Mitochondrial DNA variation in ranked caste groups of Maharashtra (India) and its implication on genetic relationships and origins

SANGITA ROY†, CHITRA THAKUR (MAHADIK)‡ and PARTHA P. MAJUMDER§

† Human Genetics and Genomics Division, Indian Institute of Chemical Biology, Kolkata, India
‡ Research Society, B.J. Wadia Hospital for Children, Parel, Mumbai, India
§ Anthropology & Human Genetics Unit, Indian Statistical Institute, Kolkata, India

Summary. *Background*: Polymorphisms in mitochondrial DNA have proven to be useful in studying genetic relationships and origins. The origins of caste populations have remained an enigma and genetic relationships among ranked caste groups are not uniform across geographical regions of India.
Aim: This study was undertaken to investigate the nature and extent of mtDNA variation and relationships among caste groups of the western Indian State of Maharashtra and examine the implication of the results on their origins.
Subject and methods: One population was selected from each of the three caste ranks and blood samples were obtained with informed consent from unrelated individuals. The ranked caste populations were: upper (Brahmin; $n = 31$), middle (Maratha; $n = 41$) and lower (Nava Baudh; $n = 40$). Ten relevant restriction site polymorphisms (RSPs) and one Insertion/Deletion (InDel) polymorphism were studied. The Hypervariable Segment 1 (HVS1) was sequenced in a subset of sampled individuals.
Results: Four RSP loci were found to be monomorphic in all populations. The InDel locus was monomorphic in two (Brahmin and Maratha) populations. One haplotype, constructed on the basis of the RSPs, was found to be predominant in all populations. Haplotype diversity was of similar magnitudes among the Maratha and Nava Baudh (68% and 64%, respectively), and was much higher than among the Brahmin (49%). The frequency of haplogroup M was found to be high in all three groups, but, contrary to expectations, was highest in the upper caste Brahmin. About 10% of Brahmins, however, possessed the haplogroup C. Extensive variation was found in the HVS1 region. The nucleotide diversities and mean number of mismatches were found to be of similar magnitudes in all three groups.
Conclusions: The upper caste group, Brahmin, is genetically distinct from the middle and lower caste groups. However, in view of the highest frequency of haplogroup M among the Brahmin, it appears that there may have been recruitment from other populations into this group.

1. Introduction

The contemporary people of India are culturally stratified as tribals and non-tribals. The non-tribal people primarily belong to the Hindu religious fold and are hierarchically arranged in castes. The origins of the castes in India remain an enigma (Majumder 2001). Many castes are known to have tribal origins (Kosambi 1964, Karve 1961). The most accepted theory is that this unique social system was probably introduced to India by Aryan-speaking people (Thapar 1966) who migrated from Europe, the Near East, Anatolia and the Caucasus about 3000–8000 years ago (Poliakov 1974, Renfrew 1989a,b). Multiple waves of migration into India during prehistoric and historic times and the subsequent cultural differentiation resulting in strict rules governing mating practices are two of the major causes of the genomic diversity observed among the contemporary ethnic groups of India.

The Indian State of Maharashtra occupies roughly a wedge-shaped area between the Arabian Sea on the west and the highlands of Bastar in the east and between the Satpura Mountains in the north and the plains of Andhra Pradesh and Karnataka in the south. The seaboard region, called the 'Konkan', is the west coast of India and is partly isolated from the rest of Maharashtra. Marathi is the language spoken by most of the caste groups of Maharashtra. This language belongs to the Indo-European linguistic family. To examine whether the caste groups belonging to different social rungs have different genetic origins, we selected one group from each of the three hierarchical levels (upper, middle and lower) to study mitochondrial DNA polymorphisms. Each of the selected groups practises endogamy and hence it is expected that there has been little gene flow among these groups.

2. Subjects and methods

2.1. *Populations studied*

Three populations from three different rungs of the Hindu caste hierarchy were selected. Blood samples were obtained with informed consent from unrelated individuals belonging to each population group. The selected populations were:

Brahmin (upper caste) (KBR): The Brahmins are traditionally priests and occupy the highest position in the Hindu social hierarchy. They are now engaged in various white-collar occupations. We sampled Brahmins ($n = 31$) from the Konkan region.

Maratha (middle caste) (MRT): The Marathas represent 50% of the population of Maharastra (a state in western region of India). People of this caste claim to be Kshatriya (warriors); the second highest Hindu social rank. They primarily practise agriculture now. The sample size of this group was 41.

