan 1982; Roychoudhury 1982, 1984; Sirajuddin et al. 1994;), (3) gene flow and admixture between populations in some geographic regions predicted from ethnohistories (Chakraborty et al. 1986), and (4) minimally structured genetic variation among castes groups according to social rank (Balakrishnan 1978; Papiha et al. 1982; Mukherjee et al. 1979).

The relationship of specific tribal and caste populations of South India to Europeans, Asians, and Africans remains controversial (Roychoudhury 1977). African admixture with East Indians had two potential sources: (1) aboriginal populations that were of African origin and that migrated into South India from the north or by sea (Chandler 1988) and (2) forced immigration of Africans into South India by the Portuguese in the sixteenth century (Watson 1979). Morphological features, such as "frizzly hair" and "a broad flat nose, fleshy everted lips and skin colour" (Roychoudhury 1982, p. 148), found in some Indians, encouraged exploration for a connection with African and/or Australian aboriginal populations (Guha 1944; Kirk et al. 1962; Sarkar 1954; Roychoudhury 1977, 1984). In fact, it has been suggested that at one time a "Negrito element" was widespread throughout India and was eventually forced into a more restricted location in South India (Majumder and Mukherjee 1993). Yet no genetic evidence of African admixture has been identified, leading to further speculation of "only sporadic, numerically small waves of African immigration to restricted geographical regions of India" (Majumder and Mukherjee 1993, p. 259). The genetic distance analysis of continental populations (Table 6) shows that African populations are nearly equidistant from all other continental groups, although they are slightly closer to Indians. Bootstrapping indicates that an African-Indian branch is supported by the neighbor-joining network (Figure 6) in nearly 88% of 1000 trees. This is limited but provocative genetic evidence for African admixture in caste populations of Andhra Pradesh. No specific caste population reliably clusters with any continental population (data not shown). Our first hypothesis, proposing that the Brahmin would exhibit a closer affinity to Europeans and the Relli to African and Asian populations, cannot be supported.

Why is this pattern of African affinity found using mtDNA markers and not with nuclear DNA markers? One possible explanation is that not enough neutral, highly polymorphic nuclear DNA markers have been examined in caste populations from South India to enable detection of a similar pattern. The discrepancy also may reflect real differences between mitochondrial and nuclear DNA patterns. Jorde et al. (1995) studied 30 restriction site polymorphisms and 30 microsatellite markers distributed throughout the genome and a 200-bp region of HVS2 in 241 individuals from Asia, Europe, and Africa. A Mantel comparison of the patterns of genetic variation summarized in nuclear and mitochondrial DNA distance matrices demonstrated significant concordance between the two sets of nuclear polymorphisms but discordance between the nuclear DNA and mtDNA polymorphisms. It is possible that African males and females have had different patterns or rates of gene flow into South India. Thus the African mtDNA affinity could reflect predominantly female migration from Africa to South India. Further evidence of sex-limited admixture could be found by examining Y-chromosome markers from these caste populations. Selection at the mtDNA locus for a subset of polymorphisms common to Africans and East Indians would also explain the discordance between the pattern generated by mtDNA versus nuclear DNA markers. Thus far, no substantial evidence exists for selection acting preferentially on African and East Indian haplotypes.

Although nucleotide diversity is highest in Brahmins, there is no significant difference in the amount of genetic variation within each caste group. Our second hypothesis, predicting higher levels of genetic diversity in the middle castes compared to the upper and lower castes, is not supported.

Our third hypothesis, predicting that there would be substantial differences in genetic variation between castes, is supported. The estimate of between-population differentiation is slightly higher ($G_{ST} = 0.17$) than that found between continental populations ($G_{ST} = 0.15$). This value is much higher than previously reported estimates based on nuclear DNA markers. This estimate of G_{ST} changes little if any single caste population is removed from the analyses. Therefore no single caste contributes to mtDNA differentiation disproportionately. mtDNA variation is significantly structured by social stratification in the Hindu caste system.

