Genetic Evidence on the Origins of Indian Caste Populations

Michael Bamshad,[1,10,12] Toomas Kivisild,[2] W. Scott Watkins,[3] Mary E. Dixon,[3] Chris E. Ricker,[3] Baskara B. Rao,[4] J. Mastan Naidu,[4] B.V. Ravi Prasad,[4,5] P. Govinda Reddy,[6] Arani Rasanayagam,[7] Surinder S. Papiha,[8] Richard Villems,[2] Alan J. Redd,[7] Michael F. Hammer,[7] Son V. Nguyen,[9] Marion L. Carroll,[9] Mark A. Batzer,[9,11] and Lynn B. Jorde[3]

[1]Department of Pediatrics, University of Utah, Salt Lake City, Utah 84112, USA; [2]Institute of Molecular and Cell Biology, Tartu University and Estonian Biocentre, Tartu 51010, Estonia; [3]Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA; [4]Department of Anthropology, Andhra University, Visakhapatnam, Andhra Pradesh, India; [5]Anthropological Survey of India, Calcutta, India; [6]Department of Anthropology, University of Madras, Madras, Tamil Nadu, India; [7]Laboratory of Molecular Systematics and Evolution, University of Arizona, Tucson, Arizona 85721, USA; [8]Department of Human Genetics, University of Newcastle-upon-Tyne, UK; [9]Department of Pathology, Biometry and Genetics, Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Louisiana State University Health Science Center, New Orleans, Louisiana 70112, USA

The origins and affinities of the ~1 billion people living on the subcontinent of India have long been contested. This is owing, in part, to the many different waves of immigrants that have influenced the genetic structure of India. In the most recent of these waves, Indo-European-speaking people from West Eurasia entered India from the Northwest and diffused throughout the subcontinent. They purportedly admixed with or displaced indigenous Dravidic-speaking populations. Subsequently they may have established the Hindu caste system and placed themselves primarily in castes of higher rank. To explore the impact of West Eurasians on contemporary Indian caste populations, we compared mtDNA (400 bp of hypervariable region 1 and 14 restriction site polymorphisms) and Y-chromosome (20 biallelic polymorphisms and 5 short tandem repeats) variation in ~265 males from eight castes of different rank to ~750 Africans, Asians, Europeans, and other Indians. For maternally inherited mtDNA, each caste is most similar to Asians. However, 20%–30% of Indian mtDNA haplotypes belong to West Eurasian haplogroups, and the frequency of these haplotypes is proportional to caste rank, the highest frequency of West Eurasian haplotypes being found in the upper castes. In contrast, for paternally inherited Y-chromosome variation each caste is more similar to Europeans than to Asians. Moreover, the affinity to Europeans is proportionate to caste rank, the upper castes being most similar to Europeans, particularly East Europeans. These findings are consistent with greater West Eurasian male admixture with castes of higher rank. Nevertheless, the mitochondrial genome and the Y chromosome each represents only a single haploid locus and is more susceptible to large stochastic variation, bottlenecks, and selective sweeps. Thus, to increase the power of our analysis, we assayed 40 independent, biparentally inherited autosomal loci (1 LINE-1 and 39 *Alu* elements) in all of the caste and continental populations (~600 individuals). Analysis of these data demonstrated that the upper castes have a higher affinity to Europeans than to Asians, and the upper castes are significantly more similar to Europeans than are the lower castes. Collectively, all five datasets show a trend toward upper castes being more similar to Europeans, whereas lower castes are more similar to Asians. We conclude that Indian castes are most likely to be of proto-Asian origin with West Eurasian admixture resulting in rank-related and sex-specific differences in the genetic affinities of castes to Asians and Europeans.

Present addresses: [10]Eccles Institute of Human Genetics, 15 North 2030 East, Room 2100, University of Utah, Salt Lake City, UT 84112-5330, USA. [11]Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 508 Life Sciences Building, Baton Rouge, LA 70803, USA.
[12]Corresponding author.
E-MAIL mike@genetics.utah.edu; FAX (801) 585-9148.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.173301.

Shared Indo-European languages (i.e., Hindi and most European languages) suggested to linguists of the nineteenth and twentieth centuries that contemporary Hindu Indians are descendants of primarily West Eurasians who migrated from Europe, the Near East, Anatolia, and the Caucasus 3000–8000 years ago (Poliakov 1974; Renfrew 1989a,b). These nomadic migrants may

have consolidated their power by admixing with native Dravidic-speaking (e.g., Telugu) proto-Asian populations who controlled regional access to land, labor, and resources (Cavalli-Sforza et al. 1994), and subsequently established the Hindu caste hierarchy to legitimize and maintain this power (Poliakov 1974; Cavalli-Sforza et al. 1994). It is plausible that these West Eurasian immigrants also appointed themselves to predominantly castes of higher rank. However, archaeological evidence of the diffusion of material culture from Western Eurasia into India has been limited (Shaffer 1982). Therefore, information on the genetic relationships of Indians to Europeans and Asians could contribute substantially to understanding the origins of Indian populations.

Previous genetic studies of Indian castes have failed to achieve a consensus on Indian origins and affinities. Various results have supported closer affinity of Indian castes either with Europeans or with Asians, and several factors underlie this inconsistency. First, erratic or limited sampling of populations has limited inferences about the relationships between caste and continental populations (i.e., Africans, Asians, Europeans). These relationships are further confounded by the wide geographic dispersal of caste populations. Genetic affinities among caste populations are, in part, inversely correlated with the geographic distance between them (Malhotra and Vasulu 1993), and it is likely that affinities between caste and continental populations are also geographically dependent (e.g., different between North and South Indian caste populations). Second, it has been suggested that castes of different rank may have originated from or admixed with different continental groups (Majumder and Mukherjee 1993). Third, the size of caste populations varies widely, and the effects of genetic drift on some small, geographically isolated castes may have been substantial. Fourth, most of the polymorphisms assayed over the last 30 years are indirect measurements of genetic variation (e.g., ABO typing), have been sampled from only a few loci, and may not be selectively neutral. Finally, only rarely have systematic comparisons been made with continental populations using a large, uniform set of DNA polymorphisms (Majumder 1999).

To investigate the origin of contemporary castes, we compared the genetic affinities of caste populations of differing rank (i.e., upper, middle, and lower) to worldwide populations. We analyzed mtDNA (hypervariable region 1 [HVR1] sequence and 14 restriction-site polymorphisms [RSPs]), Y-chromosome (5 short-tandem repeats [STRs] and 20 biallelic polymorphisms), and autosomal (1 LINE-1 and 39 *Alu* inserts) variation in ~265 males from eight different Telugu-speaking caste populations from the state of Andhra Pradesh in South India (Bamshad et al. 1998). Comparisons were made to ~400 individuals from tribal and Hindi-speaking caste and populations distributed across the Indian subcontinent (Mountain et al. 1995; Kivisild et al. 1999) and to ~350 Africans, Asians, and Europeans (Jorde et al. 1995, 2000; Seielstad et al. 1999).