Nava Baudh (lower caste) (NBH): The Nava Baudhs predominantly inhabit the Ratnagiri district of Maharastra. These groups have adopted Buddhism and are therefore known as the Nava (New) Baudh (Buddhist). This community, upon adoption of Buddhism, has drawn people from different lower-caste groups, e.g. Chamar, Mahar, who are traditionally menial labourers, and continue to practise the same occupation. We sampled 40 unrelated individuals from this group.

2.2. *Laboratory analyses*

DNA was isolated from each individual using the salting-out procedure (Miller *et al.* 1988). Each DNA sample was screened in respect of a set of 10 loci, using PCR amplification followed by restriction digestion, agarose gel electrophoresis, ethidium bromide staining and band visualization under UV light. The primers used for PCR amplification were according to Roychoudhury *et al.* (2000). Additionally, we analysed the COII/$tRNA^{lys}$ intergenic deletion of one of the two copies of 9 bp tandem repeat sequence (CCCCCTCTA) that occurs between np 8272–8289, previously used as a marker for populations of Asian and Asian-derived origins, such as Polynesians and Native Americans (Rickards 1995). The sequences of Hypervariable Segment 1 (HVS1) from position 16024–16380 (relative to the revised Cambridge reference sequence (CRS); Andrews *et al.* 1999) were amplified and sequenced as described in Roychoudhury *et al.* (2001).

2.3. *Statistical analyses*

Allele and haplotype frequencies were calculated and haplotype diversity was estimated according to Nei (1987). Genetic distance among the populations was estimated using the D_A distance measure (Nei 1987), and genetic affinities were estimated using the neighbour-joining algorithm (Saitou and Nei 1987). Nucleotide diversity was calculated according to Nei (1987).

3. Results

3.1. *Allele frequencies and average heterozygosities*

Locus-specific allele frequencies for the three populations and the average heterozygosities are presented in table 1. All three populations are monomorphic at the *Hae*III np 663, *Hpa*I np 3592, *Alu*I np 5176 and *Alu*I np 7025 loci; no sampled individual showed the presence of these restriction sites. However, while *Dde*I (10394), *Alu*I (10397) and *Hae*III (16517) loci are polymorphic in all three populations, the Nava Baudh population was monomorphic at the *Hin*cII (13259) and *Alu*I (13262) loci. The 9 bp COII/tRNA^{lys} intergenic length mutation was not detected among Brahmins and Marathas. This length mutation (one insertion) was observed only among Nava Baudh (1 of 40 individuals). There was considerable variability (0.016–0.396) in heterozygosities across the polymorphic loci (table 1). The *Alu*I np 10397 locus exhibited the highest heterozygosity of 0.396, and the 9 bp deletion locus showed the minimum heterozygosity (0.016).

3.2. *Haplotype frequencies and diversities*

With respect to the 11 loci (10 biallelic and 1 InDel), 11 distinct haplotypes were observed among the 112 individuals examined from three ethnic populations. However, in none of the populations were all the 11 haplotypes observed. The maximum number of haplotypes (eight) was observed among the Maratha; the Brahamin and Nava Baudh contained equal numbers (seven) of haplotypes. The frequencies of haplotypes in each study population are presented in table 2. One haplotype, 00111101010, accounted for 59.8% of all mtDNA molecules. It is also evident from the table 2 that this haplotype was modal in the pooled data set and in each of the three study populations. Most haplotypes were shared across populations and occurred with frequencies that were not significantly different across the populations. The unshared haplotypes occurred sporadically.

Haplotype diversities were calculated for each population; these are presented in table 2. The haplotype diversity in most of the populations was quite high

Table 1. Estimated allele frequencies (%), with standard errors (SEs), and heterozygosities at the RSP and InDel loci*.

Population code	*Dde*I np 10394		*Alu*I np 10397		*Hin*fI np 12308		*Hin*cII np 13259		*Alu*I np 13262		*Hae*III np 16517		9 bp del	
	(–)	SE	(–)	SE	(–)	SE	(–)	SE	(–)	SE	(–)	SE	(–)	SE
KBR	16.13	6.61	16.13	6.61	90.32	5.31	9.68	5.31	90.32	5.31	6.45	4.41	100.00	–
MRT	34.15	7.41	36.59	7.52	92.68	4.07	2.44	2.41	97.56	2.41	21.95	6.46	100.00	–
NBH	35.00	7.54	35.00	7.54	82.50	6.01	0.00	–	100.00	–	15.00	5.65	97.50	2.47
Heterozygosity	0.392		0.396		0.200		0.074		0.074		0.239		0.016	

*The *Hae*III np 663, *Hpa*I np 3592 and *Alu*I np 7025 loci were monomorphic for the (–) allele and the *Alu*I np 5176 locus was monomorphic for the (+) allele in all populations.

