

Genetic Heterogeneity in Northeastern India: Reflection of Tribe–Caste Continuum in the Genetic Structure

VIKRANT KUMAR,¹ DEBASHIS BASU,² AND B. MOHAN REDDY^{1*}

¹*Anthropology and Human Genetics unit, Indian Statistical Institute, Kolkata, India*

²*Anthropological Survey of India, North-Eastern Regional Center, Shillong, India*

ABSTRACT We critically examined the gene frequency data for 11 genetic markers commonly available in the literature for 22 populations of northeastern India in the light of their geographic, linguistic, and ethnic affiliations. The markers investigated were three blood groups (A₁A₂BO, MNS, and Rh), four serum proteins (KM, Gc, Hp, and Tf), and four enzyme systems (AP, AK, EsD, and Hb). The neighbor-joining tree and multidimensional scaling of the distance matrix suggest relatively high genetic differentiation among the Mongoloid groups, with probably diverse origins when compared to the Caucasoid Indo-European populations, which had probably come from relatively more homogeneous backgrounds. Broadly speaking, the pattern of population affinities conforms to the ethno-historic, linguistic, and geographic backgrounds. An interesting and important feature that emerges from this analysis is the reflection of the effect of the sociological process of a Tribe–Caste continuum on genetic structure. While on one end we have the cluster of Caucasoid caste populations, the other end consists of Mongoloid tribal groups. In between are the populations which were originally tribes but now have become semi-Hinduized caste groups, viz., Rajbanshi, Chutiya, and Ahom. These groups have currently assumed caste status and speak Indo-European languages. Therefore, one may infer that what appears to be a purely sociological phenomenon of a Tribe–Caste continuum may well reflect in their genetic structure. *Am. J. Hum. Biol.* 16:334–345, 2004.

The northeastern part of India is inhabited by numerous endogamous tribes and castes that have their own distinct social, linguistic, and biological identity. It has been hypothesized that a plethora of migrations, particularly through the northeast Indian corridor, has contributed to the present-day population of northeastern India. Ethnically speaking, most of the tribal groups are Mongoloids, whereas caste groups are either Caucasoids or show a mosaic of features of both the ethnic groups. The Mongoloids/Indo-Mongoloids have come to India from different directions at different times and perhaps earlier than the Caucasoids (Das et al., 1987). While the Mongoloids have migrated from eastern, southeastern (Rapson, 1955; Dani, 1960), and central Asian regions, the Caucasoids may have entered from western and northern boundaries of this region. While a majority of the Mongoloids are tribes affiliated with the Tibeto-Chinese linguistic family, excepting Khasi, most of the Caucasoids are caste groups and speak Indo-European languages. Although these groups have been broadly classified on the basis of language and

ethnicity, they show considerable variations within these broad categories. Both the Mongoloid and Caucasoid groups show a certain degree of differentiation within themselves in cultural and biological traits such as anthropometry, genetic markers, and dermatoglyphics (Das, 1971, 1973, 1979; Phookan, 1974; Das and Das, 1981; Roychoudhury, 1992; Bhasin and Walter, 2001). Further, the Caucasoid and Mongoloid populations have cohabited for a long time and presumably there was gene flow between them, as is evident from the conclusions drawn in many studies (Das, 1977; Das et al., 1980a,b).

A number of studies using traditional genetic markers (Bhasin et al., 1986; Singh

et al., 1986; Deka et al., 1988; Mukherjee et al., 1989; Roychoudhury, 1992) and hyper-variable loci such as VNTRs and STRs (Reddy et al., 2001a; Dutta et al., 2002, 2003; Langstieh et al., 2003) have been undertaken focusing on the variation in a certain group of populations from different areas in this region. However, no comprehensive analysis based on a matrix of genetic markers that reflects the nature, extent, and processes of genetic variation in northeastern India as a whole has yet been attempted. In the present article we discuss the results of a statistical analyses of the data on 11 genetic markers—blood groups, red cell enzyme, and serum proteins—available in the literature for 22 northeastern populations with reference to their linguistic, geographic, and ethnohistoric backgrounds. Further, a critical evaluation of the possibility of the effect of sociological process of the Tribe-Caste continuum on the

genetic structure of these populations will be made.