RESULTS

Analysis of mtDNA Suggests a Proto-Asian Origin of Indians

MtDNA HVR1 genetic distances between caste populations and Africans, Asians, and Europeans are significantly different from zero ($p < 0.001$) and reveal that, regardless of rank, each caste group is most closely related to Asians and is most dissimilar from Africans (Table 1). The genetic distances from major continental populations (e.g., Europeans) differ among the three caste groups, and the comparison reveals an intriguing pattern. As one moves from lower to upper castes, the distance from Asians becomes progressively larger. The distance between Europeans and lower castes is larger than the distance between Europeans and upper castes, but the distance between Europeans and middle castes is smaller than the upper caste-European distance. These trends are the same whether the Kshatriya and Vysya are included in the upper castes, the middle castes, or excluded from the analysis. This may be owing, in part, to the small sample size ($n = 10$) of each of these castes. Among the upper castes the genetic distance between Brahmins and Europeans (0.10) is smaller than that between either the Kshatriya and Europeans (0.12) or the Vysya and Europeans (0.16). Assuming that contemporary Europeans reflect West Eurasian affinities, these data indicate that the amount of West Eurasian admixture with Indian populations may have been proportionate to caste rank.

Conventional estimates of the standard errors of genetic distances assume that polymorphic sites are independent of each other, that is, unlinked. Because mtDNA polymorphisms are in complete linkage disequilibrium (as are polymorphisms on the nonrecom-

Table 1. MtDNA (HVR1 Sequence) Genetic Distances between Caste Groups from Andhra Pradesh and Continental Populations

Caste group	Africans	Asians	Europeans
Upper	.179	.037	.100 (0.106)[a]
Middle	.182	.025	.086 (0.084)[b]
Lower	.163	.023	.113
All castes	.196	.026	.077

[a]Genetic distance between upper castes and Europeans if the Kshatriya and Vysya are excluded from the analysis.
[b]Genetic distance between the middle castes and Europeans if the Kshatriya and Vysya are grouped in the middle castes.

bining portions of the Y chromosome), this assumption is violated. Alternatively, the mtDNA genome can be treated as a single locus with multiple haplotypes. However, even if this assumption is made, mtDNA distances do not differ significantly from one another even at the level of the three major continental populations (Nei and Livshits 1989), the standard errors being greater than the genetic distances. Considering that the distances between castes and continental populations are less than those between different continental populations, the estimated mtDNA genetic distances between upper castes and Europeans versus lower castes and Europeans would not be significantly different from each other. Therefore, to resolve further the relationships of Europeans and Asians to contemporary Indian populations, we defined the identities of specific mtDNA restriction-site haplotypes.

The presence of the mtDNA restriction sites *Dde*I$_{10,394}$ and *Alu*I$_{10,397}$ defines a haplogroup (a group of haplotypes that share some sequence variants), M, that was originally identified in populations that migrated from mainland Asia to Southeast Asia and Australia (Ballinger et al. 1992; Chen et al. 1995; Passarino et al. 1996) and is found at much lower frequency in European and African populations. Most of the common haplotypes found in Telugu- and Hindi-speaking caste populations belong to haplogroup M (Table 2) and do not differentiate into language-specific clusters in a phylogenetic reconstruction (Fig. 1). Furthermore, these Indian haplogroup-M haplotypes are distinct from those found in other Asian populations (Fig. 2) and indicate the existence of Indian-specific subsets of haplogroup M (e.g., M3). As expected if the lower castes are more similar to Asians than to Europeans, and the upper castes are more similar to Europeans than to Asians, the frequencies of M and M3 haplotypes are inversely proportional to caste rank (Table 2).

Of the non-Asian mtDNA haplotypes found in Indian populations, most are of West Eurasian origin (Table 2; Torroni et al. 1994; Richards et al. 1998). However, most of these Indian West-Eurasian haplotypes belong to an Indian-specific subset of haplogroup U, that is, U2i (Kivisild et al. 1999), the oldest and second most common mtDNA haplogroup found in Europe (Torroni et al. 1994). In agreement with the HVR1 results, the frequency of West Eurasian mtDNA haplotypes is significantly higher in upper castes than in lower castes ($p < 0.05$), the frequency of U2i haplotypes increasing as one moves from lower to higher castes. In addition, the frequency of mtDNA haplogroups with a more recent coalescence estimate (i.e., H, I, J, K, T) was fivefold higher in upper castes (6.8%) than in lower castes (1.4%). These haplotypes are derivatives of haplogroups found throughout Europe (Richards et al. 1998), the Middle East (Di Rienzo and Wilson 1991), and to a lesser extent Central Asia (Comas et al. 1998). Collectively, the mtDNA haplotype evidence indicate that contemporary Indian mtDNA evolved largely from proto-Asian ancestors with Western Eurasian admixture accounting for 20%–30% of mtDNA haplotypes.

Table 2. MtDNA Haplogroup Frequencies in Dravidic and Hindi-Speaking Indians

	Dravidic (%)			Hindi (%)	
Haplogroup	upper	middle	lower	total	
Asian	61 (5.5)	64.6 (3.8)	71.4 (5.3)	65.7 (2.7)	55.7 (2.9)
A	0	0	0	0	0.3 (.32)
B	0	0	0	0	0
F	0	0	0	0	2.7 (.94)
M	61 (5.5)	64.6 (3.8)	71.4 (5.3)	65.7 (2.7)	52.7 (2.9)
M3	18.6 (4.4)	3.5 (1.5)	1.4 (1.4)	6.6 (1.4)	6.0 (1.4)
M-C	0	0	0	0	0.7 (.48)
M-D	0	0	0	0	1.0 (.57)
M-G	0	0.9 (.74)	0	0.4 (.36)	0
M-E	0	1.8 (1.1)	0	0.8 (.51)	0
West Eurasian	23.7 (4.8)	14.2 (2.8)	7.1 (3.0)	14.5 (2.0)	27.4 (2.6)
U2i[b]	16.9 (4.2)	9.7 (2.3)	5.7 (2.7)	10.3 (1.7)	15.3 (2.1)
W	1.7 (1.5)	0	0	0.4 (.36)	3.7 (.29)
H	3.4 (2.0	0	0	1.2 (.62)	2.3 (.87)
I	0	0	0	0	1.3 (.65)
J	0	0.9 (.75)	0	0.4 (.36)	0.7 (.48)
K	1.7 (1.5)	0	0	0.4 (.36)	0
T	0	2.7 (1.3)	1.4 (1.4)	1.7 (.73)	1.7 (.75)
X	0	0	0	0	0.7 (.48)
Others	15.3 (4.1)	21.2 (1.3)	21.4 (4.8)	19.8 (2.3)	16.7 (2.2)

± standard errors are in parentheses.
[a]These haplotypes belong to super-haplogroup R (ancestral to haplogroups B, F, H, T, J, V, and U) but do not belong to any previously recognized haplogroup.
[b]U2i is differentiated from haplogroup U by the presence of a transition at np 16051.

