Hum Genet (2004) 115: 221–229
DOI 10.1007/s00439-004-1130-x

ORIGINAL INVESTIGATION

Stephen Wooding · Christopher Ostler · B. V. Ravi Prasad · W. Scott Watkins · Sandy Sung · Mike Bamshad · Lynn B. Jorde

Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data

Received: 15 December 2003 / Accepted: 22 March 2004 / Published online: 1 July 2004

Abstract Genetic, ethnographic, and historical evidence suggests that the Hindu castes have been highly endogamous for several thousand years and that, when movement between castes does occur, it typically consists of females joining castes of higher social status. However, little is known about migration rates in these populations or the extent to which migration occurs between caste groups of low, middle, and high social status. To investigate these aspects of migration, we analyzed the largest collection of genetic markers collected to date in Hindu caste populations. These data included 45 newly typed autosomal short tandem repeat polymorphisms (STRPs), 411 bp of mitochondrial DNA sequence, and 43 Y-chromosomal single-nucleotide polymorphisms that were assayed in more than 200 individuals of known caste status sampled in Andrah Pradesh, in South India. Application of recently developed likelihood-based analyses to this dataset enabled us to obtain genetically derived estimates of intercaste migration rates. STRPs indicated migration rates of 1–2% per generation between high-, middle-, and low-status caste groups. We also found support for the hypothesis that rates of gene flow differ between maternally and paternally inherited genes. Migration rates were substantially higher in maternally than in paternally inherited markers. In addition, while prevailing patterns of migration involved movement between castes of similar rank, paternally inherited markers in the low-status castes were most likely to move into high-status castes. Our findings support earlier evidence that the caste system has been a significant, long-term source of population structuring in South Indian Hindu populations, and that patterns of migration differ between males and females.

S. Wooding · C. Ostler · W. S. Watkins · S. Sung · M. Bamshad · L. B. Jorde (✉)
Department of Human Genetics, University of Utah,
15 North 2030 East,
Salt Lake City, UT 84112-5330, USA
e-mail: lbj@genetics.utah.edu
Tel.: +1-801-5854566
Fax: +1-801-5859148

B. V. R. Prasad
Department of Anthropology, Andhra University,
Visakhapatnam, Andhra Pradesh,
India

B. V. R. Prasad
Anthropological Survey of India,
Calcutta, India

M. Bamshad
Department of Pediatrics, University of Utah,
Salt Lake City, UT, USA

Introduction

The Hindu caste system plays a major role in the social and economic organization of nearly one-sixth of the world's population. In the caste system, individuals are born into one of five *varnas*, which are ranked in status from low to high: *Panchama*, *Sudra*, *Vysya*, *Kshatriya* and *Brahmin*. Historically, caste has been an important determinant of access to education and occupational opportunity. Caste has also been an important determinant of marital choice. Marriage between partners of equal status is preferred, and reproduction in the caste system is largely endogamous (Beals 1962; Heinz 1999).

Ethnographic and genetic evidence both support the suggestion that Hindu castes have been highly endogamous for a considerable length of time (Bamshad et al. 2001; Karve 1968; Misra 2001). Although the level of genetic differentiation between castes is relatively small, genetic distances observed in several studies suggest that gene flow is limited (Bamshad et al. 1998, 2001; Bhattacharyya et al. 1999; Char et al. 1989; Dutta et al. 2002; Lakshmi et al. 2002; Papiha et al. 1996). However, some intercaste migration does occur. Caste distinctions are blurred in many modern communities, especially in urban areas, where intercaste marriages are more common (Heinz 1999). Intercaste marriages are also permitted by Hindu doctrine under some circumstances. The concept of *anuloma*, for example, permits some men to marry a woman of lower caste, while women are forbidden from

marrying a man of lower caste (Misra 2001). The latter mechanism of mobility between castes is a point of particular interest because it may explain differences in diversity observed between maternally and paternally inherited genetic markers. Previous studies have found that castes are more similar with respect to maternally inherited markers than with respect to paternally inherited markers, as would be expected if females marry across caste boundaries more often (Bamshad et al. 1998).

The varying means of gene flow across caste boundaries raise basic questions about the long-term structure of the caste system and its impact on the historical demography of India. Given earlier evidence for low but significant movement between castes, one would like to know: (1) What are the rates of gene flow between Hindu castes? (2) How do males and females differ in their migrational patterns? (3) How are sex differences in migration related to broader patterns in the caste hierarchy? Inferences about these aspects of intercaste migration are potentially valuable as a source of insight into the impact of long-term, culturally-imposed subdivision on one of the world's major populations. Genetic data are ideal for answering questions about long term population structure in these populations. However, the close relationships among caste groups have been difficult to resolve using small datasets and standard statistical tools.

In this study, we used a newly assembled dataset of 45 STRP markers and recently developed likelihood analyses based in coalescence theory to test hypotheses about intercaste migration. By comparing patterns of genetic diversity in maternally, biparentally, and paternally inherited markers, we were able to estimate migration rates between low-, middle-, and high-status castes. We used these estimates to address two hypotheses about general and sex-specific patterns of intercaste gene flow. First, we tested the hypothesis that rates of migration between groups of similar status in the caste hierarchy (i.e., between high- and middle-, and middle- and low-status castes) have historically been greater than rates of migration between groups of dissimilar status (i.e., between high- and low-status castes). This hypothesis is relevant to understanding the basic properties of gene flow between the castes. Second, we tested the hypothesis that patterns of intercaste migration in maternally inherited markers differ from those of paternally inherited markers. This hypothesis is relevant to the proposition that females in the caste system are "upwardly mobile" in status, while males are not. We found that while some aspects of migration in our sample are consistent with longstanding assumptions about mechanisms of social mobility in Hindu society, others are not.