MATERIALS AND METHODS

Figure 1 shows the map of northeastern India along with the distribution of the studied populations. A survey of the literature suggests that the gene frequency data were available for a maximum of 11 loci only for the 22 populations from this region. Although data on some of these loci were available for a couple more populations, we had to restrict our study to these 22 populations in order to maximize the number of loci. The studied loci included three blood groups (A_1A_2BO , MNS, and Rh), four serum proteins (KM, Gc, Hp, and Tf), and four enzyme systems (AP, AK, EsD, and Hb). For the sake of brevity, we are not presenting allele frequency data for



Fig. 1. Map of northeastern India showing geographic distribution and the names of populations analyzed.

these loci (the data can be supplied upon request).

Both intra- and intergroup analyses were performed using the gene diversity measure of Nei (1987). Based on these 11 loci, average heterozygosity, G_{ST} , and pair-wise genetic distances using the modified Cavalli-Sforza distance (D_A) of Nei et al. (1983) were computed using the program Dispan (Ota, 1993). Although this genetic distance measure is not linear with evolutionary times, it is observed to be most efficient for obtaining correct phylogenetic trees under various evolutionary conditions and is least affected by small sample size (Takezaki and Nei, 1996). To assess the strength of support of data, a total of 1,000 bootstrap replication were performed and subsequently a neighbor-joining (NJ) tree (Saitou and Nei, 1987) was drawn using Mega v. 2.1 (Kumar et al., 2001) software in addition to a multidimensional scaling of the distance matrices (using SPSS software, Chicago, IL). Further, we computed the Mantel correlation between the geographic and genetic distance matrices using MANTEL v. 3.1 software (Relethford, 2003). The Harpending and Ward (1982) model was applied to understand the patterns of gene flow.

Population backgrounds

The name, ethnic status, and linguistic affiliation of the studied populations are given in Table 1. While 10 populations each from Assam and Sikkim were taken into consideration, the remaining two are from Manipur. Eight communities out of the 22 are affiliated with the Indo-European linguistic family, while the remaining groups speak Tibeto-Chinese languages. The Indo-European groups comprise the Brahmins, Kalita, Kaibarta, Muslim, and Rajbanshi of Assam, Brahmin, and Chetri of Sikkim and Brahmin of Manipur. The migrations of Indo-European-speaking people to this region started from protohistoric period (Majumdar, 1980) and continued. The Kalitas, who claim to be Rajputs or Kshatriyas, were probably one of the earliest to arrive in this region. The Brahmins are the priestly class, which occupies the highest position in the caste hierarchy and by 15th century AD the Brahmins almost reached the easternmost fringes of this region and settled in Manipur (Singh, 1963). These early settlers married Meitei women (Mongoloid group) and as a consequence the Brahmins of Manipur reveal both Caucasoid and Mongoloid morphological

features (Chakraverti and Mukherjee, 1963; Singh and Malhotra, 1970, 1971). While Chatterjee (1950) considers Meiteis a part of the Kuki-Chin (a linguistic branch of the Assam-Burmese family) group of immigrants from Burma, Pemberton (1835) describes them as the descendants of a colony of Tartar immigrants from northwest China during the 13th and 14th centuries AD. The Brahmins of Sikkim, unlike the Brahmins of Assam and Manipur, have recently migrated (around the 19th century) from Nepal, along with many other groups of Nepal.

The State of Sikkim was peopled at different points in time. Lepcha, a Mongoloid group, is considered to be the aboriginal population of this state. When and how they came to Sikkim is not known. Bhutia is another group that came to this region quite early (~ 700 years ago) from Tibet. The Sherpas of Sikkim migrated to this region about 250 years ago from Tibet through Nepal (R. Gupta, pers. commun.) and represent an offshoot of the larger Tibetan population. The remaining populations from Sikkim considered in our study migrated recently (19th century) from Nepal. Except for the Newars, the ethnic status of these populations is well established. From the historical and traditional standpoint, they are considered to be a conglomeration of many ethnic groups of India that have intermingled with a particular group, probably the Nayar of south India (Nepali, 1965), thus resulting in the present-day Newars. The non-Brahmin Indo-European populations of Assam are Kaibarta, Muslims, and Rajbanshi. The Muslims came to Assam from the 13th century to the 16th century as part of several Muslim invasions and are scattered throughout the state. The Kaibartas are the lower caste of Assam and their settlements are on the riverbanks or near sources of water. The Rajbanshi are an interesting population. While Waddel (1975) considers them a heterogeneous Mongoloid population, Risley (1915) identifies them as Kshatriya, an upper-caste Hindu population. According to Dalton (1872), they belong to the Dravidian stock that came in contact with the Mongoloids of Assam. Das (1981) considers them a conglomeration of various tribes that were converted into Hindus and in the process became admixed with certain caste groups.