Y-Chromosome Variation Confirms Indo-European Admixture

Genetic distances estimated from Y-chromosome STR polymorphisms differ significantly from zero ($p < 0.001$) and reveal a distinctly different pattern of population relationships (Table 3). In contrast to the mtDNA distances, the Y-chromosome STR data do not demonstrate a closer affinity to Asians for each caste group. Upper castes are more similar to Europeans than to Asians, middle castes are equidistant from the two groups, and lower castes are most similar to Asians. The genetic distance between caste populations and Africans is pro-

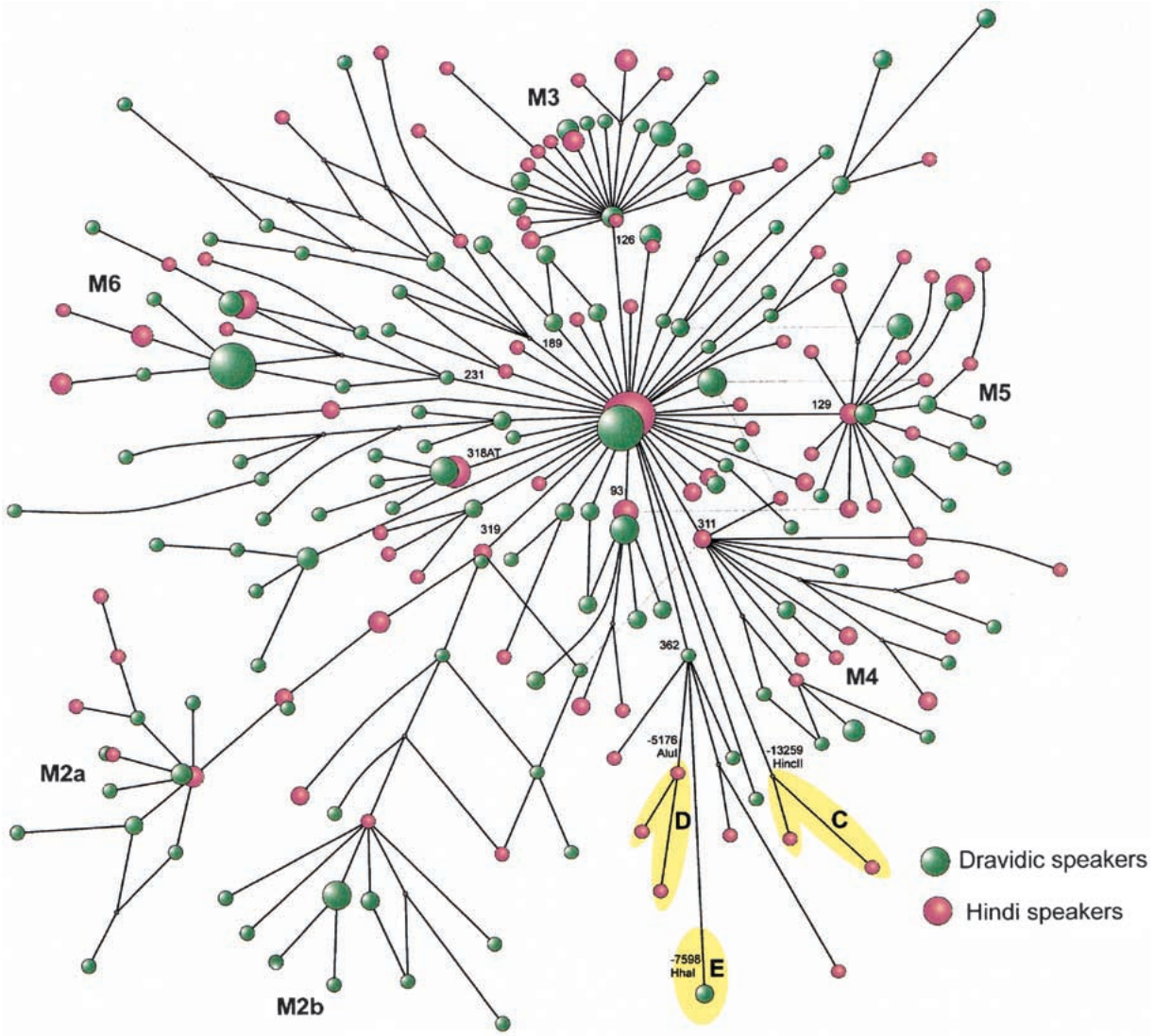

Figure 1 Phylogeny of haplogroup M in India. Phylogenetic relationships between HVR1 haplotypes were estimated by constructing reduced median networks. The size of each node is porportional to the haplotype frequency. Reticulations indicate parallel mutational pathways or multiple mutations. The identities of HVR1 mutations (numbered according to the Cambridge reference sequence +16000; Anderson et al. 1981) that define major haplogroup subsets are depicted along selected internodes. The coalescence estimate of Indian haplogroup-M haplotypes is 48,000 ± 1500 yr, suggesting that Indian-specific mtDNA haplotypes split from a proto-Asian ancestor in the late Pleistocene.

gressively larger moving from lower to middle to upper caste groups (Table 3).

Genetic distances estimated from Y-chromosome biallelic polymorphisms differ significantly from zero ($p < 0.05$), and the patterns differ from the mtDNA results even more strikingly than the Y-chromosome STRs. For Y-chromosome biallelic polymorphism data, each caste group is more similar to Europeans (Table 4), and as one moves from lower to middle to higher castes the genetic distance to Europeans diminishes progressively. This pattern is further accentuated by separating the European population into Northern, Southern, and Eastern Europeans; each caste group is most closely related to Eastern Europeans. Moreover, the genetic distance between upper castes and Eastern Europeans is approximately half the distance between Eastern Europeans and middle or lower castes. These results suggest that Indian Y chromosomes, particularly upper caste Y chromosomes, are more similar to European than to Asian Y chromosomes. This underscores the close affinities between Hindu Indian and Indo-European Y chromosomes based on a previously reported analysis of three Y-chromosome polymorphisms (Quintana-Murci et al. 1999b).

Overall, these results indicate that the affinities of Indians to continental populations varies according to

