

Gene Differentiation Among Ten Endogamous Groups of West Bengal, India

R. CHAKRABORTY, H. WALTER, B. N. MUKHERJEE,
K. C. MALHOTRA, P. SAUBER, S. BANERJEE, AND M. ROY
Center for Demographic and Population Genetics, University of Texas
Graduate School of Biomedical Sciences, P.O. Box 20334, Houston, Texas
77225 (R.C.); Department of Human Biology/Physical Anthropology,
University of Bremen, D-2800 Bremen 33, West Germany (H.W., P.S.); and
Anthropometry and Human Genetic Research Unit, Indian Statistical
Institute, 203 B. T. Road, Calcutta 700 035, India (B.N.M., K.C.M., S.B.,
M.R.)

KEY WORDS Gene differentiation, Biochemical markers, West Bengal, Genetic distance, F_{ST} analysis, Caste and tribes

ABSTRACT Ten endogamous populations of West Bengal, India have been surveyed for genetic variation in 12 systems. These populations encompass all social ranks in the caste hierarchy and cover almost the entire geographic area of the state. Gene diversity analysis suggests that these groups exhibit significant allele frequency variation at all but three loci. The overall genetic difference is not, however, in accord with the classification based on caste. Two low-ranking scheduled caste groups are, in fact, in close proximity with the high-caste ones, suggesting evidence of past generations of gene flow among them. Three different clusters of groups emerge from the present data, providing support for the anthropologic assertion that in Bengal Proto-Australoid, Caucasoid, and Mongoloid racial elements generally coexist. However, these three components are not uniformly present in all groups. Geographic separation of the groups is a strong determinant of the gene differentiation that exists among these populations.

Indian populations, representing nearly one-seventh of the World's total population, offer some interesting as well as unique opportunities for genetic and anthropological studies because of the diversity that exists within them. Perhaps nowhere else can one find as much intertwined social stratification based on caste hierarchy, language, religion, and traditional occupation, each of which dictates the mating pattern within each stratum, and hence gene flow across the different strata of the Indian society. Furthermore, during the long political history of the subcontinent, the land has been invaded by many racial stocks, each of which has had some impact on the local gene pool. Thus, in India it is not unreasonable to find evidence of gene flow from ethnic groups from beyond the present geographical boundaries of the country.

The populations of the eastern states of India, namely, West Bengal and Assam, are

no exception to this description. A series of anthropological studies, based on anthropometry and somatoscopy (Majumder and Rao, 1958; Chakraborti and Mukherji, 1971) argue for the presence of Mongoloid, Caucasoid, and Proto-Australoid racial elements in West Bengal. However, since these traits are environmentally influenced, it is uncertain to what extent these features are common to the groups characterized by caste, religion, and language stratification.

Serological and biochemical markers are more ideal for identification of the genetic affinity of populations. But in the eastern states of India (e.g., West Bengal and Assam), scanty information is available with regard to the distributions of red cell antigens, iso-

Received April 28, 1986; revision accepted June 20, 1986.

Address reprint requests to Dr. Ranajit Chakraborty, Center for Demographic and Population Genetics, University of Texas Graduate School of Biomedical Sciences, P.O. Box 20334, Houston, Texas 77225.

zymes, and serum proteins. In view of this, we have undertaken a systematic study of genetic variation in West Bengal and Assam, encompassing the endogamous groups of all social strata. In this paper, we report the genetic relationships among ten groups of West Bengal, based on allele frequency analyses of 12 genetic systems. The objectives of the present analyses are: 1) to examine the extent of gene differentiation across different strata of the populations of West Bengal to determine where most differentiation exists; 2) to study the intra- and intergroup diversity among the populations; and 3) to understand the pattern of gene differentiation as a function of the geographic proximity and social position of these populations. Based on these analyses we examine to what extent the different racial elements are common to all caste denominations.

MATERIALS AND METHODS

Design of the survey

The genetic survey conducted for the present study is a part of a collaborative research program of the Indian Statistical Institute, Calcutta and the University of Bremen, West Germany. Ten population groups from West Bengal, India, were sampled on the basis of 1) their caste endogamy; 2) at least one group from each social strata representing high caste, scheduled caste (a common term to designate the low ranking caste groups that are protected by the Government), and tribes; and 3) representation of the populations with low and high probability of Mongoloid admixture. Accordingly, four groups were chosen from the northern part of West Bengal where the chance of Mongoloid admixture is high, and three each from the southwest and southeast zones of West Bengal, where the chance of Mongoloid admixture is low. The distribution of the population groups, with their social rank, language spoken, total population size in the districts where they are found, the location of the sampling sites, and the sample sizes are given in Table 1. Figure 1 shows the geographic position of the groups on a map of West Bengal.

A total of 1,000 blood samples was collected. The populations encompass the northern districts, Coochbehar and Jalpaiguri, in the *tarai* region of the Himalaya; and the two southern districts, 24 Parganas and Midnapur, which represent the coastal region of the Gangetic plains of West Bengal. The distance between these two major zones is about 500 km.

Blood collection and laboratory methods

The blood samples were collected in the field and were refrigerated before dispatching them to the Human Genetics Laboratory of the Indian Statistical Institute, Calcutta, where they were analysed. ABO phenotype testing was done by agglutination tests in test tubes with anti-A and anti-B sera; and gammaglobulin (Gm and Km) groupings were done with anti-Gm-1, -2, -5, and Km-1 sera, following Steinberg (1980). Serum protein systems, haptoglobin (Hp), transferrin (Tf), ceruloplasmin (Cp), and albumin (Alb) were studied by horizontal starch gel electrophoresis (Ashton and Beede, 1961), and the red cell systems, acid phosphatase (ACP), adenylylase kinase (AK), and esterase D (EstD), were scored for genotype determinations by starch gel electrophoresis according to Harris and Hopkinson (1977). Typings for phosphoglucomutase-1 (PGM1) for the Rabha and Rajbanshi populations were done by the same method, while for the rest of the populations PGM1 subtypes were determined by an isoelectric focusing technique on polyacrylamide gels (Mukherjee et al., 1982). The hemoglobin (Hb) screening was conducted at the B.J. Medical College, Pune, by Kate et al. (1984).

We condensed the PGM1 genotypes by ignoring their subtypes for the present analysis to preserve uniformity over all populations. For the lack of sufficient blood samples, the Gm and Km systems could not be scored in the Mech population.