Table 2. Frequencies (%) of haplotypes and estimated haplotype diversities based on 11 RSP and InDel loci in three ethnic populations of Maharashtra and in the pooled sample.

Serial no.	Haplotype*	KBR ($n=31$)	MRT ($n=41$)	NBH ($n=40$)	Total ($n=112$)
1	00110000110		1 (2.44)		1 (0.90)
2	00110001000	1 (3.23)	4 (9.76)	1 (2.50)	6 (5.40)
3	00110001010	2 (6.45)	6 (14.63)	6 (15.00)	14 (12.50)
4	00110011000	1 (3.23)	1 (2.44)	1 (2.50)	3 (2.70)
5	00110011003			1 (2.50)	1 (0.90)
6	00110011010	1 (3.23)	2 (4.88)	5 (12.50)	8 (7.10)
7	00111001010		1 (2.44)		1 (0.90)
8	00111100110	3 (9.68)			3 (2.70)
9	00111101000		4 (9.76)	3 (7.50)	7 (6.30)
10	00111101010	22 (70.97)	22 (53.66)	23 (57.50)	67 (59.80)
11	00111111010	1 (3.23)			1 (0.9)
Haplotype diversity		0.494	0.684	0.639	–

*The order of loci: *Hae*III np 663, *Hpa*I np 3592, *Alu*I np 5176, *Alu*I np 7025, *Dde*I np 10394, *Alu*I np 10397, *Hinf*I np 12308, *Hinc*II np 13259, *Alu*I np 13262, *Hae*III np 16517, 9 bp deletion. 1 = presence of restriction site, 0 = absence of restriction site.

(49.40–68.40%). The highest haplotype diversity was found among Maratha (68.40%). The haplotype diversity among the Nava Baudh was of a similar magnitude (64%). The Brahmins exhibited a markedly lower diversity (49.40%). We also carried out ANOVA analysis to partition diversity into within- and between-population components. The results showed that the maximum percentage of variation in haplotype frequencies was attributable to among-individuals (97.8%) within-populations; only 2.2% of the variation was attributable to between-populations.

3.3. *Haplogroup frequencies*

Based on the co-occurrence of mutations, mtDNA molecules have been classified into several haplogroups (Torroni *et al.* 1994). The frequencies of these haplogroups show significant geographical variation, and also variation across major human morphological groups—Caucasoid, Mongoloid, etc. (Chen *et al.* 1995). The restriction site polymorphisms (RSPs) examined in the present study permitted classification of the data into the following haplogroups: A, B, C, D, H, L, M, U. Certain haplotypes could not be classified into any of the above haplogroups. The percentage frequencies of unclassified haplogroups among Brahmin, Maratha and Nava Baudh were, 9.68, 29.27 and 17.5, respectively.

Figure 1 presents the frequencies of these haplogroups in the three study populations. Haplogroups A, B, D and H were not observed in any of the populations. Of particular interest are the frequencies of haplogroups M and U. Haplogroup M has been proposed to be an ancient east-Asian marker (Ballinger *et al.* 1992) and is virtually absent among African and Caucasoid populations (Torroni *et al.* 1994, Passarino *et al.* 1996a,b). However, Quintana-Murci *et al.* (1999) have proposed that the origin of haplogroup M is in Africa, in view of its high frequency in Ethiopia. Haplogroup U has been found in high frequencies among Caucasoid populations (Kivisild *et al.* 1999), making it suitable for identifying Caucasoid admixture in Indian populations. The Caucasian admixture in northern Indian populat[illegible]

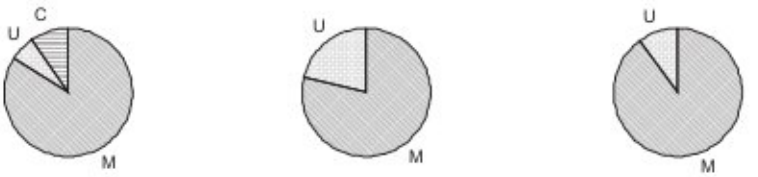

Figure 1. Frequencies of M and U haplogroups in three ethnic populations of Maharashtra.

was also highlighted by Barnabas *et al.* (1996), who had also shown the absence of Caucasian mtDNA types in southern India.

The Brahmins have the highest frequency (83.87%) of haplogroup M, which is significantly higher ($p < 0.05$) than in Maratha (63.41%) and Nava Baudh (65%). The frequencies of haplogroup M between Marathas and Nava Baudhs were not significantly different. The frequency of haplogroup U was highest among the Nava Baudh (17.5%) followed by Maratha (7.32%) and Brahmin (6.45%). These frequency differences are, however, not statistically significant ($p > 0.45$). Haplogroup C, which predominantly occurs in Siberia and in American Indians, was only found (figure 1) among the Brahmin (9.68%).