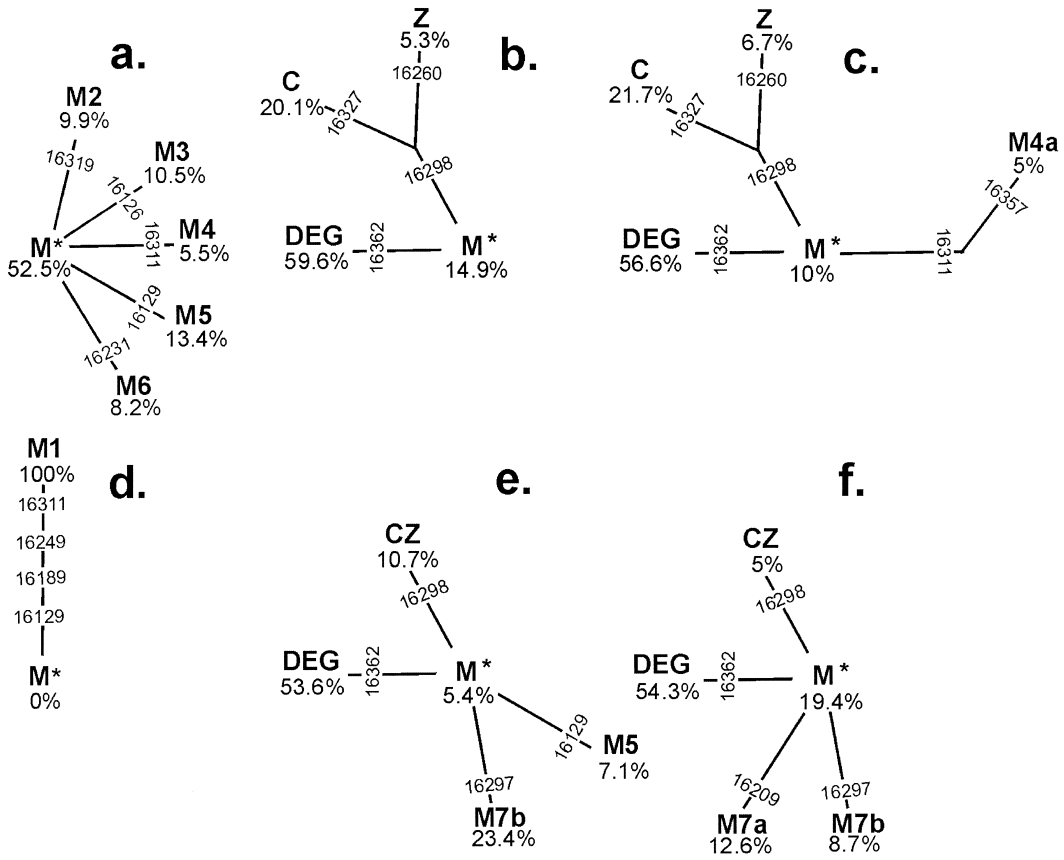

Figure 2 Major subsets of haplogroup M. Phylogenetic relationships of HVR1 haplotypes assigned to haplogroup M were estimated for: (*a*) 343 Indians (Quintana-Murci et al. 1999a; this study); (*b*) 16 Turks and 78 Central Asians (Comas et al. 1998; this study); (*c*) 60 Mongolians (Kolman et al. 1996); (*d*) 25 Ethiopians (Quintana-Murci et al. 1999a); (*e*) 56 Chinese (Horai et al. 1996; this study); (*f*) 103 Japanese (Horai et al. 1996; Seo et al. 1998). The founding node of each network (M*) differs from the CRS (Anderson et al. 1981) by transitions at np 10398, 10400, and 16223. The frequency of each subset of haplogroup M is indicated. Each phylogenetic network was pruned by eliminating branches containing haplotypes summing to a frequency of <5% (these branches were binned with the founder haplotype, M*). The identities of HVR1 mutations (numbered according to the CRS −16,000; Anderson et al. 1981) that define major haplotype subsets are depicted along selected internodes.

caste rank and depends on whether mtDNA or Y-chromosome data are analyzed. However, conclusions drawn from these data are limited because mtDNA and the Y chromosome is each effectively a single haploid locus and is more sensitive to genetic drift, bottlenecks, and selective sweeps compared to autosomal loci. These limitations of our analysis can be overcome, in part, by analyzing a larger set of independent autosomal loci. Consequently, we assayed 1 LINE-1 and 39 unlinked *Alu* polymorphisms.

Table 3. Y Chromosome (STRs) Genetic Distances between Caste Groups from Andhra Pradesh and Continental Populations

Caste group	Africans	Asians	Europeans
Upper	.0166	.0104	.0092
Middle	.0156	.0110	.0108
Lower	.0131	.0088	.0108
All castes	.0151	.0101	.0102

Affinities to Europeans and Asians Stratified by Caste Rank

Genetic distances estimated from autosomal *Alu* elements correspond to caste rank, the genetic distance between the upper and lower castes being more than 2.5 times larger than the distance between upper and middle or middle and lower castes (upper to middle, 0.0069; upper to lower, 0.018; middle to lower, 0.0071). These trends are the same whether the Kshatriya and Vysya are included in the upper castes, the middle castes, or excluded from the analysis (data not shown). Furthermore, a neighbor-joining network of genetic distances between separate castes (Fig. 3) clearly differentiates castes of different rank into separate clusters. This is similar to the relationship between genetic distances and caste rank estimated from

Table 4. Y Chromosome (Bi-Allelic Polymorphisms) Genetic Distances between Caste Groups from Andhra Pradesh and Continental Populations[a]

Caste group	Asians	Europeans	W. Europeans	S. Europeans	E. Europeans
Upper	.388	.135	.265	.168	.073
Middle	.291	.146	.249	.156	.133
Lower	.376	.173	.283	.189	.155

[a]Includes comparisons to unpublished data of M.F.H.

mtDNA (Bamshad et al. 1998). It is important to note, however, that the autosomal genetic distances are estimated from 40 independent loci. This afforded us the opportunity to test the statistical significance of the correspondence between genetic distance and caste status. The Mantel correlation between interindividual genetic distances and distances based on social rank was low but highly significant for individuals ranked into upper, middle, and lower groups ($r = 0.08$; $p < 0.001$) and into eight separate castes ($r = 0.07$; $p < 0.001$). Given the resolving power of this autosomal dataset, we next tested whether we could reconcile the results of the analysis of mtDNA and Y-chromosome markers in castes and continental populations.

Genotypic differentiation was significantly different from zero ($p < 0.0001$) between each pair of caste populations and between each caste and continental population. Similar to the results of both the mtDNA and Y-chromosome analyses, the distance between upper castes and European populations is smaller than the distance between lower castes and Europeans (Table 5). However, in contrast to the mtDNA results but similar to the Y-chromosome results, the affinity between upper castes and Europeans is higher than that of upper castes and Asians (Table 5). If the Kshatriya and Vysya are excluded from the analysis or included in the middle castes, the genetic distance between the upper caste (Brahmins) and Europeans remains smaller than the distance between the lower castes and Europeans and the distance between upper castes and Asians (Table 5). Analysis of each caste separately reveals that the genetic distance between the Brahmins and Europeans (0.013) is less than the distance between Europeans and Kshatryia (0.030) or Vysya (0.020). Nevertheless, each separate upper caste is more similar to Europeans than to Asians.

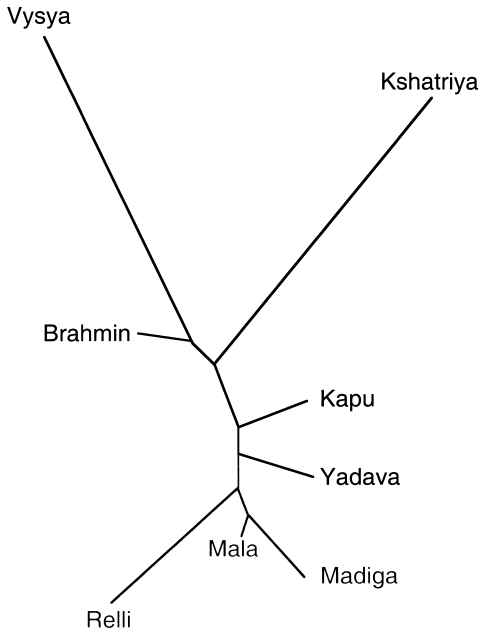

Figure 3 Neighbor-joining network of genetic distances among caste communities estimated from 40 *Alu* polymorphisms. Distances between upper castes (U; Brahmin, Vysya, Kshatriya), middle castes (M; Yadava, Kapu), and lower castes (L; Mala, Madiga, Relli) are significantly correlated with social rank.