As stated before, mtDNA polymorphisms are more sensitive than some nuclear DNA markers for detecting population differentiation. This is due to the haploid inheritance, high mutation rate, and smaller effective population size of mtDNA molecules. It is therefore reasonable to expect that estimates of G_{ST} would be higher than those calculated from nuclear DNA markers. But the structuring of mtDNA variation among castes is greater than anticipated. Given that social mobility is thought to be higher for caste females than for caste males, the pattern of mtDNA variation should be less structured than estimates derived from nuclear DNA markers.

There are numerous potential explanations for this paradox. Many of the caste populations living in urban centers have moved from neighboring rural areas. Thus our sampling strategy may have amplified ancestral patterns of geographic structuring by recruiting individuals whose ancestors were widely dispersed. This is similar to the agglomeration effect described by Chakraborty et al. (1988), the signature of which may persist for several thousand years (Neel et al. 1988). This could be tested by sampling the same caste populations in rural areas. Another demographic feature that could increase the differentiation of mtDNA variation among castes is the dispersal of females from different geographic areas to a patrilocal residence. An analysis of Y-chromosome variation could further define the role of this force. That is, Y-chromosome variation in male caste members, who would tend to remain in one location through many generations, should be less structured by caste affiliation or more geographically structured. Last, the social mobility of women may be more restricted than estimated from ethnographic data.

Our fourth hypothesis, suggesting that the middle castes would share more haplotypes than the upper and lower castes, is not supported. Of the 31 Indian mtDNA haplotypes that differ from the CRS, none are shared between the Kapu and the Yadava. Two haplotypes are shared between castes. One haplotype is shared by a Relli and a Yadava and another by a Brahmin and a Kapu. Thus haplotypes are shared only with a caste immediately above or below in the social hierarchy. This suggests that even at the resolution of this short sequence of mtDNA, genetic variation is structured between caste populations. This is reinforced by the neighbor-joining network of caste populations (Figure 7), which supports the clustering of the middle castes, the Kapu and the Yadava, in 66% of bootstrapped trees. Yet the Kapu and the Yadava share no haplotypes. Perhaps sharing an mtDNA type is too crude a measure, and nucleotide differences offer better resolution. Furthermore, in the bootstrapped haplotype tree (Figure 8) only two branch points where terminal branch castes differ and a single branch point where terminal branch castes are not contiguous in the caste hierarchy are supported by more than 50% of trees. Interestingly, this branch joins a Brahmin to a Relli, and support for intermarriage between people of high and low castes has been found previously (Chakraborty et al. 1986).

Conclusions

Genetic diversity of mtDNA polymorphisms is significantly higher in Indian caste populations than in Asians or Europeans but is lower than that found in Africans. mtDNA variation between continental populations is structured geographically, with G_{ST} estimates similar to those from nuclear DNA markers. In a neighbor-joining network the caste populations of Andhra Pradesh cluster more often with Africans than with Asians or Europeans. This

is suggestive of admixture with African populations. Genetic diversity does not vary significantly between caste populations, although the Brahmins exhibit slightly more variation than other castes. mtDNA variation among Indian populations of Andhra Pradesh is highly structured between castes. This is supported by the lack of haplotype sharing between castes and the clustering of caste haplotypes in a majority-rule neighbor-joining tree. Caste haplotypes are not shared more often between the middle castes (i.e., Kapu and Yadava) than between the upper (Brahmin) and lower (Relli) castes. The middle castes do cluster together more often when linked by the neighbor-joining algorithm. Explanations for the discordance between patterns of variation as measured by mtDNA versus nuclear DNA include (1) the higher resolving power of mtDNA, (2) sex-dependent gene flow, (3) differences in male and female effective population sizes, and (4) elements of the kinship structure. Future analyses should use a variety of mtDNA and nuclear DNA markers and examine more caste populations. This is a small but significant step toward unraveling the complex relationships between the populations of South India.

Acknowledgments We would like to thank T. Jenkins, H. Soodyall, P. Nute, and J. Kidd for providing DNA samples and S. Austin, A. Comuzzie, R. Duggirala, R. Feldman, K. Lum, A. Rogers, and S. Watkins for technical advice, critical comments, and thoughtful discussion. This work was supported in part by the National Science Foundation through grant NSF-DBS-9211255, the Clinical Research Center at the University of Utah through grant NIH RR-00064, and the Technology Access Center of the Utah Human Genome Project.