3.4. *Analysis of HVS1 sequences*

The mitochondrial HVS1 was sequenced for 10 individuals from each ethnic group, using both forward and reverse primers. Upon alignment of the HVS1 sequences from the study populations against the CRS (Cambridge Reference Sequence) (Andrews *et al.* 1999), four gaps were introduced in the CRS: one after np 16169, one after 18183 and two after 18189. These four gaps were eliminated from all the sequences prior to statistical analysis. The stretch of 357 nucleotides of the HVS1 region showed deletions of nucleotide at position 16183 in several individuals. A total of 49 polymorphic sites were observed. A list of all polymorphic sites and the corresponding nucleotide changes at each position are presented in the Appendix.

Sharing of the HVS1 region has been primarily observed between individuals within the same population. This was observed mainly among the Brahmin. Different sub-haplogroups of haplogroup M have been constructed (table 3) based on HVS1 sequence polymorphisms. Sub-haplogroups M*, M2 and M4 were found to be the most frequent sub-haplogroups in the study populations.

3.5. *Nucleotide diversities*

Nucleotide diversities and mean number of mismatches have been calculated (table 4) separately for each population. It was found that although the lower caste population, Nava Baudh, exhibited the highest nucleotide diversity and mean number of mismatches compared to the upper caste (Brahmin) and the middle caste (Maratha) populations, the differences were not statistically significant.

3.6. *Phylogenetic analysis*

To examine the genetic affinities among three ethnic populations a neighbour-joining tree was constructed using Nei's D_A distance based on the RSP haplotype frequencies given in table 2. The unrooted neighbour-joining tree (not shown) showed that the Brahmins were genetically distinct from the other two populations: Maratha and Nava Baudh.

Table 3. Frequencies of various known sub-haplogroups of haplogroups M and U in three ethnic populations of Maharashtra.

	M					U			
Population	M*	M2	M3	M4	M5	U1	U2i	U2e	U7
Brahmin	8	6		2					1
Maratha	5	1		4	1		2	1	1
Nava Baudh	7	2	1	3	1		3	1	

Table 4. Descriptive statistics and nucleotide diversities in populations of Maharashtra belonging to three different socio-cultural groups.

	Population		
	Brahmin	Maratha	Nava Baudh
No. of sequences	10	10	10
No. of polymorphic sites	25	27	27
Nucleotide diversity (π) $\pm$ 2 SD	0.019 ± 0.011	0.018 ± 0.010	0.022 ± 0.013
Mean no. of mismatches	6.96 ± 3.57	6.57 ± 3.39	8.08 ± 4.10

4. Discussion

RSP marker and sequence data of the mtDNA hypervariable segments (HVS1 and HSV2) have already been used widely and successfully to examine the origins and relationships of human populations. The high variability in the hypervariable regions makes them particularly useful markers for reconstructing recent evolutionary events. However, the high mutation rates are known to sometimes confound population relationships. RSPs and InDel polymorphisms on the other hand, despite being shared by most populations worldwide, often at similar frequencies, have a much lower mutation rate and provide a way of determining long-term evolutionary relationships.

Populations belonging to the Hindu caste fold in India are primarily endogamous, and there is a social rank-ordering of these populations. The origins of the populations have remained an enigma (Kosambi 1964). The Brahmins occupy the highest social rung. They have traditionally been the torch-bearers of Aryan rituals, which were introduced to India by Indo-European speakers who arrived in India from Central Asia and the Middle East (Thapar 1966). A recent study (Bamshad *et al.* 2001) examined the question of caste origins and, by studying some population groups of Andhra Pradesh, concluded that the upper castes were most similar to the Europeans, followed by the middle and lower castes. The present study was conducted to examine the origins of the ranked caste populations of Maharashtra. Since only the maternally inherited mtDNA has been investigated, our study is restricted to a maternal perspective of genetic origins.

The mtDNA RSP haplotype data (table 2) showed that there is extensive sharing of a single haplotype (00111101010) across the three caste groups of Maharashtra which belong to distinct ordered social ranks. This reveals a commonality of mtDNA lineages among the ranked caste groups. This haplotype has also been found at fairly high frequencies among many populations, both tribal and caste, from disparate geographical regions of India (Roychoudhury *et al.* 2000, 2001 and our unpublished data). Thus, it appears that there may have been a small ancient

founding group of Indian populations. In view of its wide distribution at high frequencies among geographically and ethnically disparate populations of India, we also believe that this haplotype is the most ancient in India. It is, however, also possible that this haplotype has recently expanded in frequency, while other haplotypes have decreased or have been lost due to drift. We do not believe that this is a parsimonious explanation, otherwise there would have been considerable local fluctuations in the frequencies of this haplotype which is not observed, and the haplotype is modal across a disparate array of ethnic populations in India (Roychoudhury *et al.* 2000, 2001 and our unpublished data).