Because historical evidence suggests greater affinity between upper castes and Europeans than between lower castes and Europeans (Balakrishnan 1978, 1982; Cavalli-Sforza et al. 1994), it is appropriate to use a one-tailed test of the difference between the corresponding genetic distances. The 90% confidence limits of Nei's standard distances estimated between upper castes and Europeans (0.006–0.016) versus lower castes and Europeans (0.017–0.037) do not overlap, indicating statistical significance at the 0.05 level. Significance at 0.05 is not achieved if the Kshatriya and Vysya are excluded. These results offer statistical support for differences in the genetic affinity of Europeans to caste populations of differing rank, with greater European affinity to upper castes than to lower castes.

DISCUSSION

Previous genetic studies have found evidence to support either a European or an Asian origin of Indian caste populations, with occasional indications of admixture with African or proto-Australoid populations (Chen et al. 1995; Mountain et al. 1995; Bamshad et al. 1996, 1997; Majumder et al. 1999; Quintana-Murci et al. 1999a). Our results demonstrate that for biparentally inherited autosomal markers, genetic distances between upper, middle, and lower castes are significantly correlated with rank; upper castes are more similar to Europeans than to Asians; and upper castes are significantly more similar to Europeans than are lower castes. This result appears to be owing to the amalgamation of two different patterns of sex-specific genetic variation.

The majority of Indian mtDNA restriction-site haplotypes belong to Indian-specific subsets (e.g., M3)

Table 5. Autosomal Genetic Distances[a] between Caste Groups from Andhra Pradesh and Continental Populations

Caste group	Africans	Asians	Europeans
Upper	.140 (0.074 ± .018)	.058 (0.024 ± .009)	.032[b] (.011 ± .003)
Middle	.149 (0.082 ± .018)	.032 (0.013 ± .005)	.057[c] (.020 ± .006)
Lower	.147 (0.083 ± .017)	.044 (0.017 ± .005)	.073 (.027 ± .006)
All castes	.147	.039	.045

[a]Nei standard distances ± standard errors are in parentheses.
[b]If the Kshatriya and Vysya are excluded, the genetic distance between the upper castes and Europeans is 0.038.
[c]If the Kshatriya and Vysya are grouped in the middle castes, the genetic distance between the middle castes and Europeans is 0.050.

of a predominantly Asian haplogroup M, although a substantial minority of mtDNA restriction site haplotypes belong to West Eurasian haplogroups. A higher proportion of proto-Asian mtDNA restriction-site haplotypes is found in lower castes compared to middle or upper castes, whereas the frequency of West Eurasian haplotypes is positively correlated with caste rank, that is, is highest in the upper castes. For Y-chromosome STR variation the upper castes exhibit greatest similarity with Europeans, whereas the lower caste groups are most similar to Asians. For Y biallelic polymorphism variation, each caste group is more similar to Europeans than to Asians, and the affinity to Europeans is proportional to caste rank, that is, is highest in the upper castes.

Importantly, five different types of data (mtDNA HVR1 sequence, mtDNA RSPs, Y-chromosome STRs, Y-chromosome biallelic polymorphisms, and autosomal *Alu* polymorphisms) support the same general pattern: relatively smaller genetic distances from European populations as one moves from lower to middle to upper caste populations. Genetic distances from Asian populations become larger as one moves from lower to middle to upper caste populations. It is especially noteworthy that the analysis of Y biallelic polymorphisms, which involved an independent set of comparative Asian, European, and African populations, again indicated the same pattern. Additional support is offered by the fact that the autosomal polymorphisms yielded a statistically significant difference between the upper-caste–European and lower-caste–European genetic distances. With additional loci, other differences (e.g., the distances between different caste groups and Asians) may also reach statistical significance.

The most likely explanation for these findings, and the one most consistent with archaeological data, is that contemporary Hindu Indians are of proto-Asian origin with West Eurasian admixture. However, admixture with West Eurasian males was greater than admixture with West Eurasian females, resulting in a higher affinity to European Y chromosomes. This supports an earlier suggestion of Passarino et al. (1996), which was based on a comparison of mtDNA and blood group results. Furthermore, the degree of West Eurasian admixture was proportional to caste rank. This explanation is consistent with either the hypothesis that proportionately more West Eurasians became members of the upper castes at the inception of the caste hierarchy or that social stratification preceded the West Eurasian incursion and that West Eurasians tended to insert themselves into higher-ranking positions. One consequence is that shared Indo-European languages may not reflect a common origin of Europeans and most Indians, but rather underscores the transfer of language mediated by contact between West Eurasians and native proto-Indians.

West Eurasian admixture in Indian populations may have been the result of more than one wave of immigration into India. Kivisild et al. (1999) determined the coalescence (~50,000 years before present) of the Indian-specific subset of the West Eurasian haplotypes (i.e., U2i) and suggested that West Eurasian admixture may have been much older than the purported Dravidian and Indo-European incursions. Our analysis of Indian mtDNA restriction-site haplotypes that do not belong to the U2i subset of West Eurasian haplotypes (i.e., H, I, J, K, T) is consistent with more recent West Eurasian admixture. It is also possible that haplotypes with an older coalescence were introduced by Dravidians, whereas haplotypes with a more recent coalescence belonged to Indo-Europeans. This hypothesis can be tested by a more detailed comparison to West Eurasian mtDNA haplotypes from Iran, Anatolia, and the Caucasus. Alternatively, the coalescence dates of these haplotypes may predate the entry of West Eurasians populations into India. Regardless of their origin, West Eurasian admixture resulted in rank-related differences in the genetic affinities of castes to Europeans and Asians. Furthermore, the frequency of West Eurasian haplotypes in the founding middle and upper castes may be underestimated because of the upward social mobility of women from lower castes (Bamshad et al. 1998). These women were presumably more likely to introduce proto-Asian mtDNA haplotypes into the middle and upper castes.

Our analysis of 40 autosomal markers indicates clearly that the upper castes have a higher affinity to Europeans than to Asians. The high affinity of caste Y chromosomes with those of Europeans suggests that the majority of immigrating West Eurasians may have

been males. As might be expected if West Eurasian males appropriated the highest positions in the caste system, the upper caste group exhibits a lower genetic distance to Europeans than the middle or lower castes. This is underscored by the observation that the Kshatriya (an upper caste), whose members served as warriors, are closer to Europeans than any other caste (data not shown). Furthermore, the 32-bp deletion polymorphism in CC chemokine receptor 5, whose frequency peaks in populations of Eastern Europe, is found only in two Brahmin males (M. Bamshad and S.K. Ahuja, unpubl.). The stratification of Y-chromosome distances with Europeans could also be caused by male-specific gene flow among caste populations of different rank. However, we and others have demonstrated that there is little sharing of Y-chromosome haplotypes among castes of different rank (Bamshad et al. 1998; Bhattacharyya et al. 1999).

