Coll. Antropol. **25** (2001) 2: 425–442
UDC 572.087:575.17(540)
Original scientific paper

Patterns of Variation in a Caste-Cluster of Dhangars of Maharashtra, India

B. M. Reddy[1], D. A. Demarchi[2] and K. C. Malhotra[1]

[1] Anthropology and Human Genetics Unit, Indian Statistical Institute, Calcutta, India
[2] Department of Clinical Biochemistry, Faculty of Chemical Sciences, National University of Córdoba, Córdoba, Argentina

A B S T R A C T

We study patterns of variation among the 20 endogamous groups of Dhangars, a caste-cluster from Maharashtra State of India, who are semi-nomadic shepherds and cattle herders. To understand patterns of variation, we subjected the data on fourteen anthropometric measurements of about 2,500 adult males and data on 6 genetic markers, published among 13 of the 20 Dhangar castes, to R-matrix analysis, Harpending and Ward model of regression of heterozygosity on the distance from centroid of the populations, spatial autocorrelation analysis and Mantel statistics of matrix correspondence of the distances – geographic, anthropometric and genetic. Results of multiple regression analysis suggest a high degree of association between allele frequencies and the geographic longitude and latitude; R^2 value suggests that about 70% of the variance in RH7 and ACP can be assigned to geographic distribution of groups. In case of anthropometry, this association with body size is found to be even stronger. Results of spatial autocorrelation analysis, as suggested by Moran's (I), are somewhat complementary to those based on multiple regression analysis. Mantel test indicates significant association between anthropometric distances and the geographic distances, not between geographic and genetic distances. The extent of differentiation of Dhangar sub-castes is much higher in anthropometric traits (F_{ST} = 0.068) when compared to the genetic markers (F_{ST} = 0.023). Yet, the F_{ST} value obtained for genetic markers is larger than the average for the Indian populations, based on similar class of markers. The positioning of the groups in the multivariate space reflects primarily geographic proximity of the groups with reference to anthropometric dimensions while no tangible pattern is evident for genetic markers. The plot of average heterozygosity of the groups versus their distance from the gene frequency centroid seems to reflect population size variation, rather than-group variation in external gene flow.

Received for publication June 24, 2001.

Introduction

Because of many historical events and different waves of migration, the Indian population exhibits enormous cultural diversity with many of the major ethnic/racial groups and linguistic families represented. In each geographical and linguistic area, the population is also subdivided into a number of endogamous castes, tribes and religious communities. Many of the castes are large and widely distributed with further subdivisions or-castes within them. These sub-castes vary in size, mating patterns and even adaptive strategies. While it is possible that many of these sub-castes have common origin, Karve and Malhotra[1] suggest sociological phenomenon, for example, sanskritization, rather than genetic fission, as responsible for the existence of the sub-castes and recommend that they be called caste-clusters. India abounds in such caste-clusters or sub-castes in every part and/or linguistic area, and each of those clusters are identified with, broadly speaking, an occupation. Given that the caste system is approximately 3000 years old, this scenario provided a variety of situations to address questions pertaining to the microevolution of the groups.

Although a large number of endogamous groups of India have been studied for various biological and genetic parameters, only a few systematic regional[2–4] and caste-cluster studies[1] have been conducted. A few efforts were also made in collating the existing data and deriving patterns of variation at regional and national levels[5–7]. Among the large-scale systematic studies in India, the multidisciplinary project among semi-nomadic Dhangars of Maharashtra undertaken during 1969 –1974 by the Deccan College, Pune and Indian Statistical Institute, Kolkota under the leadership of K.C. Malhotra stands as unique. As part of this project, data on a wide variety and battery of variables- cultural, ecological, demographic and biological- were collected among all the 20 endogamous groups of Dhangars. Further the samples were drawn from all the districts of Maharashtra state, using rigorous stratified random sampling design developed by T.V. Hanurao and R. Chakraborty. Findings from this project have been published in a number of papers, covering ecology[8–11] population structure[12–15] and variation- genetic markers[16–20] dermatoglyphs[21–23] and anthropometry[24]. In the present study, we attempt to elucidate the patterns of variation among the Dhangars with reference to anthropometric and genetic marker data, applying four different methods: (i) Harpending and Jenkin's R-matrix analysis[25], (ii) Harpending and Ward method of correlating genetic heterozygosity (H) to the distance from the centroid of the gene frequency array (rii)[26] and its extension to quantitative traits by Relethford and Blangero[27], (iii) Mantel test[28] and (iv) spatial autocorrelation[29–31]. The latter two tests give insights into geographic patterning and interaction between geography and biological variables.

Several earlier papers[8–24] dealing with Dhangars provided very detailed description of these populations and, therefore, we shall provide only a brief account here. The Dhangars of Maharashtra are traditionally semi-nomadic pastorals (maintaining a variety of live stock), distributed through out heterogeneous environments of Maharashtra. The archaeological evidence and ethnographic data suggest that the contemporary Dhangar castes are the result of more than one migration from North-West India, between 4000 and 10000 BC[32]. They are estimated to number about 3 millions. The density and distribution patterns of the different groups of Dhangars seem to have been guided by the suitability of the region for the sustenance of the animals that they

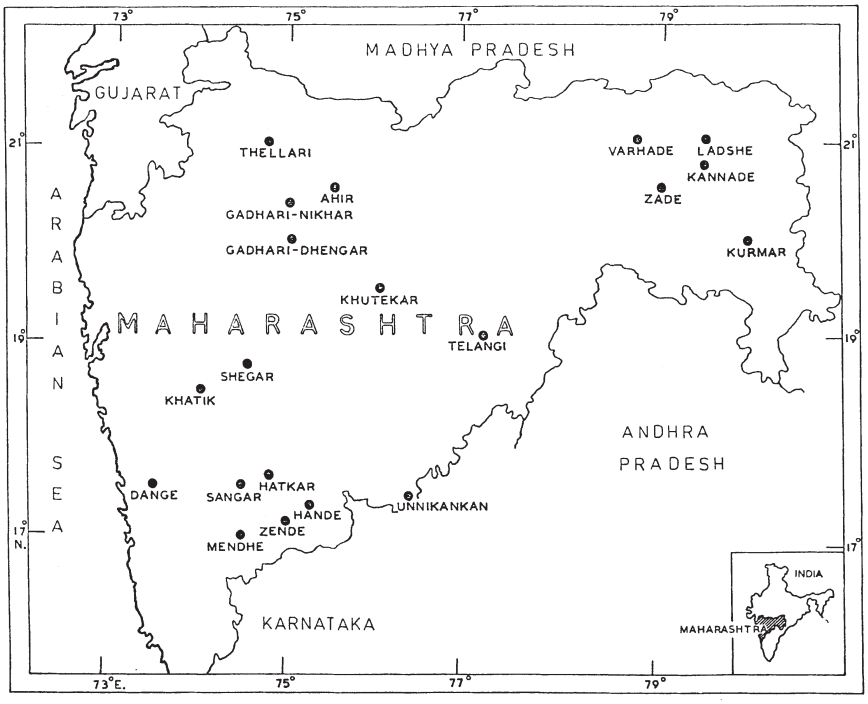

Fig. 1. Map of Maharashtra showing the core areas of the distribution of individual Dhangar castes.

