

Independent Origins of Indian Caste and Tribal Paternal Lineages

Richard Cordaux,^{1,5,*} Robert Aunger,²

Gillian Bentley,³ Ivane Nasidze,¹

S.M. Sirajuddin,⁴ and Mark Stoneking¹

¹Max Planck Institute for Evolutionary Anthropology

D-04103 Leipzig

Germany

²Environmental Health Group

London School of Hygiene and Tropical Medicine

London WC1E 7HT

England

³Department of Anthropology

University College London

London WC1E 6BT

England

⁴Anthropological Survey of India

Mysore 570002, Karnataka State

India

Summary

The origins of the nearly one billion people inhabiting the Indian subcontinent and following the customs of the Hindu caste system [1, 2] are controversial: are they largely derived from Indian local populations (i.e. tribal groups) or from recent immigrants to India? Archaeological and linguistic evidence support the latter hypothesis [2–4], whereas recent genetic data seem to favor the former hypothesis [5]. Here, we analyze the most extensive dataset of Indian caste and tribal Y chromosomes to date. We find that caste and tribal groups differ significantly in their haplogroup frequency distributions; caste groups are homogeneous for Y chromosome variation and more closely related to each other and to central Asian groups than to Indian tribal or any other Eurasian groups. We conclude that paternal lineages of Indian caste groups are primarily descended from Indo-European speakers who migrated from central Asia ~3,500 years ago. Conversely, paternal lineages of tribal groups are predominantly derived from the original Indian gene pool. We also provide evidence for bidirectional male gene flow between caste and tribal groups. In comparison, caste and tribal groups are homogeneous with respect to mitochondrial DNA variation [5, 6], which may reflect the sociocultural characteristics of the Indian caste society.

Results and Discussion

Earlier genetic studies discussing the origins of Indian caste groups have included at best a few tribal groups for comparison [5, 7–9]. Although they constitute only ~8% of the total Indian population, tribal groups are

generally considered to be the aboriginal inhabitants of the Indian subcontinent, present in the region before the arrival of Indo-European speakers [2]. As such, they represent a unique source for estimating the in situ pre-Indo-European genetic diversity of India as well as for investigating the origins of caste populations of India. Recently, a qualitative comparison of presence versus absence of Y chromosome haplogroups in just two tribal and six caste groups led to the conclusion that both Indian caste and tribal Y chromosomes largely derive from the same Pleistocene genetic heritage, with only limited recent gene flow from external sources. This conclusion implies an in situ origin of paternal lineages of caste groups [5], which is at odds with nongenetic evidence [2–4].

To determine if Indian caste paternal lineages are derived from local ancestors (i.e., tribal groups) or from other Eurasian source(s), we obtained new Y chromosome data from 155 individuals from nine tribal groups and one caste group and compared these to published data [5, 9, 10]. The total dataset consists of 931 Y chromosomes from 15 tribal and 12 caste groups and constitutes the most extensive dataset of Indian Y chromosomes to date. The studied caste groups originate from all over India, whereas the tribal groups were sampled only from southern India. Tribal groups from elsewhere were not included as most of them inhabit the Indo-European-speaking sphere and hence may not reflect the pre-Indo-European genetic composition of India.

Thirteen haplogroups were observed in India (Figure 1). The seven most frequent haplogroups account for 80%–90% of both caste and tribal Y chromosomes, suggesting an extensive overlapping of caste and tribal Y chromosome variation, as found previously [5]. However, the frequency distribution of these haplogroups is significantly different between tribal and caste groups ($\chi^2 = 134.1$, degrees of freedom [d.f.] = 7, $p < 0.001$). At a finer scale, the same trend is observed in that tribal groups differ significantly from both northern ($\chi^2 = 132.5$, d.f. = 7, $p < 0.001$) and southern ($\chi^2 = 94.1$, d.f. = 7, $p < 0.001$) Indian caste groups.