The affinity of caste populations to Europeans is more apparent for Y-chromosome biallelic polymorphisms than Y-chromosome STRs. This could be attributed to the use of different European populations in comparisons using STRs and biallelic polymorphisms. Alternatively, it may reflect, in part, the effects of high mutation rates for the Y-chromosome STRs, which would tend to obscure relationships between caste and continental populations. A lack of consistent clustering at the continental level has been observed in several studies of Y-chromosome STRs (Deka et al. 1996; Torroni et al. 1996; de Knijff et al. 1997). The autosomal *Alu* and biallelic Y-chromosome polymorphisms, in contrast, have a slower rate of drift than Y-chromosome STRs because of a higher effective population size, and their mutation rate is very low. Thus, the Y-chromosome biallelic polymorphisms and autosomal *Alu* markers may serve as more stable markers of worldwide population affinities.

Our analysis may help to explain why estimates of the affinities of caste groups to worldwide populations have varied so widely among different studies. Analyses of recent caste history based on only mtDNA or Y-chromosome polymorphisms clearly would suggest that castes are more closely related to Asians or to Europeans, respectively. Furthermore, we attempted to minimize the confounding effect of geographic differences between populations by sampling from a highly restricted region of South India. Because of the ubiquity of the caste system in India's history, it is reasonable to predict similar patterns in caste populations living in other areas. Indeed, any genetic result becomes more compelling when it is replicated in other populations. Therefore, comparable studies in caste populations from other regions of India must be completed to test the generality of these results.

The dispersal and subsequent growth of Indian populations since the Neolithic Age is one of the most important events to shape the history of South Asia. However, the origin and dispersal route of the aboriginal inhabitants of the Indian subcontinent is unclear. Our findings suggest a proto-Asian origin of the Indian-specific haplogroup-M haplotypes. Haplogroup-M haplotypes are also found at appreciable frequencies in some East African populations— ~18% of Ethiopians (Quintana-Murci et al. 1999a) and 16% of Kenyans (M. Bamshad and L.B. Jonde, unpubl.). A comparison of haplogroup-M haplotypes from East Africa and India has suggested that this southern route may have been one of the original dispersal pathways of anatomically modern humans out of Africa (Quintana-Murci et al. 1999a). Together, these data support our previous suggestion (Kivisild et al. 1999) that India may have been inhabited by at least two successive late Pleistocene migrations, consistent with the hypothesis of Lahr and Foley (1994). It also adds to the growing evidence that the subcontinent of India has been a major corridor for the migration of people between Africa, Western Asia, and Southeast Asia (Cavalli-Sforza et al. 1994).

It should be emphasized that the DNA variation studied here is thought to be selectively neutral and thus represents only the effects of population history. These results permit no inferences about phenotypic differences between populations. In addition, alleles and haplotypes are shared by different caste populations, reflecting a shared history. Indeed, these findings underscore the longstanding appreciation that the distribution of genetic polymorphisms in India is highly complex. Further investigation of the spread of anatomically modern humans throughout South Asia will need to consider that such complex patterns may be the norm rather than the exception.

METHODS

Sample Collection

All studies of South Indian populations were performed with the approval of the Institutional Review Board of the University of Utah, Andhra University, and the government of India. Adult males living in the district of Visakhapatnam, Andhra Pradesh, were questioned about their caste affiliations and surnames and the birthplaces of their parents. Those who were unrelated to any other subject by at least three generations were considered eligible to participate.

We classified caste populations based upon the traditional ranking of these castes by *varna* (defined below), occupation, and socioeconomic status. According to various Sanskrit texts, Hindu populations were partitioned originally into four categories or *varna*: Brahmin, Kshatriya, Vysya, and Sudra (Tambia 1973; Elder 1996). Those in each *varna* performed occupations assigned to their category. Brahmins were priests; Kshatriya were warriors; Vysya were traders; and Sudra were to serve the three other *varna* (Tambia 1973; Elder 1996). Each *varna* was assigned a status; Brahmin, Kshatriya, and Vysya were considered of higher status than the Sudra because the Brahmin, Kshatriya, and Vysya are considered the twice-born

castes and are differentiated from all other castes in the caste hierarchy. This is the rationale behind classifying them as the upper group of castes (Tambia 1973).

The Kapu and the Yadava are called once-born castes that have traditionally been classified in the Sudra, the lowest of the original four *varna*. However, the status of the Sudra was actually higher than that of a fifth *varna*, the Panchama. This fifth *varna* was added at a later date to include the so-called untouchables, who were excluded from the other four *varna* (Elder 1996). The untouchable *varna* includes the Mala and Madiga. The position of the Relli in the caste hierarchy is somewhat ambiguous, but they have usually been classified in the lower caste group. Therefore, prior to the collection of any data, males from eight different Telugu-speaking castes ($n = 265$) were ranked into upper (Niyogi and Vydiki Brahmin, Kshatriya, Vysya [$n = 80$]), middle (Telega and Turpu Kapu, Yadava [$n = 111$]), and lower (Relli, Madiga, Mala [$n = 74$]) groups (Bamshad et al. 1998). This ranking has been used by previous investigators (Krishnan and Reddy 1994).

After obtaining informed consent, ~8 mL of whole blood or 5 plucked scalp hairs were collected from each participant. Extractions were performed at Andhra University using established methods (Bell et al. 1981).

MtDNA Polymorphisms

The mtDNA data consisted of 68, 116, and 73 HVR1 sequences and 79, 159, and 72 restriction-site haplotypes from largely the same individuals in upper, middle, and lower castes, respectively. These data were compared to data from 143 Africans (15 Sotho-Tswana, 7 Tsonga, 14 Nguni, 24 San, 5 Biaka Pygmies, 33 Mbuti Pygmies, 9 Alur, 18 Hema, and 18 Nande), 78 Asians (12 Cambodians, 17 Chinese, 19 Japanese, 6 Malay, 9 Vietnamese, 2 Koreans, and 13 Asians of mixed ancestry), and 99 Europeans (20 unrelated males of the French CEPH kindreds, 69 unrelated Utah males of Northern European descent, and 10 Poles) (Jorde et al. 1995, 1997). Mitochondrial sequence data from these 597 individuals are available at: http://www.genome.org/supplemental/.

In addition to our samples, the phylogenetic analyses also included data from 98 published HVR1 sequences from two castes (48 Havlik and 43 Mukri), and a tribal population (7 Kadar) living in south-western India (Mountain et al. 1995) and restriction-site haplotypes from one caste (62 Lobana) from Northern India, three tribal populations from Northern (12 Tharu and 18 Bhoksa) and Southern (86 Lambadi) India, and 122 individuals from various caste populations in Uttar Pradesh (Kivisild et al. 1999). Phylogenetic relationships of HVR1 sequences assigned to haplogroup M were estimated for Indians (this study), Turks (this study), Central Asian populations (Comas et al. 1998), Mongolians (Kolman et al. 1996), Chinese (Horai et al. 1996), and Japanese (Horai et al. 1996; Seo et al. 1998).