Methods

Population samples

DNA samples were collected from Telegu-speaking males in the vicinity of Visakhaptnam, Andhra Pradesh, India, as part of an ongoing study of the genetics of the caste system (Bamshad et al. 1996, 1998, 2001; Jorde et al. 2002; Watkins et al. 1999, 2003). These samples were collected and analyzed with the approval of the government of India, Andhra University, and the Institutional Review Board of the University of Utah. Subjects were asked about their caste affiliations, surnames, and parental birthplaces. Only individuals unrelated to each other by at least three generations were included. The sample is described in detail by Bamshad et al. (2001).

Caste populations were grouped based on their traditional ranking by *varna* (Tambia 1973), as shown in Table 1. Historically, four *varnas* were recognized: *Brahmin*, *Vysya*, *Kshatryia*, and *Sudra*. By tradition, these *varnas* held different occupations. *Brahmins* were the priestly class, *Vysya* were merchants, *Kshatryia* were warriors, and *Sudra* served the other three *varnas* (Tambia 1973). A fifth *varna*, the *Panchama* (which includes the so-called untouchables), was added at a later date (Bamshad et al. 2001). For analysis, we divided these groups into three status categories: high (*Brahmin*, *Kshatryia*, and *Vysya*), middle (*Sudra*, including *Kapu* and *Yadava*), and low (*Panchama*, including *Madiga*, *Mala*, and *Relli*). This ranking scheme has been used elsewhere (Krishnan and Reddy 1994; Lakshmi et al. 2002).

Laboratory procedures

Mitochondrial DNA (mtDNA) data consisted of 411 bp HVS-I sequences, corresponding to positions 16,000 through 16,410 in the corrected Cambridge reference mitochondrial sequence (Andrews et al. 1999). These sequences were determined by fluorescent Sanger sequencing using an American Biosystems Incorporated (ABI) 377 automated DNA sequencer. Sequence data were analyzed using ABI DNA analysis software and Genecodes SEQUENCHER software.

Forty-five STRPs were also assayed in the sampled population. The Genome Database (http://www.gdb.org) accession numbers for these systems are: UT1091, UT1201, UT1205, UT1220, UT1227, UT1228, UT1232, UT1239, UT1243, UT1257, UT1313, UT1352, UT1357, UT1376, UT1674, UT1708, UT1740, UT1747, UT1880, UT1885, UT1917, UT1950, UT1985, UT2021, UT2081, UT2092, UT2127, UT2203, UT5022, UT5027, UT5029, UT5030, UT5033, UT5048, UT5492, UT6507, UT6516, UT6540, UT7131, UT8067, UT868, UT871, UT901, UT919, VWFII. These systems were chosen because they are spread broadly across the human genome and thus provide nearly independent loci for statistical analysis.

Y-chromosomal data consisted of 43 biallelic sequence polymorphisms on the nonrecombining portion of the Y chromosome reported by Hammer et al. (2001). These data were composed of previously published and new polymorphisms that were detected using a combination of single-stranded conformation polymorphism (SSCP) and denaturing high-performance liquid chromatography (DHPLC). New polymorphisms were identified in ascertainment panels of 20 and 57 individuals and then screened in 2,858

Table 1 Hierarchical grouping of sampled castes, and their sample sizes

Rank	Varna	Caste	Sample size		
			mtDNA	STRP	Y
High	Brahmin	Brahmin	48	59	58
	Kshatryia	Kshatryia	10	11	–
	Vysya	Vysya	10	9	–
Middle	Sudra	Kapu	52	58	56
	Sudra	Yadava	48	53	51
Low	Panchama	Madiga	28	29	28
	Panchama	Mala	25	26	26
	Panchama	Relli	20	19	19
Total			241	264	238

individuals drawn from populations distributed across Africa, Asia, Europe and North America. This sampling strategy was sufficient to identify the vast majority of polymorphic sites, but did result in a slight ascertainment bias against rare alleles (Hammer et al. 2001). Ascertainment bias can have strong effects on some population genetic analyses. However, such bias is ordinarily small when ascertainment panels are large (Kuhner et al. 2000). Although ascertainment bias may exist, the large size of the panels used for SNP identification in our sample suggests that such bias, if any, is minimal. Furthermore, a slight bias against rare alleles would have little effect on the gene flow rates estimated in this study.

Statistical analysis

Levels of population substructuring were measured using F_{ST} and R_{ST}, which measure the departure of genetic variation from expectation under panmixia. Pairwise F_{ST} values were calculated for the mtDNA and Y-chromosome data using the method of Slatkin (1991). Pairwise R_{ST} values were calculated for the STRP data (Slatkin 1991, 1995).

To determine whether maternally and paternally inherited markers showed similar overall levels of population substructuring, F_{ST} values were calculated using high-, middle-, and low-status caste groups as the unit of subdivision. The null hypothesis that the observed F_{ST} values equal zero was tested using a bootstrap resampling procedure in which simulated populations were generated by randomly resampling from the observed data. One-thousand bootstrap replicates were used for each analysis.