Methods for statistical analysis

The statistical analysis of the present data consists of four parts. First, determination of allele frequencies for systems other than the ABO and Gm were done by the gene count method which gives the maximum likelihood estimates of the allele frequencies. The allele frequencies at the ABO and Gm systems were determined by the maximum likelihood scoring method, following an updated version of Reed and Schull's (1968) algorithm. Secondly, heterogeneity of allele frequencies among the ten groups was tested first by a heterogeneity chi-square statistic (Rao, 1965), to show that these populations are differentiated enough to conduct a gene diversity analysis. Subsequently, the extent of gene differentiation was studied by Wright's F_{ST} statistic, to examine the pattern of gene diversity among the two high caste groups (Rarhi, Brahmin, and Vaidya), among the three scheduled caste groups (Rajbanshi,

TABLE 1. Populations of West Bengal sampled—their location, social rank, and sample size

Group	Population size in district	Location of sampling sites				Social rank	Language	N
		Village/town	District	Lat. (°N)	Long. (°E)			
Berahi	—	Barasat, Sodepur, Medhyamgram, Baranagore	24 Pgs.	22.50	88.25	H. caste	Bengali	100
Brahmin	—	Medhyamgram, Barasat, Baranagore	24 Pgs.	22.40	88.30	H. caste	Bengali	103
Vaidya	418,893	Pitonsai, Dinhaata, Baranagore	Coochbehar	26.18	89.32	S. caste	Bengali	115
Rajbanshi	119,187	Makrapur, Hirapur, Bhara, Middapara, Narayanagarh, Chatur, Bidisa, Pakursani	Midnapur	22.25	87.24	S. caste	Bengali	100
Jatia	5,561	Dum Dun, Belaghata, Chowdar, Barasatbari, Taliguri, Tulanganj town	Calcutta	22.35	88.21	S. caste	Bengali	101
Bhadra	1,608	Garopara, Baro Aulbari, Phalkata	Coochbehar	26.25	89.50	Tribe	Bengali & Thakur-Burman	114
Garo	1,279	Mahakaiguri, Nawabgung	Coochbehar	26.22	89.42	Tribe	Bengali & Thakur-Burman	97
Mech	153	Belda, Doharpur, Bidisa, Katighora	Jalpaiguri	25.30	88.50	Tribe	Thakur-Burman	96
Munda	15,960	Makrapur, Hirapur, Narkunda, Beldia, Bidisa, Shaidanga, Kuki, Naryanagarh	Midnapur	22.50	87.40	Tribe	Bengali & Austro-Asiatic	100
Lodha	11,205		Midnapur	22.25	87.28	Tribe	Bengali	74

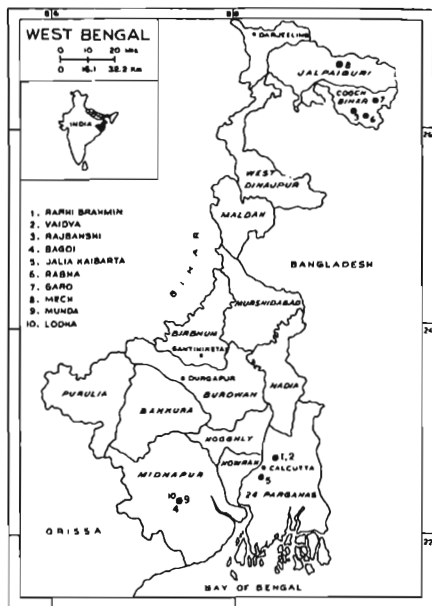


Fig. 1. Geographic locations of the populations studied in relation to the district boundaries of the State of West Bengal, India.

Bagdi, and Jalia Kaibarta), among the five tribes (Rabha, Garo, Mech, Munda, and Lodha), and among all ten groups, following the procedure adopted in Chakraborty et al. (1977). Thirdly, genetic distances among all ten groups were computed by Nei's standard genetic distance (Nei, 1972), and their standard errors were computed by Nei and Roychoudhury's (1974) method. The distance matrix was subjected to a cluster analysis by the modified unweighted pair (UPG) method of Nei (1975). Fourthly, the relationship between genetic proximity and geographic distance of these populations was examined by relating the normalized gene identity (I_r) be-

tween two populations with their geographic distance (r). If the genetic relationship between populations is governed by mutation, gene flow, and random drift, it is known that I_r is exponentially related with r (Maruyama, 1973; Blanco and Chakraborty, 1975; Malhotra et al., 1978). We fitted a regression equation

$$I_r = \alpha + \beta e^{-r}, \quad (1)$$

where α and β were estimated by considering all 45 pairs of the ten populations. Since these observations are interdependent, instead of

using a least square approach to estimate the parameters of the above regression equation, we employed a technique advocated by Smouse et al. (1986) which circumvents the complications due to the dependency of the 1, values.

ETHNOHISTORICAL BACKGROUND OF THE POPULATIONS

In order to interpret the pattern of gene differentiation among the sampled populations, it may be worthwhile to note a few features of the ethnohistoric background of the major populations of West Bengal. Risley (1891, 1908) noted that to some extent all of the seven physical types of Indian populations, namely, 1) Turko-Iranian, 2) Indo-Aryan, 3) Scytho-Dravidian, 4) Aryo-Dravidian, 5) Mongolo-Dravidian, 6) Mongoloid, and 7) Dravidian, are visible among the Bengalees of West Bengal. In earlier studies, the high caste Brahmins and Kayasthas of lower Bengal have been typed as Mongolo-Dravidian (Risley, 1908). Based on broad head shape, dark complexion, plentiful face hair, medium stature, and medium nose shape with a slight tendency for broadness, the high caste Bengalees were thought to be admixed groups with Dravidian and Mongoloid elements in them. In the upper Bengal and Himalyan region, Mongoloid features extend beyond the high caste groups, perhaps in a purer form, for example, in the Gurung of Nepal, and the Bodo and Cochi in Assam, where broader head shape, yellowish dark complexion, scanty facial hair, small stature, and characteristic flat face with often oblique eyelids are more commonly observed. The Dravidian type is found in some tribes like the Munda and Santals of south Bengal, who are originally from the Chotanagpur region. Their stature is short, their complexion is very dark, and they have plentiful facial hair with an occasional tendency to curly hair, dark eyes, long head, and a broad nose. None of these features are totally exclusive of one another and considerable overlapping exists within any given endogamous group, and thus, a classification based on these anthropometric typings is not free of error. Since in the current ethnohistoric literature the terms Caucasoid and Proto-Australoid are commonly used to indicate Indo-Aryan and Dravidian ancestry, in this paper we will use the terminology of Caucasoid for Indo-Aryan and Proto-Australoid for Dravidian interchangeably.