An analysis of the frequencies of phylogenetically distinct haplogroups provides further insights. It has been found in earlier studies (Roychoudhury *et al.* 2000, Bamshad *et al.* 2001) that haplogroup U, which is a Caucasian haplogroup, is prevalent among the caste populations, particularly in the north Indian upper caste populations. Its frequency decreases as one descends the ladder of social rank, from the upper caste to the lower caste populations. In the present study no such pattern was observed (figure 1) and the upper caste population harboured the lowest frequency of haplogroup U. An interesting feature that emerged from this study was that about 10% of the upper caste Brahmins possessed haplogroup C. Haplogroup C is predominantly found in Siberia. It is, therefore, possible that there has been a specific flow of genes into upper caste populations of Maharashtra, because this haplogroup has not been found in the middle and lower caste samples in our study.

On the other hand, the frequency of haplogroup M, which is an ancient Asian haplogroup (Ballinger *et al.* 1992), was the highest among Brahmins. In India, haplogroup M occurs at highest frequency in tribal populations and is predominant in central and southern India. It is, therefore, striking that the frequency of this haplogroup should be the significantly higher among Brahmin than among the Maratha and Nava Baudh of Maharashtra.

Our phylogenetic analysis revealed that the Maratha and the Nava Baudh were genetically more similar to each other than to the Brahmin. The main reason why the Brahmin stand out as being genetically distinct is that they possessed a high frequency of haplogroup C, which was not found in either of the two other populations.

The present study, therefore, provides evidence of somewhat different origins of the ranked caste groups of Maharashtra. Although, on the one hand, the populations belonging to distinct tiers of the caste hierarchy predominantly share one haplotype, there is at least one distinctive feature of the upper caste group—they possess haplogroup C, which is absent in the other two groups. It is puzzling that the Brahmin should possess the highest frequency of haplogroup M. From the results of previous studies (Roychoudhury *et al.* 2000, 2001, Bamshad *et al.* 2001), one would have expected the Brahmins to possess higher frequencies of haplogroup U and lower frequencies of haplogroup M compared to the Maratha and Nava Baudh. Although premature, it is tempting to speculate that there may have been large-scale gene flow from other populations harbouring high haplogroup M frequencies into the Brahmins of Maharashtra. Kosambi (1964) had raised the possibility of tribal origins of some caste groups, which is not inconsistent with the present data since most tribal populations in India possess high frequencies of haplogroup M. It may, therefore, be interesting to specifically conduct a genetic study to test this hypothesis in Maharashtra.

Acknowledgements

This study was partially funded by grants from the Department of Biotechnology, Government of India. We also thank the B.J. Wadia Research Society for their help and support.

Appendix

Raw data: nucleotides at various positions in the HVS1 segment for 30 individuals drawn from three ethnic populations of Maharashtra. A dot (.) in a particular nucleotide position indicates that the nucleotide is the same as the nucleotide at this position occurring in the Cambridge Reference Sequence (CRS). The CRS is provided in the first line of the table below.