Patterns of migration were analyzed using the MIGRATE software package (Beerli 2001). MIGRATE uses likelihood methods based in coalescence theory to simultaneously estimate migration rates and effective population sizes for a set of subpopulations (Beerli 2001; Beerli and Felsenstein 1999, 2001). This method was applied to the mtDNA sequences, autosomal STRPs and Y-chromosomal SNPs separately in order to estimate effective population sizes and migration parameters for maternally, paternally, and biparentally inherited markers. Each run of the MIGRATE program used 10,000 burn-in iterations, 20 long chains (5,000 iterations each), 40 short chains (500 iterations each). Final parameter estimates were obtained by combining the estimates of the 20 long chains. These options are described in detail by Beerli (2001)

Two models of migration were analyzed: a restricted model and an unrestricted model, shown in Fig. 1. Under the unrestricted model, migration was allowed between any pair of caste groups (high-, middle-, and low-status). Under the restricted model, migration was only allowed between caste groups with similar status (high- and middle-status castes or middle- and low-status castes).

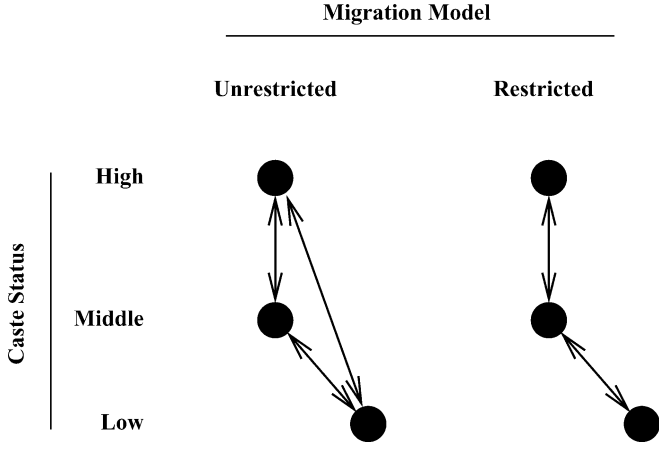

Fig. 1 Migration models used for parameter estimation and hypothesis testing. In the unrestricted model, migration is allowed between any pair of caste groups. In the restricted model, migration is only allowed between caste groups of adjacent status: high- and middle-status castes, and middle- and low-status castes

Inferred migration rates were measured as $2N_Fm_F$ for mtDNA, $2N_Mm_M$ for Y-chromosome data and $4N_em$ for STRPs. N_M, N_F and N_e are the effective sizes of the male, female, and combined populations receiving the migrants, respectively; m_M, m_F and m are the proportions of the male, female, and combined populations that are composed of new migrants, respectively (Beerli 2001). Rates inferred from the STRP data assumed that the sex ratio is 1:1 and that males and females migrate at equal rates.

The probability of a single, random migrant moving from one caste group into another was calculated from the estimated migration rates by dividing each rate by the θ (=$2N_F\mu$ for mtDNA, $2N_M\mu$ for Y-chromosomes, and $4N_e\mu$ for STRPs, where μ is the infinite sites mutation rate for mtDNA and Y-chromosomes, and the infinite alleles mutation rate for STRPs) of the receiving population and then normalizing the resulting distribution (see documentation for the MIGRATE computer program; Beerli 2001). The probability of a given migrant moving upward in caste status was calculated by summing the probabilities of moving from the low- to middle-, low- to high-, and middle- to high-status castes. Similarly, the probability of a given migrant moving downward in caste status was calculated by summing the probabilities of moving from the high- to middle-, high- to low-, and middle- to low-status castes.

The null hypothesis that patterns of migration differ between males and females was tested in three steps. First, maximum likelihood demographic parameters were estimated separately for the mtDNA and Y-chromosome datasets under the unrestricted migration model. The likelihood of the Y-chromosome data was then calculated under the maximum likelihood parameters of the mtDNA dataset, and the likelihood of the mtDNA data was calculated under the maximum likelihood parameters of the Y-chromosome data set. Finally, each pair of likelihoods was compared in a likelihood ratio test by rerunning the MIGRATE program using one set of migration parameters as the restricted model, as described in Beerli (2001).

The hypothesis that migration rates between castes of dissimilar rank are zero was tested by estimating migration rates under the constrained model and then comparing this likelihood with the likelihood of the data under the unconstrained model. These tests were performed separately on each data set.

Results

F_{ST} and R_{ST} values showed two main features (Table 2). First, the genetic distances between high- and middle-status and middle- and low-status caste groups were generally smaller than the distances between high- and low-status groups. However, in the Y-chromosome data, the distance between high- and low-status castes was smaller than the distance between the high- and middle-status castes. Some negative genetic distances were observed. Such distances can occur by chance when distances are close to zero.