The most frequent haplogroups in tribal groups are H-M52 and F-M89. By contrast, these haplogroups are significantly rarer in caste groups (Figure 1). Haplogroup H-M52 is largely restricted to the Indian subcontinent [5], and given its high frequency in tribal groups, it likely has a local Indian origin. Haplogroup F-M89 is a “default” haplogroup potentially comprising several lineages. Since Y chromosome markers have been identified from nontribal individuals [11, 12], markers for tribal-specific Y sublineages of F-M89 would not have been detected in the screening process, suggesting that a significantly higher proportion of tribal than caste Y chromosomes cannot be allotted to haplogroups determined from nontribal individuals. Haplogroup O-M95 also shows a significantly higher frequency in tribal than caste groups (Figure 1) and might be associated with the spread of Austro-Asiatic languages [5], which are exclusively spoken by tribal groups in India.

*Correspondence: cordaux@lsu.edu

⁵ Present address: Department of Biological Sciences, 202 Life Sciences Building, Louisiana State University, Baton Rouge, Louisiana 70803

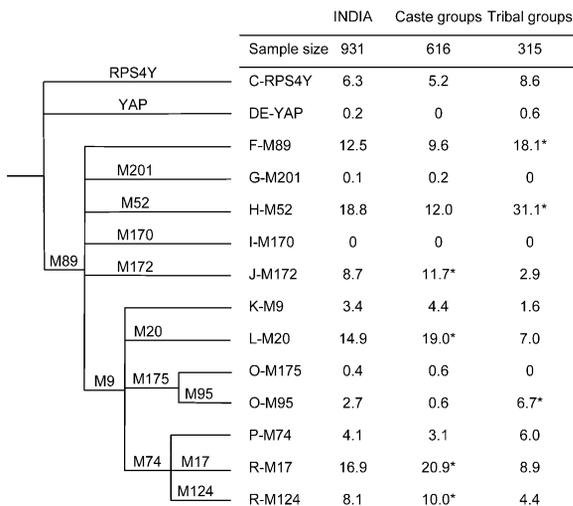


Figure 1. Y Chromosome Haplogroup Frequencies in 931 Indian Males from 27 Populations

Haplogroup relationships are shown with haplogroup-defining markers along the relevant branches of the tree. Haplogroup frequencies are given as a percentage. Data from [10] were excluded from the calculation of frequencies of haplogroups P-M74 and R-M124 because of missing information. *, significantly higher frequency ($p < 0.01$) in a χ^2 test comparing caste versus tribal groups.

Haplogroups R-M17, J-M172, R-M124, and L-M20 are among the most frequent Y lineages in caste groups. They are all significantly more frequent in caste than in tribal groups (Figure 1). The average frequency of R-M17 in 15 tribal groups from four different states of India is only 9% (or 6% if the Chenchus are excluded). Thus, the unusually high frequency of R-M17 in the Chenchu tribe (27%) is not representative of other tribal groups and hence cannot be taken as evidence for an Indian origin of R-M17, as claimed previously [5]. By contrast, R-M17 is present in all Indian caste groups and reaches a frequency of 40% in north caste groups [5]. Given the high frequency of R-M17 in central Asia (typically 20%–40% [9]), its rarity in west Asia [9, 13] and its absence in east Asia [14], Indian R-M17 Y chromosomes most probably have a central Asian origin [8, 9]. Haplogroup J-M172 in India may have a west Asian origin [8]. However, it was noted that the M67 marker, which is common in west Asian J-M172 chromosomes, is almost absent from Indian J-M172 chromosomes [5]. Given that J-M172 is rare in Indian tribal groups (Figure 1), absent in east Asia [9], and typically found in central Asia at frequencies of 10%–20% [9], it is possible that Indian J-M172 chromosomes originate from central Asia rather than west Asia. Haplogroup R-M124 is restricted to the Indian subcontinent, Iran, and central Asia [5]. It generally occurs at low frequencies (1%–4%) except in Indian caste groups and Indo-European speakers from central Asia (8% [9]). Haplogroup L-M20 is found predominantly in India and Pakistan (15%) and has tentatively been associated with the expansion of farming, thus implying a nonIndian origin [15].