traditionally maintained and/ or the products of those animals on which the specific groups subsisted[9]. While some of these groups are highly localized, some are widely distributed, this variation in distribution being largely dependent on their numerical strengths. The core areas of the distribution of individual castes are given in Figure 1. On the basis of this, the twenty Dhangar castes can be broadly categorized into four regional groups, viz. southern, central, northwestern and northeastern. Given the characterization of Khatik, Shegar, Khutekar and Telangi as 'central groups' the regional identity of the other groups is clear-cut on the map (Figure 1). The names of the populations, their sizes, and linguistic, geographical and occupational backgrounds are given in Table 1.

Ethno-historic investigations among the Dhangars (not published as yet) suggest that the Kannade, Unnikankan and Kurmar who speak Kannada were originally from Karnataka and might have migrated to the present habitats in Maharashtra at different points of time. Whereas Hatkars, Zende, Thellari and Dange trace their origin to a single caste in the remote past, Shegars claim that they have nothing to do with the Dhangars and are descendents from Rajaputs of Rajasthan, a Northwestern state of India. Ahirs who speak 'Ahrani' a mixed dialect of Gujarathi and Marathi should show closer affinity to the Ladshe and Dange who have supposedly come from Gujarat. On the other hand, Gadhari-Nikhar and Gadhari-Dhengar, having migrated from North India, speak Hindi

TABLE 1
THE LIST OF POPULATIONS AND SAMPLE SIZES (FOR ANTHROPOMETRIC TRAITS), AND LINGUISTIC, GEOGRAPHICAL AND OCCUPATIONAL BACKGROUNDS OF THE POPULATIONS STUDIED

Population No Name	Population size	Sample size	Geographic location LAT	LON	Language	Occupation*
1. Ahir	300,000	241	75.00	20.50	Marathi	SK/WW
2. Dange	100,000	154	73.50	17.50	Marathi	BK
3. Gadhari Dhengar	20,000	66	75.00	20.00	Hindi	SK
4. Gadhari Nikhar	5,000	50	75.00	20.33	Hindi	SK/WW
5. Hande	4,000	58	75.25	17.25	Kannada	SK
6. Hatkar	573,000	449	74.75	17.67	Marathi	SK
7. Kannade	15,000	52	79.50	20.75	Marathi	SK
8. Khatik	15,000	125	74.00	18.50	Marathi	MS
9. Khutekar	550,000	368	76.00	19.50	Marathi	SK/WW
10. Kurmar	15,000	88	80.00	20.00	Kannada	SK/WW
11. Ladshe	6,000	92	79.50	21.00	Marathi	SK/CW
12. Mendhe	30,000	113	74.50	17.00	Marathi	SK/WW
13. Sangar	10,000	73	74.50	71.50	Marathi	WW
14. Shegar	40,000	63	74.50	18.75	Marathi	SK/WW
15. Telangi	5,000	55	77.25	19.00	Telugu	SK/WW
16. Thellari	7,000	94	74.75	21.00	Marathi	SK/CK
17. Unnikankan	6,000	57	76.50	17.33	Marathi	SK/WW
18. Varhade	100,000	57	78.75	21.00	Marathi	SK/CW
19. Zade	15,000	62	79.00	20.50	Marathi	SK/WW
20. Zende	80,000	120	75.00	17.17	Marathi	SK/HK

*SK = sheep keeping; WW = woolen weaving; MS = meat selling; CW = cotton weaving; CK = cattle keeping; HK = horse keeping; BK = buffalo keeping

and should be genetically distinct from other Dhangar castes. Among the others, while the Telangi speaks Telugu and probably migrated from Andhra Pradesh, the remaining groups speak Marathi. The Khatiks are said to have derived from Khutekars and should show close affinity to them. Reflecting these historical backgrounds, naturally, they speak four different languages- Marathi, Hindi, Kannada and Telugu. The rate of admixture among the different Dhangar castes is estimated to be only about 1 in 1000 marriages[18], suggesting high degree of subgroup endogamy. These groups also vary enormously in terms of population size and mating patterns, providing suitable frame for studying population structure and patterns of variation.

Material and Methods

Fourteen anthropometric measurements, useful in delineating ethnic characteristics, were collected during 1970–1974 on 2437 adult males from the 20 endogamous sub-groups, distributed in 177 villages spread over 82 Tahsils of all the 26 districts of Maharashtra. Twelve of these 14 measurements are on head

and face: head length (HL), head breadth (HB), minimum frontal breadth (MFB), bizygomatic breadth (BZB), bigonial breadth BGB), upper facial height (UFH), nasal height (NH), nasal breadth (NB), biorbital breadth (BOB), inter orbital breadth (IOB), orbitonasal arc (ONA) and head circumference (HDC). The other two measurements are vertex height (HT) and height tragus (HTRG). The measurements were taken following the standard methods of Martin and Saller[33]. The measurements were taken by the same set of investigators on all the subjects.

In addition to this we have analyzed published data on 6 genetic loci (ACP, HP, ABO, MN, P, and Rh), that are commonly available for 13 of the 20 Dhangar groups[19,34]. For the remaining 7 groups not all the 6 loci were typed and hence the analyses restricted to 13 populations. The total number of blood samples typed varied enormously among the loci, about 1400 for HP to 2400 for A1A2BO blood groups. However, for majority of the groups the minimum sample size is nearly 100 or above. Only for 5 of the 13 groups the sample size is in the range of 60–70.

Statistical analyses

In order to understand the relationship among the sub-populations of Dhangars we have applied R-matrix analysis of Harpending and Jenkins[25] to the genetic marker data. This is basically a topological approach of representing population structure, and relationships among the groups are represented graphically in two dimensions by eigen vectorial reduction of the covariance matrix. The Relethford and Blangero method[26], which extends the above to quantitative variables was used for anthropometric data. The centroids of the populations are projected on to the first two vectors/principal coordinates.

The relative roles of systematic vs. non-systematic forces in the differentiation of sub-structured populations can be inferred using the model of Harpending and Ward[27] for genetic data and its extension by Relethford and Blangero[26] for anthropometric data. According to this model, given the uniform systematic pressure (gene flow) from outside genetic heterogeneity is negatively correlated with the genetic distance from the centroid of the gene frequency array (rii). Higher than average gene flow into any of the subgroups is expected to reflect in the higher than average heterozygosity/phenotypic variance than predicted by regression. Conversely, populations experiencing isolation and drift show lower values and lie below the regression line.

