Current Biology, Vol. 14, 231–235, February 3, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.cub.2004.01.024

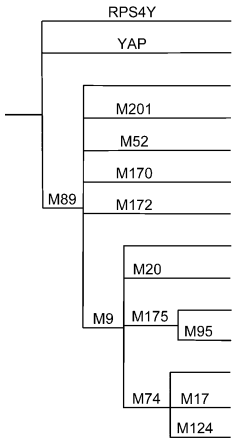
Independent Origins of Indian Caste and Tribal Paternal Lineages

Richard Cordaux,[1,5,*] Robert Aunger,[2] Gillian Bentley,[3] Ivane Nasidze,[1] S.M. Sirajuddin,[4] and Mark Stoneking[1]
[1]Max Planck Institute for Evolutionary Anthropology
D-04103 Leipzig
Germany
[2]Environmental Health Group
London School of Hygiene and Tropical Medicine
London WC1E 7HT
England
[3]Department of Anthropology
University College London
London WC1E 6BT
England
[4]Anthropological 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 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



	INDIA	Caste groups	Tribal groups
Sample size	931	616	315
C-RPS4Y	6.3	5.2	8.6
DE-YAP	0.2	0	0.6
F-M89	12.5	9.6	18.1*
G-M201	0.1	0.2	0
H-M52	18.8	12.0	31.1*
I-M170	0	0	0
J-M172	8.7	11.7*	2.9
K-M9	3.4	4.4	1.6
L-M20	14.9	19.0*	7.0
O-M175	0.4	0.6	0
O-M95	2.7	0.6	6.7*
P-M74	4.1	3.1	6.0
R-M17	16.9	20.9*	8.9
R-M124	8.1	10.0*	4.4

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].

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 Fst 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 Fst values between groups of populations, we performed t tests using average Fst standard errors calculated by resampling over populations (Table 2) via jackknife and permutations procedures (see the Supplemental Data section for addi-

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,
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: Duxburry 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 matrilocal vs. patrilocal 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.