In sum, although largely the same haplogroups are found in tribal and caste groups, they exhibit significantly different distributions in that the most frequent

Table 1. Estimated Indigenous and Nonindigenous Contributions to Indian Caste and Tribal Y Chromosome Gene Pools

	Nonindigenous Contribution	Indigenous Contribution
Caste groups	74%	26%
North caste groups	88%	12%
South caste groups	68%	32%
Tribal groups	29%	71%

These estimates are based on the frequencies of seven haplogroups (which account for >80% of Indian Y chromosomes), assuming that haplogroups H-M52, O-M95, and F-M89 have indigenous origins, whereas J-M172, L-M20, R-M17 and R-M124 have nonindigenous origins given their putative phylogeography (see text).

haplogroups in tribal groups are significantly rarer in caste groups and vice versa. Moreover, haplogroups that are likely to be of indigenous origin are in higher frequency in tribal groups, whereas haplogroups that are likely to be of nonindigenous origin are higher in frequency in caste groups. Indeed, we estimate through a phylogeographic approach (Table 1) that 74% of the caste Y chromosome gene pool has nonindigenous origins, whereas 71% of the tribal Y chromosome gene pool has indigenous origins. A preferable approach would be to obtain statistical estimates of these contributions, but this requires estimates of the Y haplogroups frequencies in the ancestral populations, which in the present case are impossible to estimate because tribal groups have received Y chromosomes from caste groups and, hence, cannot provide estimates of Y haplogroup frequencies prior to contact.

To investigate the relationships of caste and tribal groups by simultaneously taking into account the information provided by all Y chromosome haplogroups, we calculated F_{st} distances between all pairs of Indian groups and performed a multidimensional scaling (MDS) analysis (Table 2, Figure 2). We also included 604 individuals from west and east Europe, and west, central, and east Asia. To compare average F_{st} values between groups of populations, we performed t tests using average F_{st} standard errors calculated by resampling over populations (Table 2) via jackknife and permutations procedures (see the Supplemental Data section for additional details). Resampling over populations is appropriate because (1) F_{st} standard errors are affected by both errors from sampling populations and individuals, but in practice, the error from sampling individuals is negligible compared to the error from sampling populations [16], and (2) the null hypothesis to be tested is not $F_{st} = 0$ (which would require resampling over individuals), but $F_{st}(\text{first group of populations}) = F_{st}(\text{second group of populations})$. Hence, we find that north and south caste groups are significantly ($p < 0.001$) more closely related to each other (average $F_{st} = 0.072$) than either is to tribal groups (average $F_{st} = 0.149$; Table 2). In addition, Indian caste groups are significantly ($p < 0.001$) more similar to central Asians (average $F_{st} = 0.062$) than to Indian tribal or any other Eurasian groups (average $F_{st} \geq 0.15$). The same trend persists when Indian caste populations are subdivided into south and north caste groups (Table 2). It is noteworthy that Indian

Table 2. Y Chromosome Average Fst Distances (and Standard Errors) between Indian and Other Eurasian Groups

	West Europe	East Europe	West Asia	Central Asia	East Asia	Indian Tribal Groups	
						All Groups	Koragas Removed
India	0.307 (0.006)	0.231 (0.006)	0.155 (0.008)	0.101 (0.005)	0.300 (0.013)	—	—
Indian caste groups	0.265 (0.006)	0.158 (0.008)	0.154 (0.010)	0.062 (0.005)	0.261 (0.017)	0.149 (0.008)	0.126 (0.007)
North caste groups	0.259 (0.011)	0.108 (0.011)	0.158 (0.016)	0.040 (0.007)	0.259 (0.028)	0.165 (0.011)	0.141 (0.010)
South caste groups	0.270 (0.011)	0.198 (0.014)	0.151 (0.013)	0.081 (0.011)	0.263 (0.020)	0.136 (0.009)	0.115 (0.007)
Indian tribal groups	0.339 (0.011)	0.286 (0.010)	0.155 (0.013)	0.130 (0.009)	0.329 (0.020)	—	—
Reduced dataset ^a	0.337 (0.022)	0.298 (0.019)	0.216 (0.020)	0.165 (0.018)	0.343 (0.025)	—	—

^a Groups for which the sample size is less than ten were removed.

caste groups are equally distantly related to geographically neighboring tribal groups and to geographically distant west Asians or east Europeans. By contrast, Indian tribal groups exhibit large average Fst values (>0.13) both with Indian caste and all other Eurasian populations. Because small sample sizes for some tribal populations might have resulted in artificial affinities, the above analyses were repeated after removing groups for which the sample size is less than ten. However, the aforementioned trends did not change, except that Indian tribal groups were separated from other Eurasian groups by average Fst values higher than 0.16 (Table 2).