The basic steps involved in the application of Harpending and Ward model are given below:

Given n loci with two alleles at each locus, the expected heterozygosity of population i $E(H_i)$ is a function of the heterozygosity of the total region (H_t) and the genetic distance between population i and the regional gene frequency centroid (r_{ii}). That is,

$$E(H_i) = H_t(1 - r_{ii}). \qquad (1)$$

This equation provides an expected linear regression line of heterozygosity on genetic distance from the centroid with intercept H_t and slope $-H_t$. For two alleles at each locus the heterozygosity of the total region is computed under the assumption of complete panmixia as

$$H_t = \quad 2p_kq_k/n, \qquad (2)$$

where p_k and q_k ($= 1 - p_k$) are the weighted mean allele frequencies for locus k and summation is over all n loci. The mean allele frequencies are computed as

$$p_k = \quad w_i p_{ik}, \qquad (3)$$

$$q_k = 1 - p_k, \qquad (4)$$

Where *wi* is the ratio of the census size of population *i* to the total census size over all groups, p_{ik} is the frequency of one allele at locus *k* in population *i*, and summation is over all groups.

The genetic distance of a population to the regional centroid is computed as the diagonal of the *R* matrix of scaled variances and covariances about the regional mean allele frequencies[25]. For each allele the elements of the *R* matrix for populations *i* and *j* are computed as

$$r_{ij} = (p_i - p)(p_j - p) / p(1 - p). \qquad (5)$$

The overall *R* matrix is then averaged over all alleles. The *R* matrix provides an estimate of genetic kinship relative to the contemporary region. That is, it measures deviations from the contemporary mean allele frequencies.

The observed heterozygosity of population, *I,* is computed as

$$H_i = \quad 2p_{ik}q_{ik} / n, \qquad (6)$$

where summation is over all n loci. Under the assumption that all populations experience the same amount of gene flow from the same source (a homogeneous »outside world«), the expected heterozygosity and observed heterozygosity for population i will be the same. If either the rate or source of external gene flow is different among populations, then the expected relationship will not hold for all populations. In particular, Harpending and Ward[27] show that populations having greater than average external gene flow will have observed heterozygosities greater than expected. Comparison of expected heterozygosity with observed heterozygosity allows assessment of which populations, if any, have experienced greater than average external gene flow $H_i > E(H_i)$ or less than average external gene flow $H_i < E(H_i)$.

Mantel test

The statistical significance of the association between anthropometrics, genetics, and geographic distance matrices were estimated employing the Mantel test[28]. Given two distance matrices, A and B, this method examines association between their elements by using the statistic

$$Z_{AB} = A_{ij}B_{ij}$$

Where A_{ij} and B_{ij} are elements of row i and column j of matrices A and B, which results in an unnormalized correlation coefficient. The statistic Z_{AB} is normalized into product-moment correlation coefficient that ranges from –1 to +1. The significance of the observed correlation is tested by comparing it against the sampling distribution of Z, based on a randomized B matrix, B_R.

Spatial autocorrelation analysis

Spatial autocorrelation analysis[29–31] is used to explore spatial patterning of variate values and to gain insights into processes which structure the patterns of variation among the groups distributed in a wide geographic territory. The data for the spatial autocorrelation analysis consist of the geographic locations (Longitude and Latitude) the focal points of the 20 Dhangar castes and the 14 anthropometric measurements on the one hand, and the 17 allele frequencies on the other. The geographic distances between these populations range from 145 to about 726 kilometers. Moran's I, as opposed to Geary's C, was chosen as the coefficient of autocorrelation because of its numerical and statistical properties[30,31,35]. Compared to continental scales, the distribution of Dhangars is limited to a maximum distance of only about 730 kilometers. Hence, only about 5 distance classes with a lag of about 150 kilometers were used for the purpose of estimating I.

Multiple regression analysis

To test the correspondence between biological variation and geography, spatial correlation analysis was supplemented by the multiple regression analysis, using the square roots of allele frequencies and the anthropometric measurements as dependent variables and the longitude and latitude as the independent variables.

Results

Patterns of population differentiation

Figure 2 gives graphical representation of the Dhangars populations' centroids onto the first two principal coordinates, derived from the R-matrix analysis of anthropometric variables. Approximately 60% of the total variance is accounted for by the first two eigen values; the first axis accounts for 32.2 % and the second for 26.3 % of the total variance. The variables with highest positive correlation with the first axis are UFH, BOB, and HL, whereas BZB, HB, ONA, and MFB score highly on the 2nd axis. The dispersion of populations in the multivariate space suggests that the Dhangar populations distributed in the southern areas of Maharashtra are clearly separated from the rest of the populations on the 1st axis, although the differentiation of north-western and north-eastern regional groups of Dhangars is not very apparent. These populations are, however, scattered along the 2nd axis but with no particular pattern consistent either with the geographic positioning, migration history or linguistic background. However, the largest populations – Hatkar, Khutekar, Ahir and Dange – with wide geographic distribution, are placed centrally. The Khatik, whose distribution is restricted to Pune and certain other urban areas, stand out as an outlier, clearly separated from other groups. However, the unrooted NJ tree constructed from the D^2 distances (Figure 3) shows three major branches, broadly representing three regional groups. One of the three constitutes 7 populations of southern Maha-

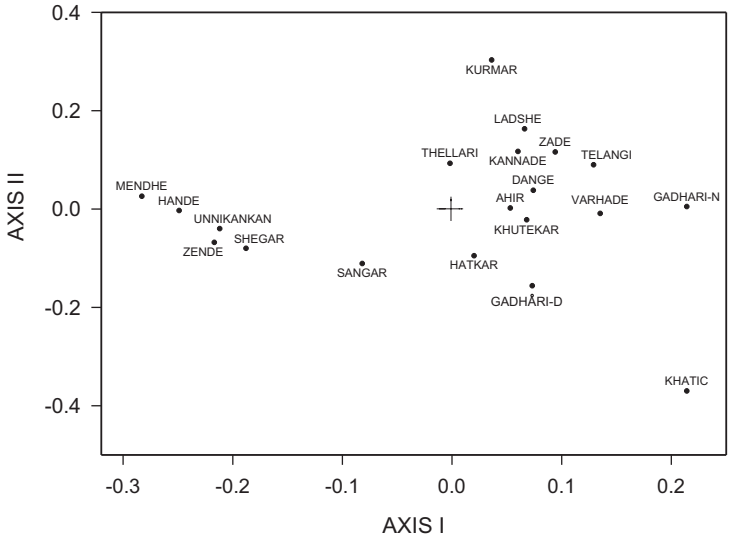

Fig. 2. R-matrix analysis of the anthropometric variables. Graphical representation of the centroids of the Dhangar populations on to the first two eigenvectors.