The Tibeto-Chinese groups of Assam are the Ahom, Karbi, Kachari, and Sonowal. The Ahoms are said to be of Thai origin and

TABLE 1. Ethnic status and linguistic affiliations of the 22 investigated populations*

Sl. no.	Name of population	Sample size (range)	Area of sampling	Ethnic status	Linguistic affiliation
1	Brahmin ^{a-d}	106-72	Assam	Caucasoid	Indo-European
2	Kaibarta ^{a-d}	102-52	Assam	Caucasoid	Indo-European
3	Kalita ^{a-d}	105-57	Assam	Caucasoid	Indo-European
4	Muslim ^{a-d}	113-34	Assam	Caucasoid	Indo-European
5	Rajbanshi ^{a-d}	105-30	Assam	Mixed	Indo-European
6	Ahom ^{a-d}	123-62	Assam	Mongoloid	Originally Tibeto-Chinese (Siamese-Chinese, Tai), currently Indo-European
7	Chutia ^{a-d}	64-42	Assam	Mongoloid	Originally Tibeto-Chinese (Tibeto-Burman, Assam-Burmese), currently Indo-European
8	Kachari ^{a-d}	113-58	Assam	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Assam-Burmese)
9	Karbi ^{a-d}	111-60	Assam	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Assam-Burmese)
10	Sonowal ^e	107-64	Assam	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Assam-Burmese)
11	Brahmin ^f	109-62	Manipur	Caucasoid	Indo-European
12	Meiteis ^g	104-79	Manipur	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Assam-Burmese)
13	Brahmin ^h	34-25	Sikkim	Caucasoid	Indo-European
14	Chetri ^h	70-40	Sikkim	Caucasoid	Indo-European
15	Newar ^h	36-33	Sikkim	Mixed	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
16	Bhutias ^h	138-120	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
17	Gurungs ^h	39-36	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
18	Lepchas ^h	98-81	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
19	Rais ^h	44-38	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
20	Sherpas ^h	45-18	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
21	Subba ^h	52-39	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)
22	Tamang ^h	38-31	Sikkim	Mongoloid	Tibeto-Chinese (Tibeto-Burman, Tibeto-Himalayan)

^aWalter et al., 1966.

^bDas et al., 1987.

^cDeka et al., 1988.

^dMukherjee et al., 1989.

^eBhasin et al., 1986.

^fSingh et al., 1986.

*The primary source for these data was Bhasin et al., 1992, although they are verified from the original sources as quoted.

migrated around 1200 AD through the Shan state of Upper Burma (Majumdar, 1980). Chutia is the tribal group of the upper valley of Assam and are alleged to be also of Shan origin. While the Ahom has a subdivision called "Chutia," the Chutia have a section known as "Ahom." The Kachari belong to the Bodo linguistic group, which probably came from China (Thanhlira, 1969) and are presently found in Lower Assam. The Sonowal is a division of a Kachari tribe but inhabits upper Assam, unlike the latter. The Karbi is ethnically Mongoloid and linguistically placed in the Tibeto-Burman subfamily.