Table 2 Genetic distances. The *mtDNA* and *Y-chromosome* columns show F_{ST} values; the *STRP* column shows R_{ST} values

Between	mtDNA	Y chromosome	STRP
High—Middle	−0.0015	0.0760	0.0022
High—Low	0.0076	0.0670	0.0066
Middle—Low	0.0046	−0.0042	0.0012

The overall F_{ST} value estimated from the Y-chromosome data, 0.048, was more than ten times greater than the F_{ST} observed for the mtDNA (0.0038) data and the R_{ST} for the STRP data (0.0042). Further, bootstrap analyses showed that the F_{ST} value for the Y-chromosome data was significantly greater than zero ($P<0.01$), but the F_{ST} value for the mtDNA data and the R_{ST} value for the STRP data were not ($P>0.05$). Some of the difference between these F_{ST} values might be explained by higher levels of homoplasy (due to back mutation) in mtDNA and STRP data than in the Y-chromosome data. Low F_{ST} values are often observed in genetic systems with high mutation rates. However, an estimate of F_{ST} in 100 *Alu* inserts, typed in our sample by (Watkins et al. 2003), was 0.0082, just two times higher than in mtDNA and STRPs. *Alu* elements are known to evolve slowly, and have very low levels of homoplasy. Thus, homoplasy is unlikely to account for the tenfold difference in F_{ST} values observed here.

The inferred θ values for the mtDNA and Y-chromosome datasets were proportioned similarly across caste groups, as shown in Table 3. The similar proportions of these values suggest that population structures within each data type are roughly similar. For example, the mtDNA and Y-chromosome data both implied much greater θ values in the low-status castes than in the middle- or high-status castes. The STRP data implied a smaller θ in the low-status castes than in the middle- and high-status caste group. All of these θ values should be interpreted with caution because population substructuring is known to occur within caste groups. For example, the low-status caste group includes three castes (*Madiga*, *Mala*, and *Relli*) which may themselves be subdivided. Such subdivision could account for the large θ value observed in the low-status castes. Similarly, the lack of Y-chromosomal data for the Kshatryia and Vysya castes might account for the very small θ in the high-status caste group. Little is known about patterns of subdivision at the level of the caste, and we know of no evidence to suggest that low-status castes are likely to be more subdivided than are high-status castes. Our findings suggest that such subdivision may be an interesting avenue for further investigation.

The migration rates in Table 3 are not directly comparable across genetic systems because they are given in units of $2N_Fm_F$ and $2N_Mm_M$ for mtDNA and Y-chromosomes, and $4N_em$ for STRPs, where m is the proportion of the receiving population that is composed of migrants. The migration parameter, m, was difficult to isolate for the Y-chromosome data because N_M is unknown. In principle, N can be obtained from the θ values in Table 3, which equals $2N_F\mu$, where μ is the mutation rate. However μ is uncertain for the Y-chromosome data because the number of nucleotides that were assessed in obtaining the polymorphisms is unknown (Hammer et al. 2001). For this reason, we compared migration rates for mtDNA and Y-chromosomes under the assumption that the N values of mtDNA and Y-chromosomes are equal, which is roughly equivalent to assuming that the effective sizes of the male and female breeding populations are the same. Under this assumption, the migration rates inferred for mtDNA would be substantially greater than for Y-chromosomes, a finding in agreement with the disparate F_{ST} values of the two systems.

Table 3 Migration rates and θ-values. Migration rates are expressed as $2N_Fm_F$ and $2N_Mm_M$ for the mtDNA and Y-chromosome data, where N is the effective population size of the population receiving the migrants and m is the proportion of the receiving population that is composed of migrants. Migration rates are expressed as $4N_em$ for the STRP data

To	From			θ
	High	Middle	Low	
mtDNA				
High		15.55	20.61	0.12
Middle	264.66		1128.19	0.11
Low	1034.31	1660.41		1.44
STRP				
High		25.75	20.15	0.73
Middle	22.63		22.02	0.87
Low	18.34	22.53		0.54
Y chromosome				
High		9.39	40.11	0.01
Middle	17.22		5.93	0.01
Low	13.82	815.65		0.22

The migration parameter, m, was also difficult to determine for the mtDNA data. Multiple runs of the MIGRATE program showed that parameter estimates for both the mtDNA and Y-chromosome data had high variances, most likely because they are based on sets of linked polymorphisms obtained from single loci. While this variance was taken into account in the likelihood ratio tests comparing these systems, and the parameters could thus be compared, migration rates could not be determined precisely. Parameter estimates obtained from the STRP data, which were based on 45 independently segregating loci, had much lower variances. For this reason, and because estimates of μ are available for STRPs, these loci provided direct estimates of m.

Estimates of mutation rates in STRPs vary, with typical values ranging from 10^{-4} to 10^{-3} mutations per locus per year (Leopoldino and Pena 2002; Weber and Wong 1993). Higher rates have been reported for some loci (Brinkmann et al. 1998; Leopoldino and Pena 2002). Under the assumption that the mutation rate for STRPs is $10^{-3.5}$ per locus per generation, our data suggest that, historically, 1–2% of any caste group changed in status each generation, as shown in Table 4. The total fraction of the population that changed in caste status was estimated to be 1.94% per generation. This estimate ranged from 0.6% under the assumption that the mutation rate is 10^{-4} per locus per year to 6.0% under the assumption that the mutation rate is 10^{-3} per locus per year.