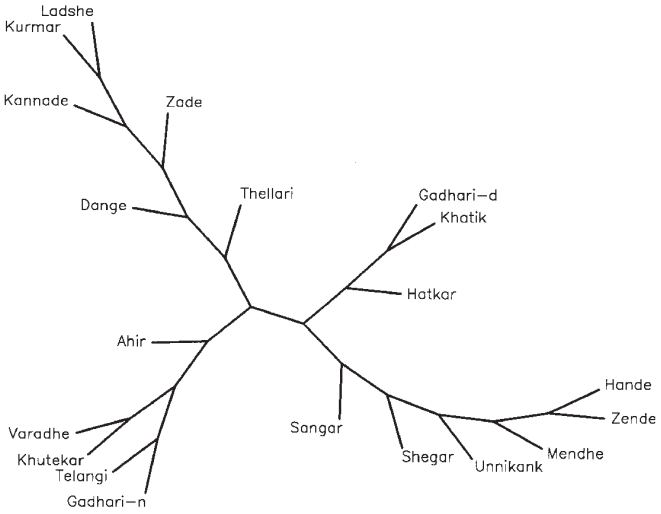

Fig. 3. Unrooted NJ tree constructed from the D^2 distances based on anthropometric variables.

rashtra, whereas the 2nd branch constitutes 4 of the northeastern populations, which are joined by Thellaries from northwest and Dange from the southwest. The other branch constitutes Ahir, Khutekar, Varhade, Telangi and Gadhari-Nikhar, including two populations from northwestern region. The central groups, true to their ambiguous geographic position, integrated into either southern or northwestern clusters.

Genetic markers

Figure 4 is a true least square reduction of an R-matrix into a genetic map of 13 Dhangar populations based on 6 loci (ABO, MN, P, Rh, AcP and Hp) with a total of 17 alleles. The first two eigen vectors account for about 51% of the total variance, about 28% and 23%, respectively, by the first and second eigen vectors. On the first axis, ABO*O and RH*3 present high positive correlation, whereas ABO*A and ABO*B show high negative correlation. While ABO*O and P*1 show high positive score, P*2 shows high negative correlation on the second axis. The migrant Telugu population, the Telangi, is clearly separated from most of the southern and some northeastern populations on the first axis, while on the 2nd axis the southern populations mostly lie on the lower half, separated from the northern populations which lie in the upper half of the plot. The Kannada speaking Hande and Kurmar, although presently geographically far apart, are placed proximate to each other in the genetic map, adhering to their linguistic affiliation. Hatkar and Zende who trace their origin to a common caste are also placed relatively closer in the genetic map.

Given vastly different effective population sizes of Dhangar castes weighting against population sizes might blur the action of genetic drift[36], hence population weights were not used in the computation. The unweighted F_{ST} values suggest moderate genetic differentiation in the

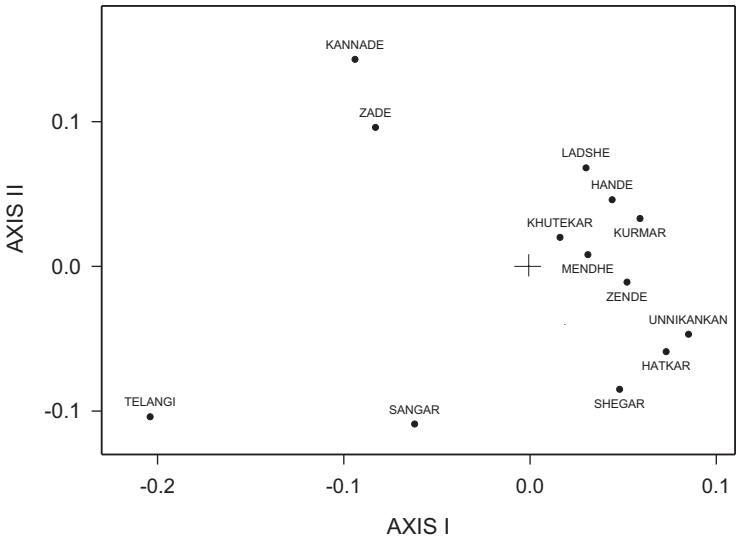

Fig. 4. Eigenvector plot of R-matrix showing the relationships among Dhangar populations based on genetic markers.

Dhangar caste- cluster with reference to genetic markers (F_{ST} = 0.023). A relatively much higher differentiation (almost 3 times) is observed in anthropometric dimensions (F_{ST} = 0.068 0.002), especially when average heritability of 0.55 is considered as approximation for Dhangars and used in the computation. The relatively much greater degree of differentiation in anthropometric dimensions is also evident in the way the populations are dispersed in the eigen vector plots made on the comparable scale (Figures 2 and 4). Given the wide geographic distribution of the Dhangar castes and due to plasticity of anthropometric dimensions the larger F_{ST} value, despite short evolutionary history, is probably expected when compared to the genetic markers.

Heterozygosity and rii

Figures 5 and 6 are the plots of the regression of phenotypic variance and mean per locus heterozygosity, respectively, against the *rii* values in case of anthropometry and genetic markers. The observed and expected values of phenotypic variance in case of anthropometry and observed and expected heterozygosities for genetic markers along with Rii values are furnished in Tables 2 and 3. Most populations in the regression plot of anthropometry (Figure 5) lie close to the theoretical line, only the migrant and non-Marati speaking Telangi and Kurmar populations with small phenotypic variance lie as distinct outliers below the line. This suggests greater isolation of the migrant populations and less than average gene flow into them from outside. While the Kurmar shows very large *rii*, the Telangi exhibits intermediate *rii* value, which probably reflects isolation and small population size, with different histories of population formation of the two groups. The Hande and Mendhe from southern Maharashtra show relatively high phenotypic variance but intermediate value of *rii*, suggesting some effect of external (from outside Dhangar castes) gene flow and reproductive isolation from the other Dhangar castes. Another inter-

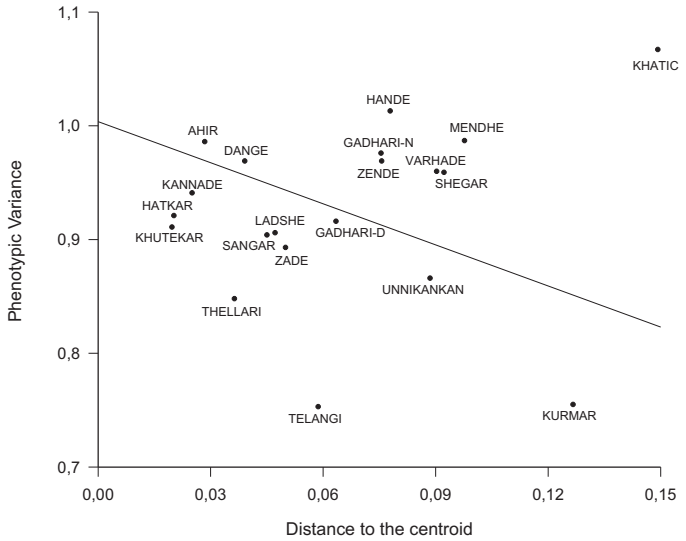

Fig. 5. Regression plot of average phenotypic variance on the distance of each group from the centroid (Rii) of the Dhangar populations and the theoretical regression line.