Rarhi Brahmin

A subcaste of Brahmins, the Rarhi Brahmins, are supposed to be direct descendants of five Brahmins and their wives who were brought to Bengal from Kanauj (Oudh of Uttar Pradesh) during the eleventh century A.D. to perform some priestly Vaidic duties about which the local priests were ignorant. According to this theory, the Rarhi Brahmins are of comparatively pure Aryan descent, deriving their ancestry from the five immigrated Brahmin families from the north-central part of India (Dutt, 1969). The name is derived from the region of their principal concentration in West Bengal, *Rarhi*—the high lying alluvia tract on the western banks of the Ganges.

This notion is contested by Dutt (1969), who claims that the historical documents of the genealogy writers are conflicting regarding the dates of arrival of the immigrant Brahmin families, their origin, and designation of clans (*gotras*). Furthermore, there are several stories in the *Pauranic* literature, suggesting that several families carrying Rarhi Brahmin gotra designation may be the result of intermixing of Brahmin and nonBrahmin (e.g., Bagdi and Kaibarta) families (Dutt, 1969). The significance of such gene admixture will be clear from our analysis later in this paper.

Nevertheless, like other high caste groups, Rarhi Brahmins do not practice consanguineous marriage; their mother tongue in Bengal is Bengali; and while in the past they exclusively performed priestly services and teaching duties, at present they are mostly engaged in white collar jobs. Even though marriages among Rarhi Brahmins were mostly within the caste group (but not within the same *gotra*) in the past, at present intercaste marriage is not prohibited and is not uncommon.

Vaidya

This is a typical caste found mostly in Bengal, generally occupying a high rank in society. Often associated with the profession of physicians, the exact time and origin of this functional name as a separate caste group is under dispute among different scholars (Dutt, 1969). Some consider the Vaidya to be an offshoot of the Brahmins who were intermarried with other castes like Vaishya, Sudra, etc., but there are also claims that they are

direct descendents of Aryans who immigrated to Bengal. Though in these later claims Vaidyas are regarded as more monophyletic than Brahmans, we shall see that genetic analysis provides some indication of their mixed ancestry.

It is difficult to distinguish the Vaidyas physically from other high castes. The Vaidyas' mother tongue is Bengali and they are mostly in white collar jobs. According to the areas of residence of their ancestors in Bengal, there are four subcastes of the Vaidyas: Rarhi, Banga, Barendra, and Panchakati, which is parallel to the classification of the Brahmans.

Rajbanishi

Rajbanishis constitute the largest scheduled caste community in West Bengal (about 17.5% of the total scheduled caste of the state). They are mostly concentrated in Coochbehar, Jalpaiguri, West Dinajpur, and 24 Parganas districts. Risley (1908) suggested that they could be the converted form of the Koch tribe. Like the Ahoms of the Brahmaputra valley, the Rajbanishis also probably settled in North Bengal from the northeast (Census of India, 1931), and hence they may have a Mongoloid affinity. In lower Bengal (e.g., in 24 Parganas) Rajbanishis are also known as Tiyar or Koyots. They are mainly agriculturists, and in some area they also practice fishing. Consanguineous marriage is absent among them; they speak Bengali and practice Hinduism. Their social status is generally low even though recently their position has improved to some extent.

Bagdi

Bagdis are also a major scheduled caste in West Bengal. Except in the northern districts where their concentration is very low, they are widely distributed all over the state. According to mythology, they are believed to have descended from the offspring of a *Kshatriya* father and a *Vaisya* mother. Dalton (1872), however, suggested that the Bagdis are the remnant of an aboriginal tribe who, by intermarrying with lowcaste Hindus, left their tribal way of life and became fishermen and carriage bearers. Several subdivisions exist within them, e.g., Tentulia, Dule, Matia, etc. Their low social rank is due to their primary occupations—cultivating, fishing, and menial work. Their religion is a mixture of Hinduism and nature worship, their mother tongue is Bengali, and they are prohibited to practice cross-cousin and uncle-niece marriages.

Jalia Kaibarta

Earlier known as the Jalia Kaibarta, this subcaste of Kaibartas derive their name from their profession—fishing. At present, their counterpart, *Chasi Kaibartas*, is regarded as a practically separate caste (Dutt, 1969). They are distributed widely throughout the state, and are often considered to be one of the earliest inhabitants of Bengal. Risley (1891) reports that the nucleus of the group was probably Dravidian, but their original features may have been refined to some extent by infusion of Aryan blood. They are Hindus by religion, speak Bengali, and occupy a low position in the social hierarchy. Consanguineous marriage is prohibited in this community as well. Some scanty accounts are available regarding their mixing with Brahmin families, suggesting that some gene flow between this scheduled caste and the Rarhi Brahmin may have taken place in the past (Dutt, 1969).

Rabha

The Rabhas are mostly concentrated in the two northern districts, Jalpaiguri and Coochbehar, and constitute a small segment of the total tribe population of the state. In these districts they are also subdivided into two subgroups: *Koch Rabha* and *Pati Rabha* (Das and Raha, 1967). Dalton (1872) thought the Rabha to be an offshoot of the Kachari tribe and suggested an affiliation with the Garos. Risley treated them as a branch of the Bodo group of Assam. Whatever may be their origin, this is also a recent migratory group in Bengal, and they exhibit some Mongoloid features. They are mostly settled agriculturists at present, and they speak Bengali and a Tibeto-Burman language with a striking similarity to the *Atong* dialect of the Garo language.

Garo

A small, recent immigrant tribe from Assam to North Bengal, they are mostly settled in the Coochbehar district. It is believed that they found their passage through Assam from Tibet, and their present composition is mostly due to admixture of Tibetans with Hindus. Linguistically they are close to the Rabhas, speaking Bengali as well as a Tibeto-Burman form of the Bodo Assamese language. Ethnohistoric report of their close affiliation with the Rabhas is also available (Dalton, 1872). Today they are mostly agriculturists, and a portion of them profess Hinduism and the rest Christianity.