RESULTS

Genetic diversity among the northeastern Indian populations

The measures of genetic diversity within and between the populations (H_T , H_S , and G_{ST}) were estimated using the method of Nei (1987). G_{ST} is a ratio of the extent of gene differentiation among populations (D_{ST}) to the extent of the total genetic variation in the entire population (H_T), while H_S is a measure of the total intrapopulation heterozygosity. We computed average heterozygosity and their standard errors for each of the populations

TABLE 2. Average heterozygosity and its standard error based 10 on loci for 22 populations

Name of population	Average heterozygosity	Standard error
Brahmin (Manipur)	0.4563	0.0507
Meitei	0.4617	0.0463
Brahmin (Assam)	0.4091	0.0597
Kalita	0.4224	0.0542
Kaibarta	0.4076	0.0537
Muslim (Assam)	0.4145	0.0548
Ahom	0.4193	0.0573
Karbi	0.4236	0.0536
Kachari	0.3936	0.0519
Sonowal	0.4187	0.0493
Chutia	0.4029	0.0623
Rajbanshi	0.4005	0.0554
Lepcha	0.3807	0.0545
Bhutia	0.4056	0.0617
Sherpa	0.4337	0.0681
Tamang	0.4237	0.0484
Brahmin (Sikkim)	0.4151	0.0598
Chetri	0.4423	0.0583
Newar	0.4585	0.0575
Rais	0.4075	0.0538
Subba	0.3622	0.0578
Gurung	0.4313	0.0517

and for different categories of populations based on 11 loci. The G_{ST} value for the hemoglobin (Hb) locus is observed to be extremely high (0.2585) when all the 22 populations are considered, which suggests that this is an influential marker that may override or mask the effect of the remaining 10 loci; hence, the overall analysis may depict a picture which is overwhelmingly structured on the basis of allele frequencies of Hb. Therefore, we repeated all the analysis using 10 loci. The population-specific average heterozygosity values are given in Table 2. The ranges and averages for different categories of populations are furnished in Table 3. The values of average heterozygosity, hence H_S , is observed to be almost the same for all the categories, although the range seems to be relatively wider for Tibeto-Himalayan groups

when compared to the Assam-Burmese or Indo-European. The G_{ST} values, locus-wise as well as averages for different categories of populations, based on 10 loci are furnished in Table 4. The G_{ST} value is highest for the Gc locus (0.0517) and least at the AP locus (0.0144). While the average G_{ST} value is 0.0303 when all 22 populations are considered, it is lowest for the Assam-Burmese group (0.0155) compared to the Tibeto-Himalayan (0.0260) and Indo-European (0.0211) groups. It is interesting to note that the G_{ST} value increases only marginally (0.0269) when the Assam-Burmese groups are pooled with the Tibeto-Himalayan groups, thus forming a Tibeto-Chinese linguistic category. Since all of them are ethnically Mongoloid, this may suggest that the genetic heterogeneity among the Tibeto-Himalayan groups probably represents the overall genetic variation of the Tibeto-Chinese groups present in India.

The average G_{ST} value of our study is found to be higher than 8 of the 10 values obtained for regional populations of India (Table 5). It is, however, pertinent to note that these G_{ST} values are not strictly comparable, as the different studies were neither based on a uniform set of markers nor represented by the similar number of the populations. Nevertheless, we note that at least one of the two G_{ST} values (Chakraborty et al., 1977) that are higher than the present one is primarily because of the inclusion of certain influential markers such as Hb, LDH, and Cp. This situation was also reflected in our study when we considered the Hb locus for computing G_{ST} . Further, the average G_{ST} value from various studies in India using traditional genetic markers turns out to be less than 1.5% (Reddy et al., 2001b). Therefore, the G_{ST} value obtained for the northeastern populations in general (3%) and particularly for the Tibeto-Chinese groups (2.7%) can be considered as

TABLE 3. Range and the average values of heterozygosity for different groups of populations

Groups of populations	No. of populations	Range	Average heterozygosity
All groups	22	0.3622–0.4617	0.4109
Indo-European*	8	0.4005–0.4563	0.4149
Tibeto-Chinese**	14	0.3622–0.4617	0.4086
Tibeto-Himalayan	8	0.3622–0.4585	0.4036
Assam-Burmese	5	0.3936–0.4617	0.4152

*All Caucasoid groups and Rajbanshi.