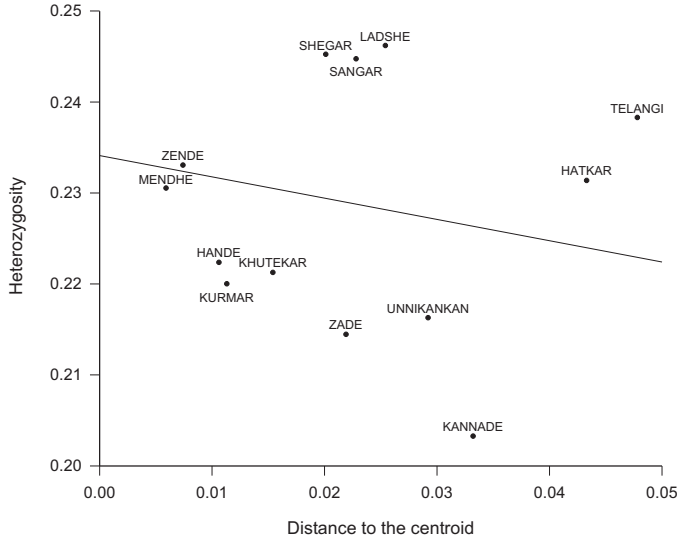

Fig. 6. Regression plot of average heterozygosity on the distance of each population from the gene frequency centroid (Rii) of the Dhangar populations and the theoretical regression line.

esting feature of this graph is the numerically predominant groups – Ahir, Dange, Hatkar and Khutekar – all show relatively high phenotypic variance but are close to the centroid. Of all the populations, Khatik shows itself as the most glaring

TABLE 2
OBSERVED AND EXPECTED PHENOTYPIC VARIANCES AMONG DHANGAR POPULATIONS: RELETHFORD-BLANGERO ANALYSIS (R_{II} = DISTANCE TO THE CENTROID)

Population	R_{ii}	Within-group phenotypic variance		
		Observed	Expected	Residual
Ahir	0.028	0.986	0.965	0.021
Dange	0.039	0.969	0.954	0.015
GadhariDengar	0.063	0.916	0.930	–0.014
GadhariNikhar	0.075	0.976	0.918	0.058
Hande	0.077	1.013	0.916	0.097
Hatkar	0.020	0.921	0.973	–0.052
Kannade	0.025	0.941	0.968	–0.027
Khatik	0.209	1.067	0.785	0.282
Khutekar	0.019	0.911	0.973	–0.063
Kurmar	0.126	0.755	0.867	–0.112
Ladshe	0.045	0.904	0.947	–0.043
Mendhe	0.097	0.987	0.896	0.091
Sangar	0.047	0.905	0.946	–0.041
Shegar	0.092	0.959	0.901	0.058
Telangi	0.058	0.753	0.935	–0.181
Thellari	0.036	0.848	0.957	–0.108
Unnikankan	0.088	0.866	0.905	–0.039
Varhade	0.090	0.960	0.903	0.056
Zade	0.049	0.893	0.943	–0.050
Zende	0.075	0.969	0.918	0.051

Mean within-group phenotypic variance = 0.925

outlier above the line, showing not only the largest phenotypic variance but also the largest distance from the centroid.

The regression plot for the genetic markers (Figure 6) is somewhat inconsistent with the anthropometric pattern. Of the 13 populations, only the positions of Unnikankan, Khutekar and Zade are somewhat similar in the two regression plots. Broadly speaking, while the Shegar, Ladshe and Sangar are placed above the regression line, somewhat removed from the centroid, the migrant Telangi and the large group of Hatkars are placed above the line but farther away from the gene frequency centroid. While the Zende and Mendhe are placed in close vicinity of the regression line, the remaining six populations are placed below and farther from the line, suggesting less than average gene flow from outside Dhangar castes. While the Telangi, Sangar, Ladshe, Hande and Hatkar are placed as outliers above the regression line suggesting greater than average gene flow into these populations in case of genetic markers, the same populations are placed below the line suggesting less than average gene flow in case of anthropometry. The two sets of variables thus present discordant patterns of population structure. Further, the Hatkars and Kannade, which are closest to the centroid in case of body measurements, are placed farthest in genetic

TABLE 3
THE OBSERVED AND EXPECTED MEAN HETEROZYGOSITIES AND THE RII

Population	Rii	Heterozygosity		
		Observed	Expected	Residuals
Hande	0.0106	0.2223	0.2277	–0.0053
Hatkar	0.0433	0.2313	0.2292	0.0022
Kannade	0.0332	0.2033	0.2287	–0.0255
Khutekar	0.0154	0.2213	0.2279	–0.0066
Kurmar	0.0113	0.2200	0.2277	–0.0077
Ladshe	0.0254	0.2462	0.2284	0.0178
Mendhe	0.0060	0.2305	0.2275	0.0031
Sangar	0.0228	0.2448	0.2282	0.0165
Shegar	0.0201	0.2452	0.2281	0.0171
Telangi	0.0478	0.2383	0.2294	0.0089
Unnikank	0.0292	0.2163	0.2286	–0.0123
Zade	0.0219	0.2145	0.2282	–0.0138
Zende	0.0073	0.2331	0.2275	0.0055

Fst = 0.0226; Mean heterozygosity = 0.2341

markers making the interpretation somewhat difficult.

Certain geographic trends are seen in the way the populations are placed in the eigen vector plots, especially in the case of anthropometric variables. In order to gauge the nature and extent of the geographic patterning we used the results of three related analytical procedures, which are outlined below.

Multiple regression analysis

The results of the multiple regression analysis of the different anthropometric dimensions on the latitude and longitude of the populations, taken at the core point of their distribution, are presented in Table 4. It is interesting to note that nine of the 14 anthropometric measurements, barring HL, BGB, UFH, NB and BOB, show highly significant association with geography. Each of the 9 anthropometric dimensions shows strong negative correlation with longitude suggesting that the western populations are bigger in size. R-Square values suggest that while over 50% of variation in HB, MFB and HDC is explained by the geographical position of the populations, about 40–45% is similarly explained by ONA, BZA and HT. Similar analysis in case of genetic markers (Table 5) suggests high correlation between some allele frequencies and the latitude and longitude. R-Square values suggest that the proportion of variance in allele frequency attributable to geographic variation is 68% in ACP*1, 65% in RH7, and 51% and 42% in RH3 and RH5, respectively.