Mech

This is also a small tribe, mostly living below the hilly regions of the Jalpaiguri and Coochbehar districts of north Bengal. They are regarded as having a common origin with the Kacharis of Assam. Some investigators believe that the Mech are descendants of the Nepalis of Nepal. Distinct Mongoloid features are commonly seen in their physical appearance. Earlier they were mostly engaged in shifting cultivation, and now in addition, they do spinning, weaving, fishing, etc. Regarding religion, they mainly follow their own tribal deity *Bathow*, although some practice Hinduism. The Mech are generally bilingual, they have their own dialect which belongs to the Tibeto-Chinese family, akin to the dialects of the Kachari and Bodo groups in Assam. In addition, they also speak Bengali.

Munda

Mundas of West Bengal comprise about 7.9% of the total tribal population of the state and are mainly found in the districts of Burdwan, Midnapur, 24 Parganas, West Dinajpur, and Darjeeling. Their main concentration in eastern India is, however, in the west of the state, in the Chotanagpur hills and the plateau area of the districts of Singbhum, Ranchi, and Manbhum of Bihar. Linguistically they are classified as Kolarian and are close to the *Hos* and *Santals* and to some extent to the *Kandhs*. In West Bengal, the Mundas are a migratory group, engaged in agriculture, and many are employed as plantation laborers in tea gardens. The Mundas speak Mundari as their mother tongue, but most of them also speak Bengali. They are basically Animists and their deity is known as *Sing Bonga* (the Sun). The Mundas are divided into 13 endogamous groups which are again subdivided into a number of clans based on distinct totems. Generally consanguineous marriage is not allowed among the Mundas.

Lodha

The Lodhas are a small tribal group mostly found in the Midnapur district with small numbers of them also in the Hoogly district. Risley (1908) described them as allied to the *Savra* or *Savar* tribe of the Mayurbhanja of Orissa, while others claim that they are from the Madhya Pradesh (Central Province) of India. The Lodhas consider themselves as *Savra*, and their Oriya affiliation is also found in their language, which is a corrupt form of Bengali with some Oriya influence.

They still adhere to their traditional occupation of collecting jungle produce, though some now engage in agriculture and daily labor. They are divided into nine exogamous clans and do not allow consanguineous marriage.

In summary, these ethnographic accounts suggest that the ten populations may be grouped in three classes. The *Rarhi Brahmins* and *Vaidyas*, the two high-caste groups, are probably derived from Caucasian ancestry, although some admixture cannot be ruled out. If there are any other groups that may have close genetic proximity with them, the likely candidates could be the *Bagdis* and *Jalia Kaibartas*, as noted earlier (Dutt, 1969). The second class consists of a scheduled caste (*Rajbanshi*) and three tribes (*Rabha*, *Garo*, and *Mech*). All of the constituent groups of this class are most likely the result of considerable Mongoloid admixture with the local tribes of North Bengal and Assam. The third class consists of the two tribes: *Munda* and *Lodha*, both of which are migrant Proto-Australoid groups who came into West Bengal from the west or southwest of the state.

To avoid any bias in our analysis, we shall first follow the traditional classification: high caste, scheduled caste, and tribe, to determine whether heterogeneous gene diversity is found within each class, and then attempt to place them according to their genetic proximity, as derived from a combined analysis of the 12 systems studied here. Of particular interest will be the placement of the two lower caste groups, the *Bagdi* and *Jalia Kaibarta*, since from the above ethnographic accounts their affiliations do not exactly correspond to their present social position.

RESULTS

Allele frequency heterogeneity among the ten populations

Table 2 provides the basic data analyzed in this paper. The data are summarized in the form of allele frequencies, since the phenotypic distributions are not needed for the present analyses (see Mukherjee et al. in press; Chakraborty et al. in press for the phenotypic frequencies and their analyses).

These allele frequencies were subjected to a chi-square test to determine whether heterogeneity exists among the ten groups in any genetic system. Table 3 gives the summary statistic in the form of chi-square values with associated degrees of freedom. Since the allele frequencies are not available for the Cp system for the *Jalia Kaibartas* and for the Gm and Km in the *Mech* population,

TABLE 2. Allele frequencies in ten endogamous groups of West Bengal, India

Allele	RB	VA	RJ	BD	JK	RA	GA	ME	MU	LO
ABO										
A	0.219	0.207	0.161	0.252	0.211	0.268	0.092	0.116	0.248	0.291
B	0.252	0.278	0.156	0.237	0.262	0.187	0.663	0.194	0.227	0.282
O	0.629	0.515	0.683	0.461	0.527	0.577	0.663	0.720	0.527	0.427
	100	103	115	100	101	114	97	98	100	74
Hp										
Hp*1	0.194	0.120	0.250	0.147	0.168	0.186	0.181	0.158	0.124	0.184
Hp*2	0.806	0.880	0.750	0.853	0.832	0.814	0.819	0.842	0.876	0.816
	90	96	20	89	96	94	72	38	87	87
Hb										
Hb*A	0.964	0.966	0.897	0.874	0.972	0.939	0.982	0.827	0.975	0.962
Hb*B	0.036	0.034	0.103	0.026	0.028	0.061	0.048	0.173	0.025	0.048
	42	25	63	170	89	90	21	26	161	187
AK										
AK*1	0.879	0.925	0.980	0.914	0.926	0.976	1.000	0.991	0.935	0.980
AK*2	0.021	0.075	0.010	0.086	0.074	0.025	0.000	0.009	0.065	0.020
	94	87	93	99	101	98	38	53	100	101
ACP										
ACP* ^a A	0.253	0.220	0.227	0.218	0.265	0.317	0.250	0.270	0.200	0.162
ACP* ^b B	0.107	0.180	0.173	0.182	0.136	0.163	0.150	0.130	0.180	0.182
	98	100	110	108	100	112	70	74	100	117
PGM										
PGM1*1	0.906	0.789	0.726	0.858	0.858	0.779	0.706	0.676	0.685	0.686
PGM1*2	0.194	0.211	0.274	0.142	0.144	0.221	0.295	0.324	0.315	0.314
	53	26	93	120	90	113	95	94	100	113
ESD										
ESD*1	0.781	0.785	0.720	0.712	0.792	0.591	0.632	0.766	0.633	0.514
ESD*2	0.219	0.215	0.280	0.288	0.208	0.409	0.368	0.234	0.367	0.486
	96	100	100	106	101	99	38	32	98	108
TP										
TPC	0.995	1.000	1.000	0.975	1.000	1.000	1.000	0.994	1.000	1.000
TPD	0.005	0.0	0.0	0.010	0.0	0.0	0.0	0.0	0.0	0.0
TPB	0.0	0.0	0.0	0.010	0.0	0.0	0.0	0.0	0.0	0.0
TPV	0.0	0.0	0.0	0.005	0.0	0.0	0.0	0.006	0.0	0.0
	98	46	14	101	93	109	99	87	98	87
Cp										
Cp*B	1.000	1.000	1.000	1.000	—	1.000	0.995	0.995	0.990	1.000
Cp*A	0.0	0.0	0.0	0.0	—	0.0	0.0	0.005	0.010	0.0
Cp*V	0.0	0.0	0.0	0.0	—	0.0	0.006	0.0	0.0	0.0
	98	51	57	100	—	93	99	87	51	87
Gm										
Gm1	0.328	0.340	0.221	0.332	0.263	0.038	0.068	—	0.134	0.058
Gm2	0.180	0.190	0.051	0.081	0.191	0.047	0.114	—	0.015	0.074
Gm3	0.442	0.439	0.274	0.362	0.253	0.223	0.269	—	0.390	0.224
Gm4	0.049	0.041	0.394	0.330	0.164	0.262	0.262	—	0.461	0.254
	98	93	94	99	98	100	96	—	89	121
Km										
Km*1	0.040	0.082	0.131	0.068	0.099	0.169	0.184	—	0.170	0.038
Km*2	0.960	0.918	0.869	0.932	0.901	0.831	0.816	—	0.830	0.962
	61	70	84	95	101	100	97	—	88	121