**All Mongoloid groups and Newar.

TABLE 4. Locus-wise and average gene diversity (G_{ST}) values among different categories

Locus	All groups (22)	Indo-European (8)	Tibeto-Chinese (14)	Tibeto-Himalayan (8)	Assam-Burmese (5)
A ₁ A ₂ BO	0.0156	0.0162	0.0130	0.0121	0.0083
MNS	0.0234	0.0182	0.0173	0.0165	0.0067
Rh	0.0406	0.0324	0.0380	0.0409	0.0231
Ge	0.0517	0.0326	0.0461	0.0604	0.0217
Hp	0.0161	0.0109	0.0191	0.0299	0.0036
Tf	0.0375	0.0357	0.0378	0.0185	0.0204
AP	0.0144	0.0043	0.0172	0.0244	0.0039
AK	0.0256	0.0044	0.0277	0.0337	0.0206
EsD	0.0161	0.0086	0.0189	0.0119	0.0159
KM	0.0516	0.0313	0.0248	0.0064	0.0351
Average	0.0303	0.0211	0.0269	0.0260	0.0155

*Numbers in parentheses are the number of populations in each category.

TABLE 5. G_{ST}/F_{ST} values of different studies from different regions of India

Region	No. of populations	No. of loci	G_{ST}/F_{ST}	References
Northern India (Himachal Pradesh)	11	20	0.013	Papiha et al., 1996
Northern India (Himachal Pradesh)	5	24	0.013	Papiha et al., 1984
Northern India	6	18	0.011	Roychoudhury, 1974
Northwest India	14	11	0.013	Papiha et al., 1982
Western India	4	10	0.012	Roychoudhury, 1974
Western India (Maharashtra)	21	11	0.044	Chakraborty et al., 1977
Northcentral India (Uttar Pradesh)	6	24	0.009	Lanchbury et al., 1996
Central India	16	12	0.016	Das et al., 1996
Eastern India (West Bengal)	10	12	0.024	Chakraborty et al., 1986
Southern India	9	7	0.041	Sirajuddin et al., 1994

suggesting relatively high differentiation of the Mongoloid groups.

Genetic affinities

Although we have drawn both the NJ and the UPGMA trees based on Nei's D_A distance matrix, we present here only the NJ dendrogram (Fig. 2), as both of them reflect a similar pattern of population relationships. Further, the bootstrap values of the NJ tree are relatively much higher, providing greater confidence for the pattern of relationships depicted. The Gaddi-Brahmins, a non-northeastern tribal population from Himachal Pradesh, were used as an outgroup for rooting the tree. Broadly speaking, this tree depicts two major clusters, the first consisting of seven Caucasoid populations from Sikkim and Assam, including the outgroup. The second cluster consists of 13 populations that are all Mongoloids, excepting the Brahmins of Manipur and Newars of Sikkim. Within this second cluster, two broad subclusters can be discerned, one formed by the Mongoloid groups of Assam (Karbi, Kachari, and Sonowal)

affiliated with the Assam-Burmese linguistic group, and the other by the Brahmins and Meiteis of Manipur and Newars of Sikkim. The clustering of Brahmins of Manipur with the Meiteis (both of them are caste populations) is not surprising, as there has been considerable admixture between them (Chakraverti and Mukherjee, 1963; Singh and Malhotra, 1970, 1971). However, the subclustering of Newar with the groups of Manipur is intriguing. Although the ethnic status of Newars is mixed and is geographically far apart from the Manipuri groups, it is considered to be a caste population, as are the two groups from Manipur. The rest of the populations within this cluster is the Mongoloids from Sikkim. The position of Chutia, Ahom, and Rajbanshis is ambiguous in the tree, which, as explained above, is consistent with their current sociocultural status. Overall, the pattern of genetic relationships is consistent with the ethnic and linguistic affiliations of the populations, and to an extent also reflects congruence with the geographic division of the constituent groups within an ethnic group.