The mtDNA HVR1 sequence was determined by fluorescent Sanger sequencing using a Dye terminator cycle sequencing kit (Applied Biosystems) according to the manufacturer's specifications (Bamshad et al. 1998). Sequencing reactions were resolved on an ABI 377 automated DNA sequencer, and sequence data were analyzed using ABI DNA analysis software and SEQUENCHER software (Genecodes). To identify mtDNA haplotypes and haplogroups (a group of haplotypes that share some sequence variants), major continent-specific genotypes (Torroni et al. 1994, 1996; Wallace 1995) for the following polymorphic mtDNA restriction sites were determined: *Hpa*I$_{3592}$, *Dde*I$_{10394}$, *Alu*I$_{10397}$, *Alu*I$_{13262}$, *Bam*HI$_{13366}$, *Alu*I$_{5176}$, *Hae*III$_{4830}$, *Alu*I$_{7025}$, *Hinf*I$_{12308}$, *Acc*I$_{14465}$, *Ava*II$_{8249}$, *Alu*I$_{10032}$, *Bst*OI$_{13704}$, and *Hae*II$_{9052}$.

Y-Chromosome and Autosomal Polymorphisms

Y-chromosome-specific STRs (DYS19, DYS288, DYS388, DYS389A, DYS390) were amplified using published conditions (Hammer et al. 1998). PCR products were separated on an ABI 377 automated sequencer and scored using ABI Genotyper software. Y-chromosome STR data were collected from 622 males including 280 South Indians, ~200 Africans (Seielstad et al. 1999; this study), 40 Asians, and 102 Europeans. Autosomal data were collected from 608 individuals including 265 South Indians, 155 Africans, 70 Asians, and 118 Europeans.

The Y-chromosome-specific biallelic polymorphisms tested included: DYS188$_{792}$, DYS194$_{469}$, DYS211$_{105}$, DYS221$_{136}$, DYS257$_{108}$, DYS287, M3, M4, M9, M12, M15, SRY$_{4064}$, SRY$_{10831.1}$, SRY$_{10831.2}$, p12f2, PN1, PN2, PN3, RPS4Y$_{711}$, and Tat (Hammer and Horai 1995; Hammer et al. 1997, 1998, 2000; Underhill et al. 1997; Zerjal et al. 1997; Karafet et al. 1999). All individuals tested negative for the Y *Alu* insert (DYS287). A complete description of the Y-chromosome STR loci can be found in Kayser et al. (1997). A table of the biallelic Y-chromosome haplotype frequencies in the upper, middle, and lower castes is available at http://www.genome.org/supplemental/.

For the Y-chromosome biallelic dataset, comparisons were made to a different set of worldwide populations including: East Asians from Japan, Korea, China, and Vietnam ($n = 460$); Western Europeans from Britain and Germany ($n = 77$); Southern Europeans from Italy and Greece ($n = 148$); and Eastern Europeans from Russia and Romania ($n = 102$) (M.F. Hammer, unpubl.). The complete dataset of Indians consisted of 55 Brahmin, 111 Yadava and Kapu, and 74 Relli, Mala, and Madiga.

Autosomal polymorphisms were amplified using conditions specifically optimized for each system. Further information on these conditions is available at the Web site: http://www.genetics.utah.edu/~swatkins/pub/Alu_data.htm or http://www.genome.org/supplemental. With minor exceptions caused by typing failures or other causes, the same individuals from each population were used to create each dataset (i.e., mtDNA, Y chromosome, and autosomal). The complete dataset of genotypes from all 40 autosomal loci is available at: http://www.genome.org/supplemental/.

Statistical Analyses

Genetic distances for Y-chromosome STRs were estimated using the method of Shriver et al. (1995), which assumes a stepwise mutation model. Genetic distances for mitochondrial and autosomal markers were calculated as pairwise F_{ST} distances, using the ARLEQUIN package (Schneider et al. 1997). For autosomal polymorphisms, Nei's standard distances and their standard errors were estimated using DISPAN (http://www.bio.psu.edu/IMEG); and 90% confidence intervals were estimated by multiplying the standard error by 1.65. The significance of the F_{ST} distances between populations was estimated by generating a null distribution of pairwise F_{ST} distances by permuting haplotypes between populations. The p-value of the test is the proportion of permutations leading to an F_{ST} value larger than or equal to the observed one. Genotypic differentiation was estimated using GENEPOP (Raymond and Rousset 1995) vers. 3.2 (http://www.cefe.cnrs-mop.fr/). The null hypothesis tested is that there is a random

distribution of K different haplotypes among r populations (the contingency table). All potential states of the contingency table are explored with a Markov chain, and the probability of observing a table less than or equally likely to the observed sample configuration is estimated.

Estimates of significance for the correlation between interindividual caste rank differences and interindividual autosomal genetic distances were made by forming two $n \times n$ matrices, where n is the number of individuals. For the first matrix, interindividual genetic distances were based on the proportion of *Alu* insertions/deletions shared by each pair of individuals. To form the second matrix, each individual was assigned a score according to his rank in the caste hierarchy for caste groups (i.e., upper caste = 1, middle caste = 2, lower caste = 3) and also for separate castes (i.e., Brahmin = 1, Kshatriya = 2, Vysya = 3, Kapu = 4, Yadava = 5, Relli = 6, Mala = 7, and Madiga = 8). An interindividual matrix of score distances was formed by comparing the absolute value of the difference between the scores of each pair of individuals. The matrix of genetic distances was compared to 10,000 permuted matrices of score distances using a Mantel matrix comparison test (Mantel 1967).

To illustrate phylogenetic relationships we constructed reduced median (Bandelt et al. 1995) and neighbor-joining networks (Felsenstein 1989). Coalescence times were calculated as in Forster et al. (1996), using the estimator ρ, which is the average transitional distance from the founder haplotype.

ACKNOWLEDGMENTS

We thank all participants, the faculty and staff of Andhra University for their discussion and technical assistance, as well as Henry Harpending for comments and criticisms. We acknowledge the contributions of an anonymous reviewer who suggested that the Kshatriya and Vysya be analyzed separately from the other upper castes. Genetic distances between STRs were estimated by the program DISTNEW, kindly provided by L. Jin. This work was supported by NSF SBR-9514733, SBR-9700729, SBR-9818215, NIH grants GM-59290 and PHS MO1–00064, the Estonian Science Fund (1669 and 2887), and the Newcastle University small grants committee.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Anderson, S., Bankier, A.T, Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290:** 457–465.

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

———. 1982. Admixture as an evolutionary force in populations of the Indian sub-continent. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Human Genetics and Adaptation* (eds. K.C. Malhotra and A. Basu), **Vol. I:** 103–145. Indian Statistical Institute, Calcutta.

Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H., and Wallace, D.C. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* **130:** 139–152.

Bamshad, M., Fraley, A.E., Crawford, M.H., Cann, R.L., Busi, B.R., Naidu, J.M., and Jorde, L.B. 1996. mtDNA variation in caste populations of Andhra Pradesh, India. *Hum. Biol.* **68:** 1–28.

Bamshad, M., Bhaskara, R.B., Naidu, J.M., Prasad, B.V.R., Watkins, S. and Jorde L. 1997. Letters to the editor. *Hum. Biol.* **69:** 432–435.