Analyses of molecular variance also suggest that caste groups are more homogeneous for Y chromosome variation than tribal groups, since the variance among caste groups (sampled from all over India) is 3-fold less than that observed among tribal groups and 2-fold less than that observed among all Indian populations grouped together (Table 3). Moreover, if only north caste groups are considered, the variance among populations is not significantly different from zero (Table 3), indicating that although they are located up to ~1500 km away from each other, these populations have highly homogeneous Y chromosome compositions.

It is unlikely that the observed opposite patterns of Y chromosome haplogroup frequencies in Indian tribal and caste groups are solely the result of genetic drift effects. This is because we based our comparisons on pooled data from a large number of caste and tribal

groups, and even if genetic drift may have been important in some particular populations, the pooling is expected to smooth individual drift effects and reconstruct ancestral frequencies. Moreover, the caste system was elaborated only within the last 3500 years [2, 17], which constitutes a rather short period of time for observing such dramatic differences in Y chromosome variation between caste and tribal people at a pooled-data level. Furthermore, drift would be unlikely to produce the regular pattern observed in which haplogroups of inferred nonindigenous origin are always in higher overall frequency in caste than in tribal groups, whereas haplogroups of inferred indigenous origin are always in higher overall frequency in tribal than in caste groups.

It therefore appears that Indian caste and tribal paternal lineages derive from independent original sources. Most Indian caste Y chromosomes have a central Asian origin, whereas Indian tribal groups seem to have largely retained Y lineages inherited from the original Indian Y gene pool (Table 1). The central Asian Y chromosomes spread over the Indian subcontinent recently and in a rather short period of time, as suggested by the relative homogeneity and close relationship of all caste populations. A possible explanation for these patterns is that Indian caste paternal lineages are largely descended from the linguistically and archaeologically inferred dispersal of Indo-European-speaking pastoralists who migrated from central Asia some 3500 years ago [2–4, 7–9].

However, the sharing of most haplogroups between

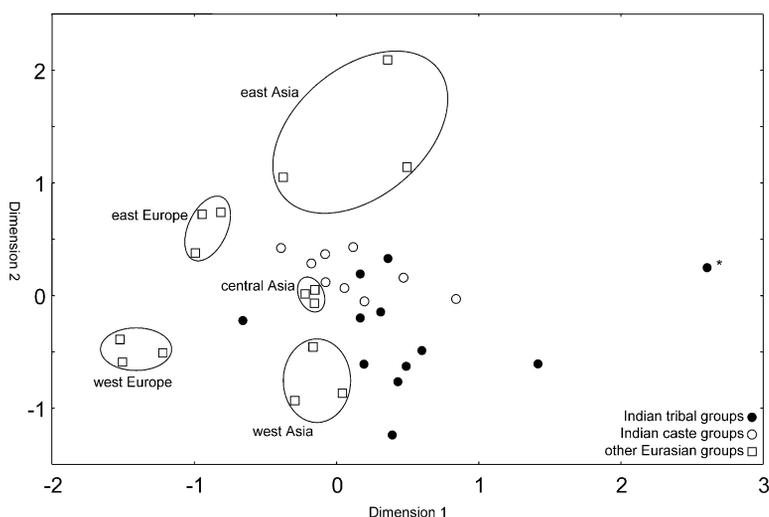


Figure 2. Multidimensional Scaling Plot of 36 Eurasian Populations Based on Fst Distances. The Koraga tribal population is identified by an asterisk (*). Stress value: 0.18.