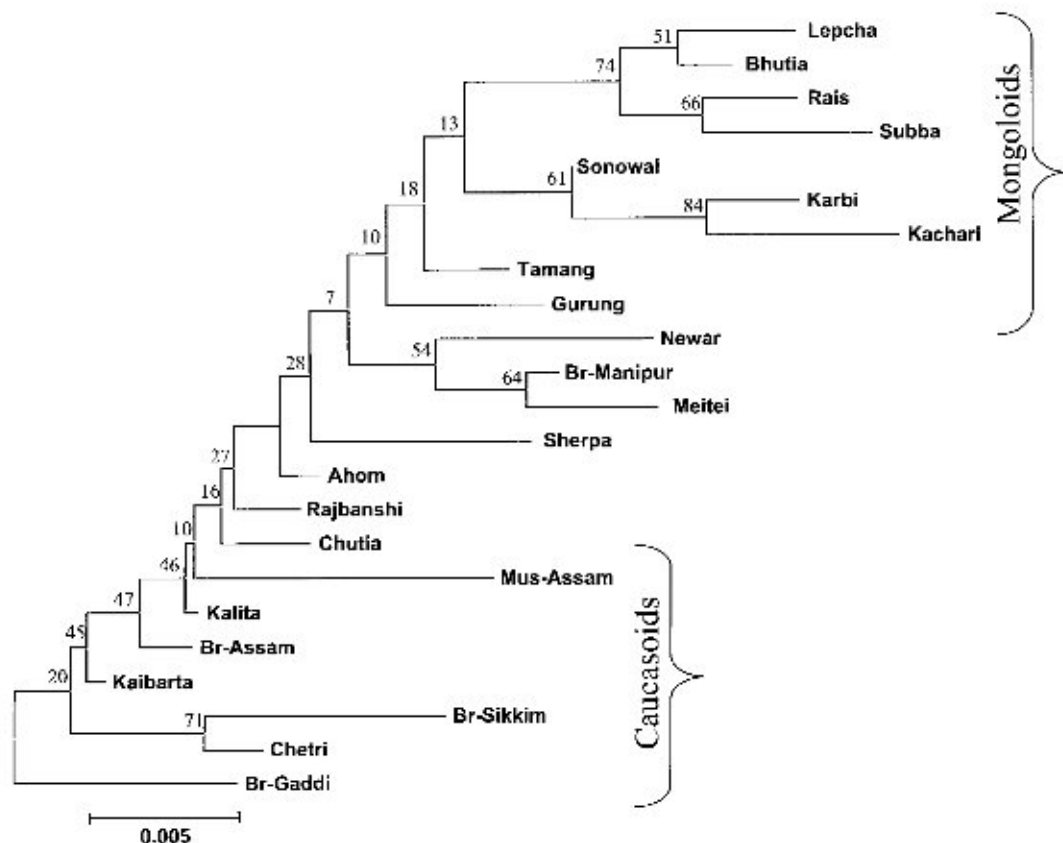


Fig. 2. Neighbor-joining tree constructed on Nei's D_A distance matrix based on 10 genetic markers among the 22 populations. Br, Brahmin; Mus, Muslim. (Source for Br-Gaddi: Chahal et al., 1982; Singh et al., 1982; Papiha et al., 1983).

Multidimensional scaling

The limitations inherent in the dendrogram resulting from the imposition of bifurcation topology can be partially overcome by using multidimensional scaling (MDS) for depicting relationships among populations based on frequency or distance data. Although MDS tends to generate conclusions similar to those obtained by the tree, they are likely to be more faithful descriptors of the data than the trees when there is considerable genetic exchange between close geographic neighbors (Cavalli-Sforza et al., 1994). In Figure 3, we present a two-dimensional plot of populations based on multidimensional scaling of Nei's D_A distance matrix. The two dimensions account for 87% of the total variation, with the stress value of 16% suggesting a reasonable fit. Broadly speaking, three major constellations of populations can be

discerned. The first comprises all the populations of Sikkim, except Brahmin and Chetri. They are differentiated from all other populations, basically on the second axis. Within this there are two subclusters, one formed by Bhutias, Lepchas, Rais, and Subba, and the second by Newars, Sherpas, and Gurungs. The Chetri of Sikkim and Brahmins, Kalibarta, Kalita, and Muslim of Assam form a second cluster; these populations not only share the same linguistic family but also have the same Caucasoid ethnic affiliation. However, the Brahmins of Sikkim are placed as outliers. The loose conglomeration of Brahmins of Manipur, Meiteis, Sonowal, Kachari, and Karbi forms the third cluster. Within this the Kachari and the Karbi are depicted as outliers. Although the Kachari and Sonowal are two divisions of the same tribe, they are not close