Mantel tests

The correlation between genetics and geography (r = 0.11) and genetics and anthropometric (r = 0.007) distance matrices are small and insignificant. However, the correlation between anthropometric and geographic distance matrices is very high (r = 0.61) and, despite limited number of the degrees of freedom, statistically highly significant ($p<0.01$). Given the

TABLE 4
MULTIPLE REGRESSION ANALYSIS OF THE ANTHROPOMETRIC VARIABLES ON LATITUDE AND LONGITUDE

Allele	R^2	F prob.	Partial correlation	
			LAT	LON
HT	0.396	0.014	–0.177	–0.466[1]
HTRG	0.327	0.035	0.092	–0.520[1]
HL	0.071	0.533	0.219	–0.255
HB	0.599	0.001	–0.338	–0.586[2]
MFB	0.539	0.002	–0.021	–0.650[2]
BZB	0.403	0.013	–0.163	–0.480[1]
BGB	0.223	0.117	0.055	–0.418
UFH	0.198	0.153	0.382	–0.424
NH	0.364	0.021	–0.276	–0.363
NB	0.225	0.115	–0.420	0.056
BOB	0.191	0.165	0.358	–0.424
IOB	0.338	0.030	0.537[1]	–0.539[1]
ONA	0.448	0.006	0.277	–0.650[2]
HDC	0.507	0.003	–0.173	–0.571[1]

1=p<0.05; 2=p<0.01

TABLE 5
MULTIPLE REGRESSION ANALYSIS OF THE SQUARE ROOT OF THE GENE FREQUENCIES (P)* ON LATITUDE AND LONGITUDE

Allele	R^2	F prob.	Partial correlation		Number of samples
			LAT	LON	
ABO1	0.146	0.453	0.096	–0.275	13
ABO2	0.215	0.289	–0.107	–0.162	13
ABO3	0.173	0.386	0.173	–0.343	13
ABO4	0.373	0.097	–0.281	0.522	13
ACP1	0.681	0.050	–0.799[1]	0.825[1]	8
HP1	0.228	0.313	0.173	0.106	12
HP2	0.066	0.737	–0.156	0.029	12
HP3	0.045	0.818	0.129	–0.024	12
M	0.249	0.276	0.499	–0.451	12
P1	0.056	0.751	–0.236	0.202	13
RH1	0.084	0.965	0.005	–0.046	13
RH2	0.059	0.965	0.098	0.031	13
RH3	0.505	0.029	0.127	0.365	13
RH4	0.279	0.194	–0.169	–0.150	13
RH5	0.420	0.065	0.635[1]	–0.635[1]	13
RH6	0.185	0.359	–0.048	–0.193	13
RH7	0.648	0.005	–0.765[2]	–0.570[1]	13

1=p < 0.05

small geographic range of population distribution, lack of systematic correlation between geographic and genetic distance matrices is not surprising. That there is no concordance in the pattern of population affinities based on anthropometry (Figure 2) and genetic markers (Figure 4) is corroborated by insignificant Mantel correlation between the distance matrices based on these two sets of variables. Furthermore, the correlation between anthropometric and geographic distance matrixes remained virtually same (r = 0.613), even after partialling out the effect of genetic distances, suggesting lack of interaction of genetic distances with either anthropometric or geographic distances. Therefore, the geographic distances between the populations can account for about 40% of the variance in anthropometric distances between them. Overall, the above results suggest that the population affinities are geographically patterned in anthropometric traits, not in genetic markers.

Spatial autocorrelation

Given that the anthropometric measurements among the Dhangars are patterned geographically spatial autocorrelation was used to examine the form of this patterning and explore the kind of spatial processes implicated. Results of the spatial autocorrelation analysis are summarized in Tables 6 and 7 for anthropometry and genetic markers, respectively. There appears to be a subtle trend of monotonic decline in the level of spatial autocorrelation – a relatively large positive values at the smallest spatial lags followed by a gradual decline to the largest negative values towards the extreme spatial lags. However, because of the limited number of spatial lags we have not attempted to draw correlograms. Eight of the 14 anthropometric dimensions show statistically significant correlation, but for the genetic markers only RH3 shows a monotonic decline with high positive correlation at the smallest lag followed by gradual decline with high

TABLE 6.
SPATIAL AUTOCORRELATION (MORAN'S ***I***) RESULTS FOR ANTHROPOMETRIC VARIABLES

Body measurements	Distance lags in kilometers					Overall P
	145	291	436	581	726	
HT	0.41[2]	–0.04	–0.67[2]	–0.70[2]	0.21	0.000
HTRG	0.41[2]	–0.14	–0.69[2]	–0.66[2]	0.37[1]	0.000
HL	–0.13	–0.02	–0.01	0.12	–0.31	0.293
HB	0.38[2]	0.10	–0.39[2]	–0.63[2]	–0.70[2]	0.000
MFB	0.33[2]	0.06	–0.16	–0.56[2]	–0.88[2]	0.000
BZB	0.25[2]	0.07	–0.28	–0.51[2]	–0.41	0.000
BGB	0.05	–0.01	0.02	–0.17	–0.53[2]	0.090
UFH	–0.12	0.11	0.01	–0.13	0.03	0.512
NH	0.14[2]	–0.11	0.08	–0.22	–0.74[2]	0.008
NB	–0.09	0.09	–0.24	0.11	–0.42	0.268
BOB	–0.09	0.01	–0.02	0.03	–0.33	0.445
IOB	–0.09	0.03	–0.16	0.10	–0.38	0.395
ONA	0.22[2]	–0.12	–0.09	–0.18	–1.04[2]	0.000
HDC	0.25[2]	0.06	–0.16	–0.33[1]	–1.14[2]	0.000

1=p<0.05; 2=p<0.01 (significance Bonferroni approximation)

TABLE 7
SPATIAL AUTOCORRELATION (MORAN'S *I*) RESULTS FOR GENETIC VARIABLES (13 POPULATIONS)

Alleles	Distance lags in kilometers					Overall P
	145	291	436	581	726	
ABO*0	–0.00	0.17	–0.19	–0.30	–0.10	0.483
ABO*A_1	0.30[1]	–0.20	–0.31	0.04	–0.23	0.137
ABO*A_2	0.05	–0.22	0.06	–0.05	–0.24	1.000
ABO*B	0.18	0.06	0.02	–0.25	–0.40[1]	0.165
ACP*1	–0.41	–0.00	–0.00	–0.21	–0.04	0.431
HP*1	–0.08	–0.29	–0.15	0.25	–0.09	0.586
HP*2	–0.05	–0.24	–0.30	0.24	–0.10	0.278
HP*3	–0.07	–0.22	–0.31	0.23	–0.09	0.312
M	–0.16	–0.16	–0.42[1]	0.20	0.03	0.198
P*1	–0.22	0.09	–0.28	–0.22	0.20	0.321
RH1	–0.26	0.05	–0.21	–0.12	0.10	0.698
RH2	–0.26[1]	–0.17	0.08	–0.02	–0.04	0.855
RH3	0.73[2]	0.11	0.13	–0.74[2]	–0.58[2]	0.000
RH4	0.17	–0.22	0.21[1]	–0.35	–0.19	0.248
RH5	–0.03	–0.03	–0.20	–0.07	–0.09	1.000
RH6	0.10	–0.19	0.23[1]	–0.51[1]	–0.02	0.084
RH7	0.14	–0.15	–0.02	–0.23	–0.15	0.657

1=p<0.05; 2=p<0.01 (significance Bonferroni approximation)

negative correlation at the extreme distance lags. None of the other alleles show any clinal pattern except for a meek and insignificant trend in ABO*B.