Table 3. Results of Analyses of Molecular Variance

Grouping	Variance within Populations	Variance among Populations
21 Indian populations	86.0%	14.0%
9 caste populations	92.6%	7.4%
4 north caste populations	98.3%	1.7%
5 south caste populations	93.5%	6.5%
12 tribal populations	79.5%	20.5%

Note: all values differ significantly from zero after Bonferroni correction for multiple tests, except the variance among 4 north caste groups.

caste and tribal groups indicates that subsequent admixture occurred between Indo-European newcomers and local tribal populations (Table 1). The evidence shows that Y chromosome admixture was limited in north Indian caste groups and more pronounced in south caste groups [10, 18]. A possible explanation for this geographic discrepancy is that the caste system comprised four classes in north India, whereas a fifth class was introduced in south India to integrate local people (those formerly called “untouchables”) in the caste system [19, 20]. This view finds support in that in south India, lower caste groups are more similar to Asians, whereas higher caste groups are more similar to west Eurasians [7]. In addition, the presence of caste-typical haplogroups at low frequency in tribal groups (Table 1) may indicate paternal admixture from Indian caste to tribal groups. Such admixture would not be unexpected during the spread of Indo-European languages and assimilation of indigenous populations in the Indian subcontinent in that Indo-European speakers were integrated in non-Indo-European speech communities, especially in south India [4].

Thus, the quantitative comparison of an extensive dataset of Y chromosome haplogroups in both Indian caste and tribal groups, as well as nongenetic information, support a scenario of independent origins of Indian caste and tribal paternal lineages, with recent immigration of caste Y lineages and subsequent bidirectional gene flow between caste and tribal groups. This conclusion contrasts with the earlier suggestion that both Indian caste and tribal Y chromosomes largely derive from the same Pleistocene genetic heritage, with only limited recent gene flow from external sources [5]. The most likely reason for this discrepancy is that this conclusion was based on the occurrence of the same haplogroups in tribal and caste groups; while we concur with this observation, we also show that haplogroup frequencies differ significantly between caste and tribal groups. The quantitative comparison allows us to distinguish between scenarios of shared ancestry versus recent gene flow when the groups share the same haplogroups.

In contrast with the Y chromosome evidence, the mtDNA evidence suggests a common origin of tribal and caste groups [5, 6]. It is likely that most maternal lineages largely represent the original mtDNA gene pool of India, implying that caste maternal lineages mainly derive from local tribal ancestors [5–7]. In comparison, west Eurasian-typical mtDNA lineages are rare or absent in tribal groups [5, 6, 21] and represent a small fraction

of caste mtDNAs [5, 7]. This led to the proposal that the Indo-European migration to India was male mediated [7]. Subsequently, these migrating males would have mixed with local females. Alternatively, we suggest that the Indo-European migration may not have been sex biased, but rather the Indo-European females were subsequently largely replaced by indigenous ones as a consequence of the combined practices of hypergyny (i.e., a higher-ranking male mating with a lower-ranking female) and preferential female infanticide, both of which are known to have occurred in India [22]. Hypergyny created an upward mobility of females [22, 23] and thus a spread of “lower-ranking” mtDNAs (i.e., those of the assimilated indigenous groups). On the other hand, female infanticide occurred preferentially (but not exclusively) in higher-ranking castes [22], leading to the removal of “higher-ranking” mtDNAs (i.e., initially those of Indo-European females) from reproduction. Such a system would ultimately result in the replacement of nonindigenous mtDNAs by indigenous ones without necessarily requiring an initial male-mediated Indo-European migration into India. The current frequency of west Eurasian-typical mtDNAs in Indian caste populations of ~10% [5] and the practice of female infanticide over centuries, perhaps at rates as high as 30%–80% in some groups [22], suggest that (1) the replacement was not completely achieved, consistent with the recent origins of the caste system [2, 17], and (2) the current frequency of west Eurasian-typical mtDNAs in the Indian caste gene pool may underestimate the ancestral frequency, consistent with the view that the Indo-European migrants were not necessarily mostly males. Analyses of additional loci would shed further light; in any event, these observations underscore the potential considerable influence of sociocultural practices in shaping human genetic variation [23–25].

Supplemental Data

Supplemental Data including a detailed description of the Experimental Procedures and a table summarizing Y chromosome haplogroup compositions of the 27 populations analyzed here are available at <http://www.current-biology.com/cgi/content/full/14/3/231/DC1/>.