The small effective population sizes estimated from STRPs (577.12, 687.80, and 426.91 for the high-, middle-, and low-status castes, respectively) are not unexpected

Table 4 Effective migration rates and effective population size in STRPs. Note that just as the effective population size is substaintially lower than the census population size, the effective number of migrants is substantially lower than the actual number. These values assume that $\mu=10^{-3.5}$ mutations per locus per year

To	From			N_e
	High	Middle	Low	
A[a]				
High		0.011	0.0087	577.12
Middle	0.0082		0.0080	687.80
Low	0.011	0.013		426.91
B[b]				
High		6.44	5.03	577.12
Middle	5.66		5.51	687.80
Low	4.59	5.63		426.91

A) Migration rates expressed as the fraction of the receiving population that is composed of migrants
B) Migration rates expressed as the effective number of individual migrants

given previous estimates of human effective population sizes. N_e values in natural populations are usually much smaller than census population sizes. For example, typical estimates of N_e in human populations sampled across Africa, Asia, and Europe fall around 10,000 (Jorde et al. 2000; Tishkoff and Verrelli 2003). Thus, the estimates of N_F, N_M, and N_e for the caste populations are reasonable.

The probabilities of migration between pairs of caste groups are shown in Table 5. These probabilities, which are conditioned on the movement of a single, random migrant, reflect the probability that a random migrant moved from one caste group into another. For example, the probability that a random mtDNA migrant originated in the high-status castes and moved into middle-status castes was 0.16, as shown in Table 5. These probabilities are consistent in several respects with the genetic distances outlined in Table 2. For example, the small genetic distance between the low- and middle-status castes inferred from mtDNA is consistent with the high probability of migration between these two groups inferred from mtDNA.

Table 5 Migration probabilities. Each probability is conditioned on the occurrence of a single migration event. For example, the probability that a randomly chosen migrant moves from the high-status castes into the middle-status castes is 0.16. Probabilities were calculated as described in Methods

To	From		
	High	Middle	Low
mtDNA			
High	–	0.0087	0.012
Middle	0.16	–	0.69
Low	0.048	0.078	–
STRP			
High	–	0.19	0.15
Middle	0.18	–	0.13
Low	0.14	0.22	–
Y chromosome			
High	–	0.085	0.36
Middle	0.16	–	0.054
Low	0.0057	0.34	–

The hypothesis that males and females have the same patterns of migration was rejected by likelihood ratio tests. The maximum likelihood migration parameters inferred from the mtDNA data did not fit the observed Y chromosome data ($P<0.01$). Likewise, the maximum likelihood migration parameters inferred for the Y chromosome data did not fit the observed mtDNA data ($P<0.01$). Under the assumption that male and female effective population sizes were the same within each caste group, the migration rates estimated from the mtDNA data were substantially higher than those implied by the Y chromosome data, as shown in Table 3.

The hypothesis that migration takes place only between castes of similar rank (i.e., between high- and middle-status, and middle- and low-status castes) was rejected for all three genetic systems, indicating that levels of migration between high- and low-status castes are significantly greater than zero. As shown in Table 5, migration probabilities were usually higher between high- and middle- and between middle- and low-status caste groups than between high- and low-status caste groups. However, the probability of migrating from low- to high-status castes slightly exceeded that of migration from low- to middle-status castes in STRPs (0.15 vs 0.13) and Y-chromosome data (0.36 vs 0.054).

Discussion

Genetic variation among Indian caste populations has long been a topic of interest (Dobzhansky 1973; Chakraborty et al. 1977; Dobzhansky 1973; Roychoudhury 1982). Because castes are highly endogamous, and because men and women of different castes may be more likely to marry if their castes are of similar social status, it is reasonable to hypothesize that inter-caste genetic distances may be correlated with similarity in caste status. Indeed, such correlations have been observed in some studies (Bamshad et al. 1998, 2001; Char et al. 1989; Dutta et al. 2002; Lakshmi et al. 2002; Papiha et al 1996). However, genetic studies show that 95% or more of genetic variation occurs within caste populations (Bamshad et al. 2001; Majumder 1998; Papiha et al. 1996), and genetic variation is strongly affected by factors such as language and geographic distance. For example, Indo-European speakers are thought to have entered India from the north, gradually percolating southward. This is expected to produce a north-south cline in gene frequencies, a pattern that is frequently observed (Majumder 1998; Malhotra and Vasulu 1993; Passarino et al. 1996). The substantial influence of geographic distance has led some investigators to conclude that geography is a major factor affecting genetic variation in India, while caste affiliation has little, if any, influence (Chakraborty et al 1986; Majumder 1998, 2001; Mukherjee et al. 1999). To the extent that caste rank may affect genetic similarity between populations, how-

ever, it is likely to be confounded with geographic location. Thus, to assess whether caste rank is correlated with genetic distance, the effect of geography should be controlled by comparing castes from the same geographic locale (Bamshad et al. 1998; 2001).

Previous analyses of migration patterns in the Hindu castes have focused on two principal trends. First, statistically significant genetic distances are found between castes with different ranks (Bamshad et al. 2001). This finding is consistent with earlier evidence for genetic differentiation among caste groups (Bamshad et al. 1998; Bhattacharyya et al. 1999; Char et al. 1989; Dutta et al 2002; Lakshmi et al. 2002; Mountain et al. 1995; Papiha et al. 1996), as well as with the ethnographic observation that intercaste marriages are rare (Heinz 1999). Second, intercaste genetic distances are greater for the Y-chromosome data than for the mtDNA data, and genetic distance is correlated with difference in caste status in mtDNA, but not in Y chromosomes (Bamshad et al. 1998). Other studies have also documented limited sharing of Y-chromosome haplotypes among castes (Bhattacharyya et al. 1999). These observations are consistent with the ethnographic observation that males are occasionally allowed to marry partners of lower caste status, with the offspring of such marriages entering the paternal caste (Misra 2001). Thus, the findings of previous genetic studies are in overall agreement with the basic ethnographic features of the caste system.