POPLN.	NUCLEOTIDE POSITION - 16000
CODE-	
SL.NO.	00000000000000111111111111111111111111111111111111
	33456678889999000112223344444455556666666777778888
	78816951262345248014694602457834682346789236890123
CRS	AAGAACTACTTTTCTCCGCTTGCTTCTGCCGTGAAAAACCCTCCTCAAAA
KBR-001	..........C.....
KBR-002	.....
KBR-003	.....
KBR-004	.....
KBR-005	..........T.....G...
KBR-006	.....
KBR-007	.....
KBR-008	.....
KBR-015	.....
KBR-016	.....
MRT-001	.....
MRT-002	...G.....T....
MRT-003	..........A.....
MRT-004	.....
MRT-005	....G.....
MRT-006	...G......C.....C.....
MRT-007	.....
MRT-008	.....
MRT-009	.....
MRT-017	.....T.....
NBH-008	...G.....A.....C.....
NBH-017	...G......C.....C.....
NBH-033	.....
NBH-034	.....
NBH-035	.....
NBH-036	..........A.....
NBH-038	.....T.....
NBH-039	.....
NBH-040	...G....T.....A.....A.....
NBH-041	...G....T.....C......A.....

```
                1111111112222222222222222222222222222222222222222333333333333333333333333333333333
                0888889990001111222223333444444455555666666777777888889999990000001111122222233344445555555666
                4567890236793457802345701459012356789046789013456902456846789012345678901245912689012457052342345679028

CRS             CCCCCTCCCCAMTGCATCACCTCAATCACAACTCAACTCACCACCCTCACACGATCACCTACCCACCCTTAACATAATAAAGCCATTCTATACTCCCTTTCTT

KBR-001         ..........................TC.....................T.A.....................................A.....C.........
KBR-002         .....C....................T.....................A.....................................AT..............
KBR-003         ..........................T.....................G.....................................A.T.............
KBR-004         ..........................T.....................G.....................................A.T.............
KBR-005         .....T....................T.....................G.....................C...............A.T.............
KBR-006         ..........................T.....................G.....T...............C..A.T.............C.
KBR-007         .....C....................T.....TT..............A.....................C..A.....................C.
KBR-008         ..........................T.....................A.....................C..C.....AT..............
KBR-015         ..........................T.....................G...T.....................G..T.....................
KBR-016         .....G....................T.....................G...T.....................G..T.....................
MRT-001         .....T....................T.....................G...T.....................G..T.....................
MRT-002         ..........................T.....................G...T.....................C..T.....................
MRT-003         ..........C...............T.....................G...T.....................C..T.....................
MRT-004         ..........................T.....................G...T.....................C..T.....................
MRT-005         .....C....................T.....................G...T.....................C..T.....................
MRT-006         ..........................T.....................G...T.....T...............C..T.....C.
MRT-007         .....C....................T.....................G...T.....T...............C..T.....TC.....
MRT-008         ..........A...............T.....................G...T.....T...............C..T.....
MRT-009         .....G....................T.....................G...T.....T...............G..T.....
MRT-017         ..........................T.....................T.A...T.....T...............C..A.....C.
NBH-008         .....C....................G.....................G...T.....T...............C..T.....
NBH-017         .....C....................T.....................G...T.....T...............C..T.....
NBH-033         ..........................T.....................G...T.....T...............T..T.....T.
NBH-034         .....C....................T.....................T.A...T.....T...............A..T.....C.
NBH-035         .....C....................T.....................G...T.....A...............C..AT....C.
NBH-036         ..........................T.....................G...T.....T...............C..T.....
NBH-038         ..........................T.....................G...T.....T...............T..T.....
NBH-039         ..........................T.....................G...T.....T...............T..T.....
NBH-040         ..........................T.....................T...T.....T...............T..T.....
NBH-041         ..........................T.....................T...T.....T...............C..T.....
```

References

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., and Howell, N., 1999, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, **23,** 147.

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 continuity of ancient mongoloid migrations. *Genetics*, **130,** 139–152.

Bamshad, M., Kivisild, T., Watkins, W. S., Dixon, M. E., Ricker, L. E., Rao, B. B., Naidu, M., Prasad, B. V. R., Reddy, P. G., Rasanayagam, A., Papiha, S. S., Villems, R., Redd, A. J., Hammer, M. F., Nguyen, S. V., Carroll, M. L., Batzer, M. A., and Jorde, L. B., 2001, Genetic evidence on the origins of Indian caste populations. *Genome Research*, **11,** 994–1004.

Barnabas, S., Apte, R. V., and Suresh, C. G., 1996, Ancestry and interrelationships of the Indians and their relationships with other world populations: a study based on mitochondrial DNA polymorphisms. *Annals of Human Genetics*, **60,** 409–422.

Chen, Y.-S., Torroni, A., Excoffier, L., Santachiara-Benercetti, A. S., and Wallace, D. C., 1995, Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *American Journal of Human Genetics*, **57,** 133–149.

Karve, I., 1961, *Hindu Society—an Interpretation* (Bombay: Deshmukh Prakashan).

Kivisild, T., Bamshed, M. J., Kaldma, K., Metspalu, E., Reidla, M. Laoss., Parik, J., Watkins, W. S., Dixon, M. E., Papiha, S. S., Mastana, S. S., Mir, M. R., Ferak, V., and Villems, R., 1999, Deep common ancestry of Indian and Western-Eurasian mtDNA lineages. *Current Biology*, **9,** 1331–1334.

Kosambi, D. D., 1964, *The Culture and Civilization of Ancient India in Historical Outline*, 1991 reprint (New Delhi: Vikas Publishing House Pvt Ltd).

Majumder, P. P., 2001, Indian caste origins: genomic insights and future outlook. *Genome Research*, **11,** 931–932.

Miller, S. A., Dykes, D. D., and Polesky, H. F., 1988, A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acid Research*, **16,** 1215.