Received 22 December 1994; revision received 9 May 1995.

Literature Cited

Anderson, S., A.T. Bankier, B.G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.

Arnold, C., and I.J. Hodgson. 1991. Vectorette PCR: A novel approach to genomic walking. *PCR Meth. Appl.* 1(1):39–42.

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

Balakrishnan, V. 1982. Admixture as an evolutionary force in populations of the Indian subcontinent. In *Human Genetics and Adaptation,* K.C. Malhotra and A. Basu, eds. New York: Plenum, v. 1, 103–145.

Ballinger, S.W., T.G. Schurr, A. Torroni et al. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 130(1):139–152.

Bell, G.I., J.H. Karem, and J.R. Rutter. 1981. Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. *Proc. Natl. Acad. Sci. USA* 78:5759–5763.

Birky, C.W., P. Fuerst, and T. Maruyama. 1989. Organelle gene diversity under migration and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121:613–627.

Birky, C.W., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103:513–527.

Bowcock, A.M., C. Bucci, J.M. Kidd et al. 1987. A study of 47 DNA markers in five populations from four continents. *Gene Geogr.* 1:47–64.

Bowcock, A.M., J.R. Kidd, J.L. Mountain et al. 1991. Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* 88:839–843.

Bowcock, A.M., A. Ruiz-Linares, J. Tomfohrde et al. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457.

Brown, W.M., M. George, and A.C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 77:3605–3609.

Brown, W.M., E. Prager, A. Wang et al. 1982. Mitochondrial DNA sequences in primates: Tempo and mode of evolution. *J. Molec. Evol.* 18:225–239.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press.

Cavalli-Sforza, L.L., A. Piazza, P. Menozzi et al. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* 85:6002–6006.

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

Chakraborty, R.H., A. Chakravarti, and K.C. Malhotra. 1977. Variation in allele frequencies among caste groups of Dhangars of Maharashtra, India: An analysis with Wright's F_{ST} statistic. *Ann. Hum. Biol.* 4:275–280.

Chakraborty, R., P.E. Smouse, and J.V. Neel. 1988. Population amalgamation and genetic variation: Observations on artificially agglomerated tribal populations of Central and South America. *Am. J. Hum. Genet.* 43:709–725.

Chakraborty, R., H. Walter, B.N. Mukherjee et al. 1986. Gene differentiation among ten endogamous groups of West Bengal, India. *Am. J. Phys. Anthropol.* 71:295–309.

Chandler, W.B. 1988. The jewel in the lotus: The Ethiopian presence in the Indus valley civilization. In *African Presence in Early Asia,* I. Van Sertima and R. Rashidi, eds. New Brunswick: Transaction, 80–105.

Char, K.S.N., and P.R. Rao. 1986. Glyoxylase I phenotypes in some endogamous populations of Andhra Pradesh, India. *Hum. Hered.* 36(2):123–125.

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.

Crow, J.F., and K. Aoki. 1984. Group selection for a polygenic behavioral trait: Estimating the degree of population subdivision. *Proc. Natl. Acad. Sci. USA* 81:6073–6077.

Di Rienzo, A., and A.C. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1597–1601.

Excoffier, L. 1990. Evolution of human mitochondrial DNA: Evidence of departure from a pure neutral model of populations at equilibrium. *J. Molec. Evol.* 30:125–139.

Felsenstein, J. 1993. *PHYLIP, version 3.5c.* Seattle, WA: University of Washington.

Giles, R.E., H. Blanc, H.M. Cann et al. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 77:6715–6719.

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

Harpending, H.C., J. Relethford, and S.T. Sherry. 1995. Methods and models for understanding human diversity. In *Molecular Biology and Human Diversity,* A.J. Boyce and C.G.N. Mascie-Taylor, eds. London, England: Cambridge University Press (in press).

Harpending, H.C., S.T. Sherry, A.R. Rogers et al. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* 34:483–496.

Horai, S., and K. Hayasaka. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* 46:828–842.

Horai, S., R. Kondo, Y. Nakagawa-Hattori et al. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Molec. Biol. Evol.* 10:23–47.

Johnson, M.J., D.C. Wallace, S.D. Ferris et al. 1983. Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J. Molec. Biol. Evol.* 19:255–271.