Bamshad, M.J., Watkins, W.S., Dixon, M.E., Bhaskara, B.R., Naidu, J.M., Rasanayagam, A., Hammer, M.E., and Jorde, L.B. 1998. Female gene flow stratifies Hindu castes. *Nature* **395:** 651–652.

Bandelt, H.J., Forster, P., Sykes, B.C., and Richards, M.B. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141:** 743–753.

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

Bhattachayya, N.P., Basu P., Das, M., Pramanik, S., Banerjee, R., Roy, B., Roychoudhury, S., and Majumder, P. 1999. Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.* **9:** 711–719.

Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, NJ.

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S., and Wallace, D.C. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57:** 133–149.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., et al. 1998. Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am. J. Hum. Genet.* **63:** 1824–1838.

Deka, R., Jin, L., Shriver, M.D., Yu, L.M., Saha, N., Barrantes, R., Chakraborty, R., and Ferrell, R.E. 1996. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res.* **6:** 1177–1184.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., et al. 1997. Chromosome Y microsatellites: Population genetic and evolutionary aspects. *Int. J. Legal Med.* **110:** 134–149.

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

Elder, J. 1996. Enduring stereotypes about South Asia: India's caste system *Edu. Asia* **1:** 20–22.

Felsenstein, J. 1989. PHYLIP—Phylogeny inference package (version 3.2). *Cladistics* **5:** 164–166.

Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.* **59:** 935–945.

Hammer, M..F. and Horai, S. 1995. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56:** 951–962.

Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., et al. 1997. The geographic distribution of human Y chromosome variation. *Genetics* **145:** 787–805.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., and Zegura, S.L. 1998. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15:** 427–441.

Hammer, M.F., Redd A.J., Wood, E.T., Bonner, M.R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim A., Jobling, M.A., Jenkins, T., et al. 2000. Jewish and middle eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci.* **97:** 6769–6774.

Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K.S., Omoto, K., and Pan, I.H. 1996. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* **59:** 579–590.

Jorde, L. B., Bamshad, M.J., Watkins, W.S., Zenger, R., Fraley, A.E., Krakowiak, P.A., Carpenter, K.D., Soodyall, H., Jenkins, T., and Rogers, A.R. 1995. Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57:** 523–538.

Jorde, L.B., Rogers, A.R., Bamshad, M., Watkins, W.S., Krakowiak, P., Sung, S., Kere, J., and Harpending, H.C. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci.* **94:** 3100–3103.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., and Batzer, M.A. 2000. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66:** 979–988.

Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., et al. 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64:** 817–831.

Kayser, M., de Knijff, P., Dieltjes, P., Krawczak, M., Nagy, M., Zerjal, T., Pandya, A., Tyler-Smith, C., and Roewer, L. 1997. Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* **18:** 1602–1607.

Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S., Dixon, M.E., et al. 1999. Deep common ancestry of Indian and western Eurasian mtDNA lineages. *Curr. Biol.* **9:** 1331–1334.

Kolman, C.J., Sambuughin, N., and Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142:** 1321–1334.

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

Lahr, M.M. and Foley, R.A. 1994. Multiple dispersals and modern human origins. *Evol. Anthr.* **3:** 48–60.

Majumder, P.P. 1999. People of India: Biological diversity and affinities. *Evol. Anthr.* **6:** 100–110.

Majumder, P.P. and Mukherjee, B.N. 1993. Genetic diversity and affinities among Indian populations: An overview. In *Human population genetics* (ed. P.P. Majumder), pp. 255–275. Plenum Press, New York.

Majumder, P.P., Roy, B., Banerjee, S., Chakraborty, M., Dey, B., Mukherjee, N., Roy, M., Thakurta, P.G., and Sil, S.K. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur. J. Human Genet.* **7:** 435–446.

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

Mlhotra, K.C. and Vasulu, T.S. 1993. Structure of human populations in India. In *Human population genetics* (ed. P.P. Majumder), pp. 207–233. Plenum Press, New York.

Mountain, J.L. Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M., and Cavalli-Sforza, L.L. 1995. Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56:** 979–992.

Nei, M. and Livshits, G. 1989. Genetic relationships of Europeans, Asians and Africans and the origin of modern *Homo sapiens*. *Hum. Hered.* **39:** 276–281.

Passarino, G., Semino, O., Bernini, L.F., and Santachiara-Benerecetti, A.S. 1996. Pre-Caucasoid and Caucasoid genetic features of the Indian population revealed by mtDNA polymorphisms. *Am. J. Hum. Genet.* **59:** 927–934.

Poliakov, L. 1974. *The Aryan Myth.* Basic Books, New York.

Quintana-Murci, L., Semino, O., Poloni, E.S., Liu, A., Van Gijn, M., Passarino, G., Brega, A., Nasidze, I.S., Maccioni, L., Cossu, G., et al. 1999a. Y-Chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: A comparative study. *Ann. Hum. Genet.* **63:** 153–166.

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K., and Santachiara-Benerecetti, A.S. 1999b. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genet.* **23:** 437–441.

Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenism. *J. Heredity* **86:** 248–249.

Renfrew, C. 1989a. Before Babel: Speculations on the origins of linguistic diversity. *Camb. Archaeol. J.* **1:** 3–23.

———. 1989b. The origins of Indo-European languages. *Sci. Am.* **261:** 82–90.

Richards, M.B., Macaulay, V.A., Bandelt, H.J., and Sykes, B.C. 1998. Phylogeography of mitochondrial DNA in Western Europe. *Ann. Hum. Genet.* **61:** 251–254.

Schneider, S., Rosslie, D., and Excoffier, L. 1997. *Arlequin ver 2.000: A software for population genetics data analysis.* Genetics and Biometry Laboratory, University of Geneva, Geneva.

Seielstad, M., Bekele, E., Ibrahim, M., Toure, A., and Traore, M. 1999. A view of modern human origins from Y chromosome microsatellite variation. *Genome Res.* **9:** 558–567.

Seo, Y., Stradmann-Bellinghausen, B., Rittner, C., Takahama, K., and Schneider, P.M. 1998. Sequence polymorphism of mitochondrial DNA control region in Japanese. *Forensic Sci.* **97:** 155–164.

Shaffer, J.G. 1982. Harappan culture: A reconsideration. In *Harappan civilization: A contemporary perspective* (ed. G.L. Possehl), American Institute of Indian Studies, pp. 41–50. Oxford and IBH Publishers, New Delhi, India.

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

Tambia, S.J. 1973. *The character of kinship* (ed. J. Goody). Cambridge University Press, Cambridge, UK.

Torroni, A., Lott, M.T., Cabell, M..F, Chen, Y.S., Lavergne, L., and Wallace, D.C. 1994. mtDNA and the origin of Caucasians: Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am. J. Hum. Genet.* **55:** 760–776.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L., and Wallace, D.C. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144:** 1835–1850.

Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., and Oefner, P.J. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7:** 996–1005.

Wallace, D.C. 1995. 1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am. J. Hum. Genet.* **57:** 201–223.

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhovel, W., Fretwell, N., Jobling, M.A., Harihara, S., et al. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60:** 1174–1183.

Received November 29, 2000; accepted in revised form March 22, 2001.