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

Passarino, G., Semino, O., Bernini, L. F., and Santachiara-Benercetti, A. S., 1996a, Pre-Caucasoid and Caucasoid genetic features if the Indian population, revealed by mtDNA polymorphisms. *American Journal of Human Genetics*, **59,** 927–934.

Passarino, G., Semino, O., Modiano, G., and Santachiara-Benerecetti, A. S., 1996b, mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals. *Annals of Human Biology*, **23,** 121–126.

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

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

Renfrew, C., 1989a, Before Babel: speculations on the origins of linguistic diversity. *Cambridge Archaeology Journal*, **1,** 3–23.

Renfrew, C., 1989b, The origins of Indo-European languages. *Scientific American*, **261,** 82–90.

Rickards, O., 1995, Analysis of the region V mitochondrial marker in two Black communities of Ecuador and in their parental populations. *Human Evolution*, **10,** 5–16.

Roychoudhury, S., Roy, S., Dey, B., Chakraborty, M., Roy, M., Roy, B., Ramesh, A., Prabhakarn, N., Usha Rani, M. V., Vishwanathan, H., Mitra, M., Sil, S. K., and Majumder, P. P., 2000, Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Current Science*, **79,** 1182–1192.

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 structure and population histories of linguistically distinct tribal groups of India. *Human Genetics*, **109,** 339–350.

Saitou, N., and Nei, M., 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4,** 406–425.

Thapar, R., 1966, *A History of India*, Vol. I (Middlesex: Penguin).

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y-S., Lavergne, L., and Wallace, D. C., 1994, Mitochondrial DNA 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. *American Journal of Human Genetics*, **55,** 760–776.

Address for correspondence: Partha P. Majumder, Anthropology & Human Genetics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India. Email: ppm@isical.ac.in

Zusammenfassung. *Hintergrund*: Die Polymorphismen bei der mitochondrialen DNA haben sich für das Studium der genetischen Beziehungen und Ursprünge als brauchbar erwiesen. Die Ursprünge von Kasten haben bis jetzt ein Rätsel bestehen lassen, und die genetischen Beziehungen zwischen den verschiedenen Kasten sind in den geographischen Regionen Indiens nicht einheitlich.
Ziel: Die vorliegende Studie wurde durchgeführt, um die Art und das Ausmaß der mtDNA-Variationen und der Beziehungen zwischen verschiedenen Kasten im westindischen Staat Maharashtra zu erfassen und die Verbindungen der Ergebnisse mit deren Ursprüngen zu untersuchen.
Material und Methoden: Jeweils eine Population der drei verschiedenen Kasten wurde ausgewählt, und Blutproben wurden von vorher informierten Personen genommen, die sich zu einer Untersuchung bereit erklärten und keine verwandtschaftlichen Beziehungen untereinander aufwiesen. Die nach der Ranghöhe ausgewählten Kasten waren folgende: 1. hochgestellte Kaste (Brahmin: n = 31); 2. mittlere Kaste (Maratha; n = 41) und 3. untere Kaste (Nava Baudh; n = 40). Zehn relevante Restriktionslängen-Polymorphismen (RSP) und ein Insertion/Deletions-Polymorphismus (InDel) wurden untersucht. Das hypervariable Segment 1 (HVS 1) wurde in einer Untergruppe der Stichprobe sequenziert.
Ergebnisse: Vier RSP-loci wurden in allen Populationen als monomorph gefunden. Der InDel-locus war in zwei Populationen monomorph (Brahmin und Maratha). Ein Haplotyp (aufbauend auf der Basis von RSP's) wurde in allen Populationen als prädominant identifiziert. Die Haplotypen-Verschiedenheit zeigte sowohl bei den Maratha (68%) als auch bei den Nava Baudh (64%) ähnliche Größenordnungen, war aber bei den Brahmin (49%) deutlich größer. Die Frequenz der Haplogruppe M wurde bei allen drei Populationen in einem hohen Prozentsatz gefunden. Aber entgegen aller Erwartungen, existierte die höchste Frequenz in der Oberkaste der Brahmin. Jedoch etwa 10% der Untersuchten der Brahmin-Kaste besitzen die Haplogruppe C. Eine extreme Variation wurde in der HVS 1- Region gefunden. Die Mannigfaltigkeiten des Nukleotids und die mittlere Anzahl von Ungleichheiten wurden bei allen drei Kasten in ähnlichem Ausmaß gefunden.
Schlußfolgerungen: Die Oberkaste der Brahmin ist gegenüber der Mittel- und der Unterkaste genetisch verschieden. Aber mit Blick auf die höchste Frequenz der Haplogruppe M innerhalb der Brahmin scheint es, daß es dort eine Rekrutierung von anderen Populationen in diese Gruppe gegeben hat.