TABLE 3. Analysis of heterogeneity of allele frequencies among ten populations of West Bengal, India

Allele	χ^2	d.f.	P
ABO			
O	63.60	9	< 0.0001
A	42.87	9 ¹	< 0.0001
B	27.80	9 ¹	0.001
Hp			
Hp*1	10.60	9 ¹	0.312
Hp*2	10.60	9	0.312
Hb			
Hb*A	37.79	9 ¹	< 0.0001
Hb*E	37.79	9	< 0.0001
AK			
AK*1	37.76	9 ¹	< 0.0001
AK*2	37.76	9	< 0.0001
ACP			
ACP*A	22.50	9 ¹	0.007
ACP*B	22.50	9	0.007
PGM1			
PGM1*1	69.15	9 ¹	< 0.0001
PGM1*2	69.15	9	< 0.0001
EstD			
EstD*1	73.69	9 ¹	< 0.0001
EstD*2	73.69	9	< 0.0001
Tf			
Tf*C	10.07	9 ¹	0.345
Tf*D	6.36	9 ¹	0.704
Tf*B	7.15	9	0.621
Cp			
Cp*A	6.98	8 ¹	0.539
Cp*B	9.25	8 ¹	0.321
Cp*V	6.24	8	0.620
Gm			
Gm ¹	163.95	8 ¹	< 0.0001
Gm ^{1,2}	73.56	8 ¹	< 0.0001
Gm ³	56.60	8 ¹	< 0.0001
Gm ^{1,3}	411.41	8	< 0.0001
Km			
Km ¹	49.54	8 ¹	< 0.0001
Km ¹	49.54	8	< 0.0001
Total	688.34	138	< 10 ⁻⁶

¹Included in the total.

the degrees of freedom for these three systems are 8 in each case. It is clear from this table that, in general, these ten populations are differentiated enough to conduct a detailed gene diversity analysis ($\chi^2 = 688.34$ with 138 d.f.; $P < 10^{-6}$ for the total of all independent alleles). Significant heterogeneity also exists for all alleles in the individual systems except for the Hp, Tf, and Cp systems.

The reasons for such allele frequency heterogeneity are not clear from these two tables. Nevertheless, a comparatively lower frequency (9–16%) of the A allele (in the ABO system) among the Rajbanshi, Garo, and Mech and a high frequency (39–69%) of the Gm^{1,3} haplotype suggests their Mongoloid affinity. The Mundas and Lodhas have a comparatively higher frequency (25 and 29%) of

the A allele; in them, the B allele also reaches a high frequency (23 and 28%). These characteristics are typical of the Dravidian populations of India (Roychoudhury, 1984), but they also exhibit a high frequency of the Gm^{1,3} haplotype (46 and 64%), suggesting that they too may have Mongoloid admixture. Any general conclusion from the trends of single allele frequencies can, however, be misleading since individual allele frequencies are known to be perturbed by past generations of random drift.

F_{ST} analysis

Since the chi-square values do not measure the extent of gene diversity directly, the results of the *F_{ST}* analysis are presented in Table 4. These computations were conducted

TABLE 4. *F_{ST}* analysis of allele frequencies in ten populations of West Bengal, India

Allele	<i>F_{ST}</i> among			
	Two high castes	Three scheduled castes	Five tribes	All ten populations
A	0.0002	0.0087	0.0391	0.0214
B	0.0009	0.0188	0.0124	0.0139
O	0.0002	0.0383	0.0374	0.0318
Hp*1	0.0106	0.0062	0.0048	0.0066
Hb*A	0.00001	0.0227	0.0218	0.0213
AK*1	0.0159	0.0194	0.0172	0.0218
ACP*A	0.0070	0.0023	0.0197	0.0114
PGM1*1	0.0263	0.0245	0.0073	0.0330
EstD*1	0.00002	0.0067	0.0203	0.0419
TP*C	0.0017	0.0058	0.0055	0.0061
TP*D	0.0017	0.0026	—	0.0038
TP*B	< 10 ⁻⁵	0.0026	0.0049	0.0043
Cp*B	—	—	0.0034	0.0048
Cp*A	—	—	0.0054	0.0054
Cp*V	—	—	0.0038	0.0043
Gm ¹	0.0001	0.0105	0.0185	0.0513
Gm ^{1,2}	0.0002	0.0321	0.0212	0.0410
Gm ³	0.0002	0.0130	0.0230	0.0315
Gm ^{1,5}	0.0004	0.0497	0.0316	0.2291
Km ¹	0.0076	0.0073	0.0360	0.0293
Average	0.0052	0.0129	0.0170	0.0241
± S.E.	± 0.0021	± 0.0024	± 0.0028	± 0.0055

Except for the alleles O, TP*B, Cp*V, and Gm^{1,5} all other alleles have been used in computing the average and its S.E.

for all ten populations taken together (last column of Table 4, and also for the three major groups individually (high caste, scheduled caste, and tribe).