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.

Jorde, L.B., M. Bamshad, S.W. Watkins et al. 1995. Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* 57:523–538.

Karve, I. 1968. Kinship organization. In *Kinship Organization in India.* Bombay, India: Asia Publishing House.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Molec. Evol.* 16:111–120.

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.

Kocher, T.D., and A.C. Wilson. 1991. Sequence evolution of mitochondrial DNA in humans and chimpanzees: Control region and a protein coding region. In *Evolution of Life: Fossils, Molecules, and Culture*, S. Osawa and T. Honjo, eds. Tokyo, Japan: Springer-Verlag, 391–413.

Kocher, T.D., W.K. Thomas, A. Meyer et al. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86:6196–6200.

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

Li, W.-H., and L.A. Sadler. 1991. Low nucleotide diversity in man. *Genetics* 129:513–523.

Lum, J.K., O. Richards, C. Ching et al. 1994. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum. Biol.* 66(4):567–590.

Majumdar, D.N. 1958. *Races and Cultures of India.* Bombay, India: Asia Publishing House.

Majumder, P.P., and B.N. Mukherjee. 1993. Genetic diversity and affinities among Indian populations: An overview. In *Human Population Genetics*, P.P. Majumder, ed. New York: Plenum, 255–275.

Majumder, P.P., B.U. Shankar, A. Basu et al. 1990. Anthropometric variation in India: A statistical appraisal. *Curr. Anthropol.* 31(1):94–103.

Malhotra, K.C. 1978. Morphological composition of the people of India. *J. Hum. Evol.* 7:45–53.

Malhotra, K.C., and T.S. Vasulu. 1993. Structure of human populations in India. In *Human Population Genetics*, P.P. Majumder, ed. New York: Plenum, 207–233.

Malhotra, K.C., R. Chakraborthy, B.V. Bhanu et al. 1980. Variation on dermal ridges in nine population groups of Maharasthra, India: Intra- and interpopulation diversity. *Hum. Hered.* 30:307–315.

Maloney, C. 1974. *The Races in Peoples of South Asia.* New York: Holt, Rinehart & Winston.

Merriwether, D.A., A.G. Clark, S.W. Ballinger et al. 1991. The structure of human mitochondrial DNA variation. *J. Molec. Evol.* 33:543–555.

Mountain, J.L., J.M. Hebert, S. Bhattacharyya et al. 1995. Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* 56:979–992.

Mukherjee, B.N., P.P. Majumder, K.C. Malhotra et al. 1979. Genetic distance analysis among nine endogamous population groups of Maharashtra, India. *J. Hum. Evol.* 8:567–570.

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.

Neel, J.V., C. Satoh, P. Smouse et al. 1988. Protein variants in Hiroshima and Nagasaki: Tales of two cities. *Am. J. Hum. Genet.* 43:870–893.

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

Nei, M., and G. Livshits. 1990. Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. In *Population Biology of Genes and Molecules,* N. Takahata and J.F. Crow, eds. Tokyo, Japan: Baifukan, 251–265.

Nei, M., and A.K. Roychoudhury. 1982. Genetic relationship and evolution of human races. *Evol. Biol.* 14:1–59.

Nei, M., and A.K. Roychoudhury. 1993. Evolutionary relationships of human populations on a global scale. *Molec. Biol. Evol.* 10:927–943.

Nei, M., G. Livshits, and T. Ota. 1993. Genetic variation and evolution of human populations. In *Genetics of Cellular, Individual, Family, and Population Variability,* C.F. Sing and C.L. Hanis, eds. New York: Oxford University Press, 239–252.

Olivo, P.D., M.J. Van De Walle, P.J. Laipis et al. 1983. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* 306:400–402.

Papiha, S.S., B.N. Mukherjee, S.M.S. Chahal et al. 1982 Genetic heterogeneity and population structure in northwest India. *Ann. Hum. Biol.* 9:235–251.

Relethford, J.H., and H.C. Harpending. 1994. Craniometric variation, genetic theory, and modern human origins. *Am. J. Phys. Anthropol.* 95:249–270.

Rogers, A.R. 1994. Genetic evidence for a Pleistocene population explosion. *Evolution* 49:608–615.