Discussion

Comparisons based on thirteen populations bring out patterns of population structure that are somewhat contradictory with reference to genetic markers and anthropometry. The different statistical analyses suggest that the pattern of anthropometric variation among Dhangars is governed primarily by the geographic affiliations of the groups than by the microlinguistic and/or ethno-historic backgrounds. However, neither geographic nor ethno-historical patterns are apparent in case of the genetic markers. There is, nevertheless, a semblance of ethnic/linguistic affinities in the way some populations formed clusters. For example, the Kannada speaking Hande and Kurmar, although presently geographically separated, are placed close to each other as is the case with Hatkars and Zende who had apparently had derived from the same caste in the recent past. Overall, discrimination based on anthropometric variables is much greater than that of the genetic markers. This may be because of the fact that the six loci studied represent very little of their genetic constitution compared to the polygenetic characters like anthropometric traits. Despite anthropometric variables usually being mediated by environmental noise, a clear geographic pattern observed may not totally belie the inherent genetic resemblance. Despite the traditional norms of endogamy it is this geographic proximity that might have provided opportunities for gene flow, whereas geographically

separated populations despite social parity often fail to get access to each other to exchange genes due to number of constraints. In this context, it is interesting to note that recent findings based on DNA and traditional markers also suggest that the Indian populations cluster according to geographic rather than ethnic affiliation[37,38].

The total lack of correspondence between anthropometric and genetic distance matrices is somewhat puzzling and prompts us to surmise whether different sets of variables evolve differently. The lack of correspondence in the distance matrices of different sets of dermatoglyphic variables has also been observed earlier among the same Dhangar groups[39].

Certain insights are provided by the regression analysis of phenotypic variance on rii, again in anthropometric variables. For example, the Khatiks, having been mostly isolated from the Dhangar castes and restricted to urban locales show very large rii as well as the highest phenotypic variance, suggesting greater degree of admixture with non-Dhangar groups as well as greater isolation from other Dhangar populations. Given that the Khatiks live mostly in towns, subsisting as butchers, they have probably had greater opportunities for genetic interaction with heterogeneous urban populations, hence their position in the regression plot may not be surprising. Similarly, large phenotypic variance and placement very close to the centroid shown by Ahir, Hatkar, and other numerically large populations, is genetically quite meaningful given that these few groups represent over 70% of the Dhangar population in the state besides having very wide geographic distribution. The relatively large phenotypic variance among these groups could be due to large population sizes rather than external gene flow, as has been recently illustrated by one of us[40]. The pattern based on genetic markers among the 13 populations does not display any of these features and the observed pattern does not lead to any rational genetic interpretation of their population structure.

Another consistent feature that has been brought out by the spatial autocorrelation analysis of the anthropometric traits is the positive value of Moran's I, at least at the first distance lag, followed by gradual decline to large negative values, although the decline is not quite monotonic in nature. Yet this may suggest that the populations, which are geographically closer tend to have greater similarity in anthropometric dimensions, when compared to the farther ones. These results almost complement the results of multiple regression analyses with very minor differences. The fact that in most cases the autocorrelation (Moran's I) becomes either zero or tend to be negative from the 2nd distance lag itself may be a pointer to the Indian population structure wherein the gene flow is restricted by geography only within the castes. Between the castes the gene flow is restricted by social norms even when their members reside the same village. Besides this, traditionally, the marital movement had been restricted to very small distances. Nevertheless, the Moran's I value in most cases indicates decreasing relatedness with increasing distance. In case of genetic markers such a pattern is seen only in one of the 16 alleles (RH3). Overall, the expectations borne out of the ethno-historical and linguistic information outlined in the last paragraph of the introduction are not adequately reflected in the results, except for very minor and subtle indications in the plot. To conclude we may say that the population structure of the Dhangars is overwhelmingly mediated by the geographic position of the groups in case of body dimensions whereas no significant pattern is brought out by the genetic markers. One of the plausi-

ble reasons for this could be the limited number of loci considered for the study and relatively short evolutionary history of these sub-castes. Increasing number of these loci and/or the analysis of hypervariable DNA markers of different kind-mtDNA, Y-based and nuclear- may offer better insights into the population structure and patterns of variation of this semi-nomadic and pastoral caste-cluster of Dhangars.

Acknowledgement

The authors thank the anonymous reviewers for their comments on the initial version of this paper, which helped in much improved presentation of the results.