It can be seen from this table that, on average, the gene diversity is the smallest for the high caste group, and the highest for the tribes. This result is consistent with Roychoudhury's (1984) analysis of gene diversity in India on a global basis. While one may infer that the tribes exchange much smaller numbers of genes among them in comparison to the caste groups, an alternative explanation is that the traditional grouping of Indian populations on the basis of caste hierarchy is not in accordance with the proximity of their gene pools. Some of the low ranking groups, may, in fact, be closer to the higher caste groups, and when they are placed with their social peers, they exhibit a dissimilarity, resulting in a higher index of gene diversity as seen in this analysis (Table 4).

Genetic distance analysis

To determine the relative position of these ten populations on the basis of their allele frequencies, we estimated the standard genetic distances among all pairs of populations and their standard errors which are

shown in Table 5. This table also shows the average heterozygosities in these populations (presented in the diagonal elements). Since for distance computations only those loci that are common to all populations can be used, we excluded the Gm and Km systems (allele frequency data on the Mech are missing for these systems). Very little variation was found in the Cp system, and it is regarded as monomorphic for these computations. The results are nearly the same if the actual allele frequencies for the Cp system are considered; however, the Jalia Kaibarta must be excluded from comparison, since we do not know the exact allele frequencies at this locus in this group.

Since the loci included in this analysis are heavily biased for polymorphic loci (eight polymorphic and two monomorphic), the average heterozygosity values may not be representative of the level of genetic variation found in humans in general. For the same reason, the absolute values of the genetic distances should not be used for estimating the time of divergence of these populations. The relative positions of the populations are, however, unaffected by this bias. Furthermore, since for this level of differentiation, almost all distance measures suggested in

TABLE 6. Standard genetic distances among ten populations of West Bengal and their S.E.¹

	RB	VA	RJ	BD	JK	RA	GA	ME	MU	LO
Rarhi (RB)	19.7 ± 6.7	3.6 ± 1.8	8.3 ± 4.5	2.8 ± 1.0	1.0 ± 0.5	7.3 ± 4.9	10.2 ± 5.9	12.8 ± 7.5	11.0 ± 5.4	18.5 ± 10.9
Brahmin (RB)		20.4 ± 6.4	7.2 ± 3.6	1.7 ± 0.9	1.0 ± 0.6	8.1 ± 5.0	7.2 ± 3.7	9.2 ± 4.9	4.6 ± 3.2	13.0 ± 9.8
Vaidya (VA)			22.8 ± 6.3	9.9 ± 5.4	7.7 ± 3.3	5.7 ± 2.7	3.1 ± 1.5	2.8 ± 1.2	6.9 ± 3.2	13.6 ± 8.6
Rajbanahi (P-J)				21.2 ± 6.6	1.5 ± 0.9	6.5 ± 2.8	9.1 ± 5.1	16.0 ± 8.0	5.3 ± 3.9	10.2 ± 6.2
Bagdi (BD)					20.3 ± 6.4	7.6 ± 5.2	8.9 ± 4.6	11.0 ± 5.7	7.7 ± 4.7	16.5 ± 10.6
Jalia Kailbarta (JK)						23.0 ± 7.0	4.1 ± 2.8	10.0 ± 5.0	4.4 ± 2.1	7.4 ± 4.1
Rabha (RA)							21.4 ± 6.8	5.4 ± 3.0	4.0 ± 2.7	6.7 ± 6.4
Garo (GA)								22.2 ± 7.1	10.1 ± 5.0	20.4 ± 12.2
Mech (ME)									22.1 ± 7.1	3.6 ± 2.1
Munda (MU)										22.9 ± 7.6
Lodha (LO)										

¹ Figures on the diagonal are the average heterozygotes expressed in percentage; the other values are in 10^{-3} codon difference per locus. The comparisons are done with eight polymorphic loci (ABo, Hp, Iib, Aic, AcP, PGM1, EstD, and Tf) and two monomorphic loci (Cp and Alb).

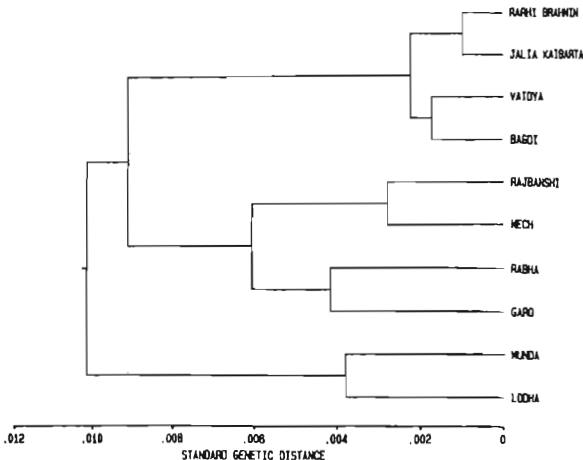


Fig. 2. Dendrogram for the ten populations of West Bengal, India, as judged from genetic distance based on eight polymorphic and two monomorphic markers.

the literature are highly correlated (Chakraborty and Tateno, 1976), the qualitative conclusions are not affected by our choice of distance measure.

One may note that even though the populations are arranged in this table according to their traditional ranking (the two high-caste groups first, then the three scheduled castes, and finally the five tribes), the matrix is hardly structured in the same fashion. For example, the genetic distances of the Bagdi (0.0028 and 0.0017) and Jalia Kaibarta (0.001 and 0.001) from the two high-caste groups (Rarhi Brahmin and Vaidya) are smaller than the same between the Rarhi Brahmin and Vaidya (0.0035). The third scheduled caste group (Rajbanshi) shows closer proximity with the Mech, Rabha, and Garo (range 0.0028–0.0057), each of which is a tribe, rather than being close to their scheduled caste compatriots. These computations show that gene diversities based on the traditional grouping or populations by caste ranking give a confused picture of the genetic differences between the populations within a group.