Our new data, which provide the most extensive survey to date of genetic variation in a Hindu caste population, confirm the presence of trends identified previously. First, genetic distances between high- and low-status caste groups were generally larger than distances between these two groups, as shown in Table 2. These findings are consistent with earlier observations (Bamshad et al. 1998, 2001; Bhattacharyya et al. 1999). Second, inferred migration probabilities suggested that movement between castes of similar status occurs more often than movement between castes of dissimilar status, as shown in Table 5. Third, the proportion of genetic variance accounted for by differences between castes was more than ten times greater when measured from Y-chromosome data (F_{ST}=0.048) than when measured from mtDNA data (F_{ST}=0.0038), and only the F_{ST} value for the Y-chromosome data was significantly greater than zero. This finding gives statistical support to earlier conclusions, based on fewer data, that migration rates are substantially higher in maternally inherited genetic markers than in paternally inherited markers (Bamshad et al. 1998).

Our new data also allowed us to address three problems that have been difficult to solve previously: (1) the estimation of overall migration rates, (2) the estimation of migration rates between low-, middle-, and high-status castes, and (3) the formal statistical comparison of migration rates in maternally, paternally, and biparentally inherited genetic systems.

Migration rates

Migration rates inferred from the STRP data indicate that gene flow involves one to two percent of the population per generation. This rate is similar to rates based on ethnographic data, which range from 0.001% in some rural populations to 3% in urban populations (Malhotra and Vasulu 1993). This rate is also consistent with earlier findings that Y-chromosomal F_{ST} values among caste groups are remarkably high (Bamshad et al. 1998, 2001), in spite of the fact that haplotype sharing among castes is common (Bamshad et al. 2001; Mountain et al. 1995; Roychoudhury et al. 2000). The low estimates of intercaste migration obtained from both ethnographic and genetic data testify to the importance of the caste system in shaping the population history of India.

Migration probabilities

In all three genetic systems, migrants leaving the high-status castes showed a greater probability of entering the middle-status castes than entering the low-status castes. Similarly, in all three systems, migrants leaving the middle-status castes showed a greater probability of entering the low-status castes than the high-status castes. However, the transition probabilities of migrants leaving the low-status castes differed strikingly between mtDNA and Y chromosome systems. While mtDNA migrants leaving the low-status castes were much more likely to move into the middle-status castes than into the high-status castes, the opposite was true for Y-chromosome migrants, as shown in Table 5. This trend was emphasized by the fact that among the genetic distances observed, only the Y-chromosome data showed smaller distances between high- and low-status caste groups than between high- and middle-status groups, as shown in Table 2.

The migration probabilities inferred from the STRP data (shown in Table 5) were similar in many respects to those estimated for the mtDNA and Y-chromosome data. For example, in STRPs the probability of migration from low- to high-status castes exceeded that of migration from low- to middle-status castes, but by only a slight margin. This pattern may reflect the countervailing effects of low probabilities of migration from low- to high-status castes by mtDNA and high probabilities of migration from low- to high-status castes by Y chromosomes. The tendency of migrants to move into castes of similar status is consistent with the observation that distinctions are sometimes unclear among castes of similar status (Malhotra and Vasulu 1993).

Male–female differences

Genetic and ethnographic data have long suggested that rates of male and female migration differ among caste groups. However, the details of these differences have remained elusive. Here, a comparison of maternally and

paternally inherited markers revealed that migration patterns are significantly different. As shown in Tables 3, 4, and 5, these migration matrices differed in two important ways. First, on average, maternally inherited markers increased in caste status, but paternally inherited markers did not; while an estimated 70% of mtDNA migrants moved upward in status, only 49% of Y-chromosome migrants moved upward in status. Thus, while the net movement of maternally inherited genes is biased strongly upward, the net movement of paternally inherited genes is near zero. Second, the single most likely migration probability for paternally inherited markers was from the low-status caste group into the high-status caste group. The higher probability of migration from low- to high-status castes for Y chromosomes does not contradict the hypothesis that females move upward in caste status more often than males. For instance, it could be that when males do leave the low-status castes they have a greater probability of entering the high-status castes than do comparable females, but that females migrate between castes more often than males and thus move upward in status more often overall.

Caste changes have traditionally been discouraged in Hindu society, so little is known about the mechanisms through which changes in caste occur. However, migration rates are affected at least partly by recognized mechanisms of movement across caste boundaries. The Hindu concept of *anuloma*, for instance, occasionally permits men to marry women of lower caste, while women are rarely permitted to marry men of lower caste (Misra 2001). This mechanism of movement implies that females should be more "upwardly mobile" than males, and it has been suggested that *anuloma* is the source of greater female intercaste mixing inferred from genetic data (Bamshad et al. 1998). The net upward migration of maternally inherited markers shown in Table 5 provides statistical support for this suggestion. Many other potential mechanisms of migration are likely to be difficult to detect. For example, the organization of the caste hierarchy is known to vary somewhat among geographic regions (Malhotra and Vasulu 1993); thus, migration from one region to another might also confer slight changes in caste status. Child adoption would provide another means of migration. The migration of genes among caste groups would not even require changes in the caste affiliation of individuals; such migration could arise as the product of unrecognized unions between individuals in different castes.