REFERENCES

1. KARVE, I., K. C. MALHOTRA, Current Anthropol., 9 (1968) 109. — 2. MAJUMDER, D. N., P. C. MAHALANOBIS, C. R. RAO, Sankhya, 9 (1949) 90. — 3. KARVE, I., V. M. DANDEKAR: Anthropometric measurements of Maharashtra. (Deccan College Monograph Series No. 8, 1951). — 4. MAJUMDER, D. N., C. R. RAO: Racial elements in Bengal: A quantitative study. (Statistical Publishing Society, Calcutta, 1960). — 5. KRISHNAN, T., B. M. REDDY, Ann. Hum. Biol., 21 (1994) 155. — 6. MAJUMDER, P. P., Evolutionary Anthropol., 6 (1998) 100. — 7. REDDY, B. M., D. A. DEMARCHI, M. H. CRAWFORD, S. BHARATI: Palmar dermatoglyphic variation among Indian populations. (Manuscript). — 8. GADGIL, M., K. C. MALHOTRA, Hom. Ecol., 10 (1982) 107. — 9. MALHOTRA, K. C., M. GADGIL, South Asian Anthropologist, 2 (1981) 49. — 10. MALHOTRA, K. C., M. GADGIL, Subsistence strategies among the pastoralists of Maharashtra, India. In: LUKACS, J. R. (Ed.): The people of South Asia. (Plenum Press, New York, 1984). — 11. MALHOTRA, K. C., M. GADGIL, Coping with uncertainty in food supply: Case studies among the pastoral and non-pastoral nomads of western India. In: DE GARINC, I., G.A. HARRISON (Eds.): Coping with uncertainty of food supply. (Clarendon Press, Oxford, 1988). — 12. MAJUMDER, P. P., K. C. MALHOTRA, Ann. Hum. Biol., 6 (1979) 17. — 13. MALHOTRA, K. C., J. Biosoc. Sci., 11 (1979) 397. — 14. MALHOTRA, K. C., Coll. Antropol., 3 (1979) 19. — 15. MALHOTRA, K. C., Population structure among the Dhangar caste cluster of Maharashtra, India. In: LUKACS, J. R. (Ed.): The people of South Asia. (Plenum Press, New York, 1984). — 16. DAS, S. R., K. C. MALHOTRA, B. N. MUKHERJEE, S. K. DAS, Jap. J. Hum. Genet., 18 (1974) 401. — 17. CHAKRABORTY, R., A. CHAKRAVORTY, K. C. MALHOTRA, Ann. Hum. Biol., 4 (1977) 275. — 18. MALHOTRA, K. C., R. CHAKRABORTY, A. CHAKRAVORTY, Hum. Hered., 28 (1978) 26. — 19. MUKHERJEE, B. N., S. K. DAS, K. C. MALHOTRA, Ind. J. Med. Res., 64 (1976) 1240. — 20. MUKHERJEE, B. N., S. K. DAS, K. C. MALHOTRA, S. R. DAS, J. Genet., 63 (1977) 39. — 21. MALHOTRA, K. C., Qualitative finger dermatoglyphics variation among 21 endogamous Dhangar castes of Maharashtra, India. In: WERTELECKI, W., C. C. PLATO (Eds.): Dermatoglyphics: Fifty years later. (Alan R. Liss, Inc., New York, 1979). — 22. CHAKRABORTY, R., K. C. MALHOTRA, J. Hum. Evol., 10 (1981) 503. — 23. KARMAKAR, B., K. C. MALHOTRA, P. NARAIN, Anthropologie, 27 (1989) 13. — 24. MALHOTRA, K. C., B. M. REDDY, B. V. BHANU, Z. Morph. Anthrop., 81 (1995) 79. — 25. HARPENDING, H. C., T. JENKINS, Genetic distances among southern African populations. In: CRAWFORD, M. H., P. L. WORKMAN (Eds.): Methods and theories of anthropological genetics. (NM: Univ. of New Mexico Press, Albuquerque, 1973). — 26. RELETHFORD, J. H., J. BLANGERO, Hum. Biol., 62 (1990) 5. — 27. HARPENDING, H. C., R. H. WARD, Chemical systematics and human populations. In: NITECKI, M. H. (Ed.): Biochemical aspects of evolutionary biology. (University of Chicago Press, Chicago, 1982). — 28. MANTEL, N., Cancer Research, 27 (1967) 209. — 29. CLIFF, A. D., J. K. ORD: Spatial autocorrelation. (Pion, London, 1973). — 30. CLIFF, A. D., J. K. ORD: Spatial processes. (Pion, London, 1981). — 31. SOKAL, R. R., N. L. ODEN, Biol. J. Linn. Soc., 10 (1978) 199. — 32. ALCHIN, F. R.: Neolithic cattle keepers of south India: A study of the Deccan Ashmounds. (Cambridge University Press, Cambridge, 1963). — 33. MARTIN, R., K. SALLER: Lehrbuch der Anthropologie. (G. Fischer, Stuttgart, 1957). — 34. MALHOTRA, K. C., S. K. DAS, B. N. MUKHERJEE, B. V. BHANU, J. V. UNDEVIA, S. YEE, P. P. MAJUMDER, Ind. J. Phys. Anthropol. & Hum. Genet., 3 (1977) 141. — 35. HAGGET, P., A. D.CLIFF, A, FREY: Locational analysis in human geography. (John Willey & Sons, New York, 1977). — 36. RELETHFORD, J. H., Hum. Biol., 68 (1996) 29. — 37. REDDY, B. M., G. SUN, J. R. LUIS, M. H. CRAWFORD, N. S. HEMAM, R. DEKA, Hum. Biol., 73 (2001) 175. — 38. REDDY, B. M., R. DUTTA, B. T. LANGSTIEH, V. K. KASHYAP, Int. J. Hum.Genet., 1 (2001) 1. — 39. KARMAKAR, B., Intern. J. Anthropol., 5 (1990) 255. — 40. REDDY, B. M., Hum. Biol., 73 (2001) 57.

B. M. Reddy

Anthropology and Human Genetics Unit, Indian Statistical Institute,
203 Barrackpore Trunk Road, Calcutta – 700 035, India

UZORCI VARIJACIJA U KLASTERU KASTA DHANGARS IZ MAHARASHTRA, INDIJA

S A Ž E T A K

Istraživan je uzorak varijacija među 20 endogamnih skupina polu-nomadskih stočara Dhangars, klastera kasti iz Maharshtra države, Indija. Korišteni su podaci o četrnaest antropometrijskih mjera za 2500 odraslih muškaraca te podaci o 6 genetskih markera koji su do sada objavljeni za 13 od 20 Dhangars kasti. U cilju razumijevanja uzorka varijacije, napravljena je R-matriks analiza, Harpending i Wardov model regresije heterozigotnosti na udaljenost od centroida populacija, prostorna autokorelacijska analiza i Mantelova statistika podudaranja matrica zemljopisne, antropometrijske i genetske udaljenosti. Rezultati analize višestruke regresije sugeriraju visok stupanj povezanosti između frekvencije alela i zemljopisne duljine i širine; R^2 vrijednosti sugeriraju da oko 70% varijance u RH7 i ACP može se pripisati zemljopisnoj raspodijeli skupina. Što se tiče antropometrije, povezanost s veličinom tijela pokazala se još i snažnijom. Rezultati prostorne autokorelacijske analize su, kao što je sugerirao Moran, donekle komplementarni onima zasnovanima na analizi višestruke regresije. Mantelov test upućuje na značajnu povezanost između antropometrijskih i zemljopisnih udaljenosti, ali ne i između zemljopisnih i genetskih udaljenosti. Razmjer diferencijacije Dhangars pod-kasta puno je veći u antropometrijskim osobinama ($F_{ST} = 0.068$) u usporedbi s genetskim markerima ($F_{ST} = 0.023$). No ipak, F_{ST} vrijednosti dobivene za genetske markere su veće od prosječnih u populacijama Indije, zasnovanih na sličnoj klasi markera. Pozicioniranje skupina u multivarijatnom prostoru odražava prvenstveno zemljopisnu bliskost skupina, a odnosi se na antropometrijske dimenzije, dok niti jedan uočljiv uzorak nije razvidan za genetske markere. Plot prosječne heterozigotnosti skupina naprema njihovim udaljenostima od centroida frekvencije gena izgleda da u većoj mjeri odražava varijacije u veličini populacije nego varijacije među skupinama u stupnju vanjskog protoka gena.