Clustering of the populations on the basis of these distance values is clearly seen in the dendrogram presented in Figure 2. In aggregate, we observe three major clusters: (1) One consisting of the two high-caste groups (Rarhi and Brahmin and Vaidya); (2) a second which consists of the three tribes (Mech, Rabha, and Garo) and one scheduled caste (Rajbanshi); and (3) the third consisting of the other two tribes, the Munda and Lodha. The two low-ranked scheduled castes, the Bagdis and Jalia Kaibarta, being genetically close to the two high-caste groups, belong to the first cluster, instead of combining with the other scheduled caste groups. This result is surely beyond the common notion of the origin of these two caste groups. An explanation for these three major clusters can be found in the fact that in these ten groups we have a conglomerate of Caucasoid (Indo-Aryan), Mongoloid (Mongolo-Dravidian), and Proto-Australoid (Dravidian) gene pools, and these three major components may be present in substantially different frequencies within the high-caste groups, tribes, and scheduled castes of the northern part of Bengal and the

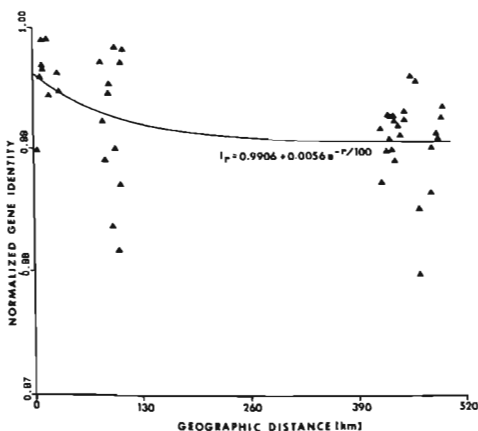


Fig. 3. Relationship of normalized gene identity with geographic distances among ten endogenous groups of West Bengal, India.

Proto-Australoid tribes (Munda and Lodha). We, thus, find the Indo-Aryan elements in the first cluster (Rarhi and Brahmin and Vaidya); the Mongolo-Dravidian elements in the cluster of Rajbanshi, Rabha, Garo, and Mech; and finally the Dravidian elements in the third cluster of Munda and Lodha. The Bagdi and Jalia Kaibarta caste groups are in the first cluster, suggesting that they may have in them a substantial remnant of past generations of gene flow from the higher caste groups, as described in the section on ethnohistory. However, this observation should be confirmed with a more extensive genetic study, since genetic clustering based on so small a number of systems is known to have substantial sampling error (Nei et al., 1983).

Relationship of genetic proximity with geographic distance

On the supposition that the ten populations examined here exchanged genes in the past as a function of their geographic proximity, we examined the empirical relationship between the normalized gene identity (I_r , Nei, 1975) and the geographic distance (r) shown in Figure 3. The solution of equation

(1), obtained by the method of Smouse et al. (1986) is given by

$$I_r = 0.9906 + 0.0056 e^{-r/100},$$

where the distance r is measured in kilometers. It should be stressed that while the regression coefficient, 0.0056 ± 0.0021 , is significant ($P < 0.05$), the fitted equation explains only 14.4% of the variation of gene identities among these populations. This is probably due to misspecification of the extract centroid of the concentration of the groups, since the sampling locations may not exactly represent the centroids of the groups within West Bengal. Most of the deviant points, however, are ascribable to the comparisons made with the Lodha group, for whom the samples were obtained from the southwestern district, Midnapur, although they extend into the Hoogly district, as noted earlier. The intercept of the fitted equation (0.9906) is not significantly different from 1.0, even though the loci studied are heavily biased for polymorphic systems. The low value of the regression coefficient signifies that the gene differences among these groups

are fairly low, perhaps due to the extent of gene flow that may have occurred in the past.

DISCUSSION AND CONCLUSION

The gene differentiation among the ten populations compared here suggests that, overall, the level of gene diversity (mean F_{ST}) is quite small (0.024 ± 0.006). This value is about one-half of the gene diversity among the 21 Dhargar caste groups of Maharashtra (0.044 ± 0.009 ; Chakraborty et al., 1977) and is somewhat higher than that among the caste groups studied here (0.005 ± 0.002 ; first column of Table 4) and elsewhere in India (e.g., Roychoudhury, 1974, 1981). The five tribal populations of West Bengal, in contrast show more differentiation. As mentioned earlier, this does not necessarily mean that the tribes in India exchange genes less commonly than the caste populations. On the contrary, we believe that some of the tribal and lower-caste populations may have accumulated genes from the high-caste groups through past generations of gene flow, and hence they may present an overall genetic profile somewhat dissimilar from their social peers. This obviously increases gene diversity of a collection of tribes or low-caste populations, for more dissimilar gene pools are compared in such events. Note that our findings that the Jalia Kaibartas are close to the high-caste groups, in spite of their low social ranking, are supported by the fact of their ethnohistoric relationship with the high-caste groups (Dutt, 1969). Roychoudhury (1984) also noted that two segments of Bagdis (Duley and Tentulia Bagdis) have close genetic proximity with the Rarhi Brahmins. His study however, does not include the Vaidyas for comparison.

In our analysis, moreover, the Rajbanshis show proximity with other tribes when the Mongoloid affinity is stronger (e.g., Mech, Rabha, and Garo). In spite of their scheduled-caste status, the genetic profile of the Rajbanshis seems to be in conformity with their affiliation with tribes having Mongoloid features, as suggested in Risley (1908) and the Census of India (1931).

The Mundas and Lodhas form a group most different from the remaining eight populations (Fig. 2). They do have some Mongoloid genes (e.g., GM^{1,2}), but they are still distant from the four populations that show Mongoloid affinity.

In addition, the overall pattern of gene differentiation is in conformity with the geo-

graphic distance between these groups (Fig. 3), although quite a bit of fluctuation from the predicted relationship exists. This could be because of misspecified locations of populations, or sampling errors in estimating the probability of allelism (I_i) from the small number of loci surveyed.

Given the history of the ten populations, it is reasonable to believe that the gene admixture that has taken place in Bengal was probably a continual process. Initially the major three components (Caucasoid, Mongoloid, and Dravidian) may have been incorporated in different segments of the Bengali society to a varying degree. Subsequent gene flow among the resultant gene pools may have transformed the genetic make-up to such an extent that the early history of gene admixture cannot be recovered from the present survey in a quantitative fashion. Nevertheless, from the individual allele frequencies, it appears that the high caste groups may not have a large Mongoloid element at present (e.g., the GM^{1,2} haplotype frequency is not significantly different from zero; Chakraborty et al., in press). The same haplotype occurs in highest frequencies in the populations belonging to the second cluster (Rajbanshi, Mech, Rabha, and Garo), suggesting that these are the groups through which the infusion of Mongoloid genes has occurred in West Bengal. The third cluster (of the Mundas and Lodhas) shows evidence of a Dravidian ethnic element in Bengal along with some Mongoloid admixture.