Rogers, A.R., and H.C. Harpending. 1986. Migration and drift in human populations. *Evolution* 40:1312–1327.

Rogers, A.R., and H.C. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molec. Biol. Evol.* 9:552–569.

Rogers, A.R., and L.B. Jorde. 1995. Genetic evidence on modern human origins. *Hum. Biol.* 67:1–36.

Roychoudhury, A.K. 1974. Gene differentiation among caste and linguistic populations of India. *Hum. Hered.* 24:317–322.

Roychoudhury, A.K. 1977. Gene diversity in Indian populations. *Hum. Genet.* 40:99–103.

Roychoudhury, A.K. 1982. Genetic relationships of Indian populations. In *Human Genetics and Adaptation,* K.C. Malhotra and A. Basu, eds. New York: Plenum Press, v. 1, 147–174.

Roychoudhury, A.K. 1984. 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.

Saccone, C., G. Pesole, and E. Sbisa. 1991. The main regulatory region of mammalian mitochondrial DNA: Structure-function model of an evolutionary pattern. *J. Molec. Evol.* 33:83–91.

Saiki, R.K., S. Scharf, F. Faloona et al. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.

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

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

Schurr, T.G., S.W. Ballinger, Y.-Y. Gan et al. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary lineages. *Am. J. Hum. Genet.* 46:613–623.

Semino, O., A. Torroni, R. Scozzari et al. 1991. Mitochondrial DNA polymorphisms among Hindus: A comparison with the Tharus of Nepal. *Ann. Hum. Genet.* 55:123–136.

Shields, G.F., K. Hecker, M.I. Voevoda et al. 1992. Absence of the Asian-specific region V mitochondrial marker in native Beringians. *Am. J. Hum. Genet.* 50:758–765.

Singh, R.H. 1978. Dermatoglyphic variation in four castes of Uttar Pradesh, India. *Hum. Biol.* 50:251–260.

Sirajuddin, S.M., R. Duggirala, and M.H. Crawford. 1994. Population structure of the Chenchu and other South Indian tribal groups: Relationships between genetic, anthropometric, dermatoglyphic, geographic, and linguistic distances. *Hum. Biol.* 66(5):865–884.

Soodyall, H., and T. Jenkins. 1992. Mitochondrial DNA studies in the South African Indian population. *Gene Geogr.* 6:127–137.

Stoneking, M. 1993. DNA and recent human evolution. *Evol. Anthropol.* 2:60–73.

Stoneking, M., L.B. Jorde, K. Bhatia et al. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733.

Sykes, B.C. 1983. DNA in heritable disease. *Lancet* 2:787–788.

Takahata, N. 1991. Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics* 129:585–595.

Takahata, N., and S.R. Palumbi. 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics* 109:441–457.

Tambiah, S.J. 1973. From varna to caste through mixed unions. In *The Character of Kinship*, J. Goody, ed. Cambridge, England: Cambridge University Press, 191–229.

Thapar, R. 1980. India before and after the Mauryan empire. In *The Cambridge Encyclopedia of Archaeology*, A. Sherratt, ed. New York: Crown, 257–261.

Vigilant, L., R. Pennington, H. Harpending et al. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA* 86:9350–9354.

Vigilant, L., M. Stoneking, H. Harpending et al. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.

Walter, H. 1986. Genetic differentiation processes among the populations of India. *Int. J. Anthropol.* 1:297.

Walter, H., K.P. Pahl, M. Hilling et al. 1981. Genetic markers in eight endogamous population groups from Andhra Pradesh (South India). *Z. Morphol. Anthropol.* 72:325–338.

Ward, R.H., B.L. Frazier, K. Dew-Jager et al. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* 88:8720–8724.

Watson, F. 1979. *A Concise History of India.* New York: Thames and Hudson.

Weiss, K.M. 1989. On the number of members of the genus *Homo* who have ever lived and some evolutionary implications. *Hum. Biol.* 56:637–650.

Wilson, A.C., R.B. Cann, S.M. Carr et al. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26:375–400.

Wright, S. 1978. *Evolution and the Genetics of Populations*, v. 4. Chicago, IL: University of Chicago Press.