Résumé. *Arrière-plan*: Les polymorphismes de l'ADN mitochondrial ont fait la preuve de leur utilité pour l'étude des origines et des parentés génétiques. Les origines des populations castées de l'Inde sont restées énigmatiques et les apparentements génétiques des castes ne sont pas uniformes d'une région géographique à une autre.
But: Cette étude a été entreprises pour analyser la nature et l'étendue des variations d'ADNmt ainsi que les relations des groupes castés de l'état indien occidental du Maharashtra et pour examiner ce que les résultats impliquent quant à leurs origines.
Sujets et méthodes: Une population a été sélectionnée dans chacune des trois castes et des échantillons sanguins ont été obtenus de sujets non apparentés pleinement informés du but de l'étude. Les populations ordonnées en castes étaient : le rang supérieur (Brahmane; n = 31), rang moyen (Maratha; n = 41) et le rang inférieur (Nava Baudh; n = 40). On a étudié les polymorphismes de dix sites de restriction (RSPs) et d'un locus d'insertion/délétion (InDel). Le segment hypervariable 1 (SHV1) a été séquencé à partir d'un sous-ensemble des individus échantillonnés.
Résultats: Quatre loci RSPs ont été trouvés monomorphiques dans les trois populations. Le locus InDel est monomorphique dans deux d'entre-elles (Brahmane et Maratha). Un haplotype construit sur la base des RSPs est prédominant dans les trois groupes. La diversité haplotypique est de magnitude similaire chez les Maratha et les Nava Baudh (respectivement 68% et 64%), nettement plus élevée que chez les brahmanes (49%). La fréquence de l'haplogroupe M est élevée dans les trois groupes, mais contrairement à ce qui était attendu, elle est plus haute chez les brahmanes. Près de 10% des brahmanes cependant portent l'haplogroupe C. Une large variation a été trouvée dans la région du SHV1. Les diversités nucléotidiques et le nombre moyen de non appariements, ont une magnitude semblable dans les trois groupes.
Conclusions: La caste supérieure Brahmane est génétiquement distincte des castes moyenne et inférieure; cependant, sa fréquence plus élevée de l'haplogroupe M, laisse penser qu'elle aurait pu recevoir des apports de population externes.

Resumen. *Antecedentes*: Los polimorfismos del ADN mitocondrial han demostrado ser útiles en el estudio de los orígenes y de las relaciones genéticas de las poblaciones. Los orígenes de las castas continúan siendo un enigma y las relaciones genéticas entre los grupos estratificados en castas no son uniformes a través de las distintas regiones geográficas de la India.
Objetivo: Este estudio se emprendió para investigar la naturaleza y la magnitud de la variación del ADNmt y las relaciones entre los grupos de castas del estado occidental indio de Maharashtra, y para examinar la implicación de los resultados sobre sus orígenes.
Sujetos y métodos: Se seleccionó una población de cada una de las 3 categorías de castas y las muestras de sangre se obtuvieron previo consentimiento informado a partir de individuos no emparentados. Las categorías de casta fueron: alta (Brahmin; n = 31), media (Maratha; n = 41) y baja (Nava Baudh; n = 40). Se estudiaron 10 polimorfismos de dianas de restricción (Restriction Site Polimorphisms, RSPs), de interés

para el objetivo del estudio, y un polimorfismo de Inserción/Deleción (InDel). El Segmento Hipervariable 1 (HVS1) fue secuenciado en un subgrupo de los individuos muestreados.
Resultados: Se encontró que cuatro loci RSP eran monomórficos en todas las poblaciones. El locus InDel era monomórfico en dos poblaciones (Brahmin y Maratha). Un haplotipo, construido sobre la base de los RSPs, se encontró que era predominante en todas las poblaciones. La diversidad haplotípica fue de magnitud similar entre los Maratha y entre los Nava Baudh (68% y 64%, respectivamente), y fue mucho más elevada que entre los Brahmin (49%). Se observó que la frecuencia del haplogrupo M era alta en los tres grupos, pero, contrariamente a lo esperado, fue más elevada en la casta alta Brahmin. Cerca del 10% de los Brahmins poseía, no obstante, el haplogrupo C. Se encontró una amplia variación en la región HVS1. Las diversidades nucleotídicas y el número medio de emparejamientos erróneos fueron de magnitud similar en los tres grupos analizados.
Conclusiones: La casta alta, Brahmin, es genéticamente diferente de las castas media y baja. Sin embargo, en vista de la elevada frecuencia de haplogrupo M entre los Brahmin, parece que ha podido haber una incorporación de individuos de otras poblaciones a este grupo.