In summary, the genetic data presented here support the ethnohistoric and anthropological findings that the present genetic constituents of the Bengali populations are truly a conglomerate of gene pools from at least three distinct sources: Dravidian, Caucasoid, and Mongoloid. These three components are evidently at very dissimilar proportions in various groups, classified by caste hierarchy. Lastly, and perhaps more significantly, the constituent genetic profile of any given population does not always correspond exactly to its present social ranking, since some low-caste groups are seen to have stronger genetic affiliation with high-ranking groups, instead of being close to groups of their own rank. The present caste hierarchy, therefore, may not be a reflection of the genetic origin of these populations.

It should be noted that while the present findings are in qualitative agreement with the postulates given in the ethnohistory of

these populations, none of the results are conclusive. This is so because none of the systems included in this study has a marker variant that is characteristic of any single group. The traditional enzyme protein systems failed to detect such a genetic variant. However, it is possible that use of restriction fragment length polymorphisms or DNA fingerprinting may provide questions that are not addressed here.

ACKNOWLEDGMENTS

The data collection for this project was funded from collaborative research funds from the Deutsche Forschungsgemeinschaft (Wa 122/29/1) and the Indian Statistical Institute, Calcutta. The statistical analysis was supported by U.S. Public Health Service Research Grants from the National Institutes of Health. We are grateful to Professor B. P. Adhikari for his guidance and encouragement during the conduct of the survey. Comments from Dr. W. J. Schull and two anonymous reviewers on an earlier draft of this paper are greatly appreciated.

LITERATURE CITED

- Ashon, CC, and Braden AWH (1961) Serum β -globulin polymorphism in mice. *Austr. J. Exp. Biol. Med. Sci.* 142:248-255.
- Bianco, R. and Chakraborty, R (1975) Genetic distance analysis of twenty-two South American Indian tribes. *Hum. Hered.* 25:177-193.
- Census of India (1931) Ethnographic Notes by Various Authors, Vol. 1, Part IIIB. Simla: Govt. of India.
- Chakraborty, M. and Mukherjee, D (1971) Indian Tribes. Calcutta: Saraswati Lib.
- Chakraborty, R. Chakravarti, A. and Malhotra, KC (1977) Variation in allele frequencies among caste groups in the Dhangars of Maharashtra, India: An analysis with Wright's F_{ST} statistic. *Ann. Hum. Biol.* 4:275-280.
- Chakraborty, R. and Tateno, Y (1976) Correlations between some measures of genetic distance. *Evolution* 30:851-853.
- Chakraborty, R. Walter, H. Mukherjee, BN, Sauber, P. Malhotra, KC, Banerjee, S. and Roy, M (1987) Immunoglobulin (Gm and Km) allotypes in nine endogamous groups of West Bengal, India. *Ann. Hum. Biol.* (In press).
- Dalton, ET (1972) Descriptive ethnology of Bengal. Calcutta: Govt. of Bengal Printings.
- Das, AK, and Raha, MK (1967) The Rabhas of West Bengal. Calcutta: Director of Cultural Research Inst., Govt. of West Bengal.
- Dutt, NK (1969) Origin and Growth of Caste in India, Vol. II. Calcutta: Firma KL Mukhopadhyay.
- Harris, H. and Hopkinson DA (1977) Handbook of Enzyme Electrophoresis in Human Genetics. Amsterdam: North-Holland.
- Kato, SL, Mookashi, GD, Khedkar, VA, and Mukherjee, BN (1984) Prevalence of hemoglobin E in ten population groups of West Bengal, India. *Ind. J. Haematol.* 11:221-223.
- Majumder DN, and Rao, CR (1956) Bengal anthropometric survey, 1945 (A statistical study). *Sankhya* 19:201-408.
- Malhotra, KC, Chakraborty, R. and Chakravarti, A (1978) Gene differentiation among the Dhanga caste-cluster of Maharashtra, India. *Hum. Hered.* 28:26-36.
- Maruyama, T (1973) Isolation by distance, genetic variability the time required for a gene substitution, and local differentiation in a finite geographically structured population. In NE Morton (ed): *Genetic Structure of Populations*. Honolulu: Univ. Hawaii Press. pp 80-81.
- Mukherjee, BN, Sauber, P. Walter, H. and Malhotra, KC (1982) Isoelectric focusing in the detection of extended genetic polymorphism of PGM1 in ten endogamous groups of West Bengal and Orissa, India. *Proc. Int. Symp. on the Recent trends in Immunohaematology*. Bombay: Indian Council of Med. Res. pp 182-188.
- Mukherjee, BN, Walter, H. Malhotra, KC, Chakraborty, R, Banerjee, S. and Roy, M (1987) Population genetic study on ten endogamous groups of West Bengal, India. *Anthrop. Anz.* (In press).
- Nei, M (1972) Genetic distance between populations. *Amer. Nat.* 106:283-292.
- Nei, M (1975) *Molecular Population Genetics and Evolution*. New York: American Elsevier.
- Nei, M. and Roychoudhury, AK (1974) Sampling variances of heterozygosity and genetic distance. *Genetics* 76:379-390.
- Nei, M, Tajima, F. and Tateno, Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* 19:153-170.
- Rao, CR (1965) *Linear Statistical Inference and its Applications*. New York: Wiley.
- Reed, TE, and Schull, WJ (1968) A general maximum likelihood estimation program. *Amer. J. Hum. Genet.* 20:579-580.
- Risley, HH (1981) Tribes and Castes of Bengal: Ethnographic Glossary. Parts I and II. Calcutta: Bengali Secretariate Press.
- Risley, HH (1908) *The People of India*. London: Thacker, Spink and Co.
- Roychoudhury, AK (1974) Gene differentiation among caste and linguistic populations of India. *Hum. Hered.* 24:317-323.
- Roychoudhury, AK (1987) The genetic composition of the people of eastern India. *J. Ind. Anthropol. Soc.* 16:153-170.
- Roychoudhury, AK (1984) Genetic relationships of Indian populations. In KC Malhotra and A Basu (eds): *Human Genetics and Adaptation, Vol. I*. Calcutta: Indian Statistical Inst. Press. pp 146-174.
- Smouse, PE, Long JC, and Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of metric correspondence. *Sys. Zool.* (In press).
- Steinberg, AG (1980) GM and Inv Studies on eight Iranian populations with distance measures among the six from the Caspian Littoral. *Amer. J. Phys. Anthropol.* 63:375-382.