Effects from outside populations

The hypothesized origin of the Hindu castes has some bearing on the interpretation of the patterns of migration inferred here. Earlier analyses of mtDNA, Y chromosomes and *Alu* polymorphisms in the castes suggested that the population of contemporary India is the product of two primary influences: a proto-Asian origin and subsequent Indo-European introgression (Bamshad et al. 2001; Basu et al. 2003). These analyses found that the modern inhabitants of India show close affiliations with other Asians, but also show affiliations with West Eurasians. The similarity between the castes and West Eurasians is greatest for Y chromosome polymorphisms in the high-status castes, a pattern consistent with the introgression of early West Eurasian migrants (especially males) preferentially into the high-status castes, possibly as a means of establishing economic and political control.

The complex origins of Hindu India could have a residual impact on the patterns of migration inferred here. For example, the simultaneous introgression of Y chromosomes into all three status levels could lead to the inference of migration between castes when migration in reality occurred between castes and outside sources. The preferential migration of genes into castes of just one status could cause a similar bias by affecting the effective population size of the castes receiving migrants. However, the low migration rates we observe, and their general concordance between males and females, are inconsistent with the effects that would be expected with a strong influence from outside populations. For instance, the high migration rates we infer from low- to high-status castes in males are inconsistent with a major introgression of Y chromosomes into the upper castes from outside sources, because such introgression would reduce the migration rate rather than increase it. Nonetheless, complicated caste-specific demographic trends, such as migration, growth, and subdivision, must be recognized as a possible factor that has affected our data.

Geographical variation in caste structure

In this study, our sample was limited to individuals residing in the province of Andhra Pradesh, India. An important question is whether the trends we have identified are present in other regions of India. Majumder and colleagues (Majumder 2001; Majumder et al. 1999) have presented useful data on a number of Indian caste and tribal populations, including three caste populations from the same location in Uttar Pradesh (Brahmins, Rajputs, and Chamars, described by the authors as upper, middle, and lower caste representatives, respectively). These populations provide a convenient comparison with the populations analyzed in the present study, particularly because allele frequencies were reported for eight autosomal *Alu* insertion polymorphisms, seven mtDNA restriction site polymorphisms, and 12 Y-chromosome polymorphisms. These allele frequencies were used to generate the genetic distances shown in Table 6. As in our data, the distance between high- and low-status caste groups in the data of Majumder et al. (1999) exceeded that between high- and middle-status groups and between middle- and low-status groups in mtDNA and autosomal markers. This pattern further supports the hypothesis that caste rank and genetic distance, when controlled for geographic location, may often be correlated. However, our results from Y chromosomes differ. Majumder et al. (1999) and Majumder (2001) found that, like mtDNA and autosomal

Table 6 Genetic distances. Each row shows Nei's distance between high-, middle-, and low-status caste groups, inferred from the data of Majumder et al. (1999) and Majumder (2001). In every system, the genetic distance between the high- and low-status castes is greater than the distance between the high- and middle-, and middle- and low-status castes

Between	mtDNA	Y chromosome	Alu
High—Middle	0.0940	0.0060	0.0120
High—Low	0.6900	0.0460	0.1600
Middle—Low	0.4200	0.0400	0.0900

polymorphisms, Y-chromosomal polymorphisms suggest that distances between high- and low-status castes are greatest (Table 6). In contrast, our data suggest that the distance between high- and middle-status castes is greatest: a finding consistent with the hypothesis that Y chromosomes move most frequently between low- and high-status castes. Further studies in additional populations are needed to determine the generality of these trends.

Conclusions

Inferences from the mtDNA sequences, autosomal STRPs, and Y-chromosomal SNPs analyzed here indicate that the ancient tradition of caste-specific endogamy has allowed small, but detectable, differences to evolve between caste groups. These differences vary between maternally and paternally inherited genetic markers. While maternally inherited genes show evidence of high rates of migration and upward social mobility, paternally inherited genes show evidence of low rates of migration and no overall upward mobility. However, paternally inherited genes in the lowest castes have historically been most likely, when migrating, to move into the upper castes.

Culturally-mediated population substructuring is distinct in India, but is also present in many other human societies. For example, the occupationally defined guilds of the Aztecs bore a number of similarities to the Hindu castes (Berdan 1982). Other populations, like the Roma (Gypsies) and certain populations in Africa, appear be socially stratified as well (Gresham et al. 2001; Tuden and Plotnicov 1970). Such substructuring might have a variety of genetic consequences. Thus, the ability to assess within- and between-group patterns of variability is a potentially powerful means of uncovering the subtle organizational features of human societies. The ability to distinguish the effects of population structure on male and female populations adds another level of resolution in efforts to define the interface between the culture and the biology of human populations.

The migration patterns inferred in Hindu castes are not limited in relevance to the study of Indian populations. The patterns we observed in a limited set of South Indian populations are potentially relevant to myriad problems in human biology. Among these, the most obvious is the problem of identifying and coping with population subdivision in population genetic analyses. The presence of such substructuring is a concern of increasing importance in linkage and association studies (Pritchard and Przeworski 2001). The approach we have used demonstrates that population substructuring, including differences between male- and female-driven gene flow, can be detected and defined both in magnitude and direction. The recognition and assessment of these patterns will be of continuing importance in the analysis of human genetic variation.