Acknowledgments

We are grateful to the original donors of samples and to Nilmani Saha for contributing DNA samples. We thank Peter Underhill for kindly providing DNA controls, Toomas Kivisild and Spencer Wells for information about some Indian groups, and Jerome Goudet for helpful comments. We also thank Silke Brauer, Axel Erler, David Hughes, Manfred Kayser, Brigitte Pakendorf, and Dominique Quinque for fruitful discussions during the course of this study and useful comments on an earlier version of this manuscript. R.C., I.N., and M.S. were supported by funds from the Max Planck Society, Germany. R.A. and G.B. were supported by funds from the LSB Leakey Foundation, National Geographic Society (4877-92) and NSF DBS-9221426.

Received: June 24, 2003

Revised: November 3, 2003

Accepted: December 18, 2003

Published: February 3, 2004

References

1. Majumder, P.P. (1998). People of India: biological diversity and affinities. *Evol. Anthropol.* 6, 100–113.

2. Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. (1994). *History and Geography of Human Genes* (Princeton: Princeton University Press).
3. Fairservis, W.A. (1975). *The Roots of Ancient India* (Chicago: University of Chicago Press).
4. Masica, C.P. (1991). *The Indo-European Languages* (Cambridge: Cambridge University Press).
5. Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**, 313–332.
6. Cordaux, R., Saha, N., Bentley, G.R., Aunger, R., Sirajuddin, S.M., and Stoneking, M. (2003). Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* **11**, 253–264.
7. Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.V.R., Reddy, P.G., Rasanayagam, A., et al. (2001). Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**, 994–1004.
8. Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S.H., Hammer, M.F., Mehdi, S.Q., Ayub, Q., Qamar, R., Mōhyuddin, A., Radhakrishna, U., et al. (2001). Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am. J. Hum. Genet.* **68**, 537–542.
9. Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., et al. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* **98**, 10244–10249.
10. Ramana, G.V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P., and Chakraborty, R. (2001). Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* **9**, 695–700.
11. Y-Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348.
12. Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonn -Tamir, B., Bertranpetit, J., Francalacci, P., et al. (2000). Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361.
13. Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvazsi, A., Limborska, S., et al. (2000). The genetic legacy of *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–1159.
14. Su, B., Xiao, C., Deka, R., Seielstad, M.T., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P., Cavalli-Sforza, L., Chakraborty, R., et al. (2000). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107**, 582–590.
15. Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Makhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C., and Mehdi, S.Q. (2002). Y-chromosomal variation in Pakistan. *Am. J. Hum. Genet.* **70**, 1107–1124.
16. Pons, O., and Petit, R.J. (1995). Estimation, variance and optimal sampling of gene diversity, I. Haploid locus. *Theor. Appl. Genet.* **90**, 462–470.
17. Gadgil, M., and Thapar, R. (1990). Human ecology in India: some historical perspectives. *Interdiscipl. Sci. Rev.* **15**, 209–223.
18. Bhattacharyya, N.P., Basu, P., Das, M., Pramanik, S., Banerjee, R., Roy, B., Roychoudhury, S., and Majumder, P.P. (1999). Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.* **9**, 711–719.
19. Majumder, P.P. (2001). Indian caste origins: genomic insights and future outlook. *Genome Res.* **11**, 931–932.
20. Misra, V.N. (2001). Prehistoric human colonization of India. *J. Biosci.* **26**, 491–531.
21. Roychoudhury, S., Roy, S., Basu, A., Banerjee, R., Vishwanathan, H., Usha Rani, M.V., Sil, S.K., Mitra, M., and Majumder, P.P. (2001). Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* **109**, 339–350.
22. Dickemann, M. (1979). Female infanticide, reproductive strategies and social stratification: a preliminary model. In *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective*, N.A. Chagnon, and W. Irons, eds. (North Scituate, MA: Duxbury Press), pp. 321–367.
23. Bamshad, M.J., Watkins, W.S., Dixon, M.E., Jorde, L.B., Rao, B.B., Naidu, J.M., Prasad, B.V., Rasanayagam, A., and Hammer, M.F. (1998). Female gene flow stratifies Hindu castes. *Nature* **395**, 651–652.
24. Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., and Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilineal vs. patrilineal residence. *Nat. Genet.* **29**, 20–21.
25. Dupanloup, I., Pereira, L., Bertorelle, G., Calafell, F., Prata, M.J., Amorim, A., and Barbujani, G. (2003). A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J. Mol. Evol.* **57**, 85–97.