References

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

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

Bamshad MJ, Fraley AE, Crawford MH, Cann RL, Busi BR, Naidu JM, Jorde LB (1996) mtDNA variation in caste populations of Andhra Pradesh, India. Hum Biol 68:1–28

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR, Rasanayagam A, Hammer MF (1998) Female gene flow stratifies hindu castes. Nature 395:651–652

Bamshad MJ, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BVR, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. Genome Res 11:994–1004

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP (2003) Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res 13:2277–2290

Beals AR (1962) Gopalpur: a South Indian village. Holt, Rhinehart and Winston, New York

Beerli P (2001) Migrate documentation and program, Version 1.1. Computer program distributed by the author, Department of Genetics, University of Washington, Seattle

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763–773

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proc Natl Acad Sci USA 98:4563–4568

Berdan FF (1982) The Aztecs of Central Mexico. Harcourt, Brace and Company, Orlando

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

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415

Chakraborty R, Chakravarti A, Malhotra KC (1977) Variation in allele frequencies among caste groups of the Dhangars of Maharashtra, India: an analysis with Wright's F_{ST} statistic. Ann Hum Biol 4:275–280

Chakraborty R, Walter H, Mukherjee BN, Malhotra KC, Sauber P, Banerjee S, Roy M (1986) Gene differentiation among ten endogamous groups of West Bengal, India. Am J Phys Anthropol 71:295–309

Char KS, Lakshmi P, Gopalam KB, Sastry JG, Rao PR (1989) Genetic differentiation among some endogamous populations of Andhra Pradesh, India. Am J Phys Anthropol 78:421–429

Dobzhansky T (1973) Genetic diversity and human equality. Basic Books, New York

Dutta R, Reddy BM, Chattopadhyay P, Kashyap VK, Sun G, Deka R (2002) Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. Hum Biol 74:33–49

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kučinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (Gypsies). Am J Hum Genet 69:1314–1331

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203

Heinz CB (1999) Asian cultural traditions. Waveland Press, Prospect Heights

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Sielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal and Y-chromosome data. Am J Hum Genet 66:979–988

Jorde LB, Bamshad MJ, Watkins WS, Ricker CE, Dixon ME, Rao BB, Prasad BVR, Naidu JM (2002) Genetic structure of South Indian caste populations: a confluence of biology and culture. In: Macbeth H, Collinson P (eds) Human population dynamics: cross-disciplinary perspective. Cambridge University Press, Cambridge

Karve I (1968) Kinship organization in India. Asia Publishing House, Bombay

Krishnan T, Reddy BM (1994) Geographical and ethnic variability of finger ridge-counts: biplots of male and female Indian samples. Ann Hum Biol 21:155–169

Kuhner M, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single-nucleotide polymorphism data for estimating population parameters. Genetics 156:439–447

Lakshmi N, Demarchi DA, Veerraju P, Rao TV (2002) Population structure and genetic differentiation among the substructured Vysya caste population in comparison to the other populations of Andhra Pradesh, India. Ann Hum Biol 29:538–539

Leopoldino AM, Pena SDJ (2002) The mutational spectrum of human autosomal tetranucleotide microsatellites. Hum Mutat 21:17–79

Majumder PP (1998) People of India: biological diversity and affinities. Evol Anthrobol 6:100–110

Majumder PP (2001) Indian caste origins: genomic insights and future outlook. Genome Res 11:931–932

Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK (1999) Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. Eur J Hum Genet 7:435–446

Malhotra KC, Vasulu TS (1993) Structure of human populations in India. In: Majumder PP (ed) Human population genetics. Plenum, New York, pp 207–233

Misra VN (2001) Prehistoric human colonization of India. J Biosci 26:491–531

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA sequence diversity. Am J Hum Genet 56:979–992

Mukherjee BN, Majumder PP, Roy B, Roy M, Dey B, Chakraborty M, Banerjee S (1999) Variation at 4 short tandem repeat loci in 8 population groups of India. Hum Biol 71:439–446

Papiha SS, Mastana SS, Purandare CA, Jayasekara R, Chakraborty R (1996) Population genetic study of three VNTR loci (D2S44, D7S22, and D12S11) in five ethnically defined populations of the Indian subcontinent. Hum Biol 68:819–835

Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS (1996) Pre-caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59:927–934

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Roychoudhury AK (1982) Genetic relationships of Indian populations. In: Malhotra KC, Basu A (eds) Human genetics and adaptation. Plenum, New York, pp 147–174

Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran N, Rani MVU, Vishwanathan H, Mitra M, Sil SK, Majumder PP (2000) Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. Curr Sci 79:1182–1192

Slatkin M (1991) Inbreeding coefficients and coalescence times. Genet Res Camb 58:167–175

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Tambia SJ (1973) The character of kinship. Cambridge University Press, Cambridge

Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genom Hum Genet 4:293–340

Tuden A, Plotnicov L (1970) Social stratification in Africa. The Free Press, New York

Watkins WS, Bamshad M, Dixon ME, Rao BB, Naidu JM, Reddy PG, Prasad BV, Das PK, Gai PB, Bhanu A, Kusuma YS, Lum JK, Fischer P, Jorde LB (1999) Multiple origins of the mtDNA 9-bp deletion in populations of South India. Am J Phys Anthrobol 109:147–158

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Batzer MA, Prasad BVR, Reddy G, Das PK, Jorde LB (2003) Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. Genome Res 13:1607–1618

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128