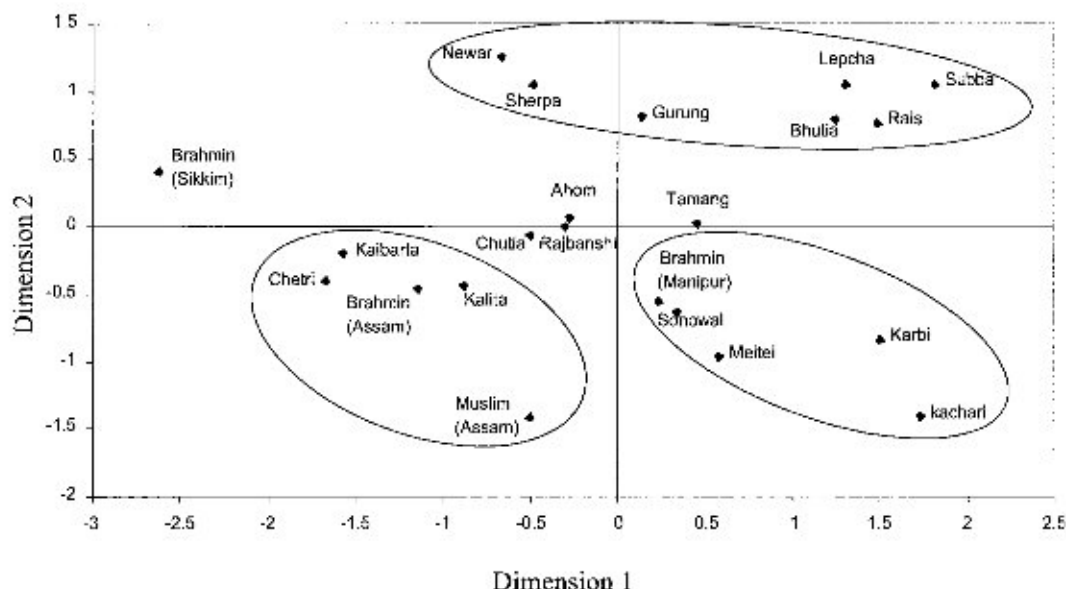


Fig. 3. Plot of 22 populations on the first two dimensions derived from the multidimensional scaling of Nei's D_A distance matrix.

to each other in the multidimensional space. This may in part be due to their geographic isolation, which might have led to genetic isolation between them (Das et al., 1987). The unexpected position of the Karbis reflects their peculiar linguistic status. According to some, the Karbi language shows a mosaic of Austro-Asiatic and Tibeto-Burman languages, while others are of the opinion that it shares similarity with the Naga and Bodo dialects, which belong to the Tibeto-Burman family. In fact, Das et al. (1980a) in their study found the Karbis to be significantly different from the members of the Bodo group to which the Kachari and Sonowal belong. The Chutia, Ahom, and Rajbanshis constitute a separate and compact cluster positioned in the center of MDS plot, while a constellation of caste populations, viz., Brahmin and Chetri of Sikkim, Kalita, Kaibarta, and Brahmin of Assam, is on one side and tribal groups of Assam (Sonowal, Karbi, and Kachari) and Sikkim (Tamang, Gurung, Bhutia, Rais, Subba, and Lepcha) are on the other.

To test the congruence between geographic and genetic distances, the Mantel correlation was obtained between the genetic and geographic distance (Great Circle distances) matrices. The Mantel correlation was observed to be very low and statistically not different

from zero (0.0076; $P = 0.5060$), although in the NJ tree and in the MDS plot we find that there is a tendency for populations of a particular region to cluster together. This can be explained by the fact that there is considerable ethnic heterogeneity within a geographic region and this heterogeneity is simultaneously reflected in the way the population clusters are formed. Furthermore, the populations are not distributed in a continuum, but make four disjoint geographic clusters (Fig. 1) and within each of them the geographic heterogeneity is minimal. A significant Mantel correlation is expected only when the genetic structure of populations is framed as well as maintained primarily due to isolation by distance, which is probably not the case in the present study.

Population structure and patterns of gene flow

To understand the local genetic differentiation and the underlying processes responsible for the present pattern of population structure among the groups of northeastern India, we used the Harpending and Ward (1982) model, somewhat analogous to the Wright's island model. According to this model, given the uniform systematic pressure (gene flow) from outside, the average genetic

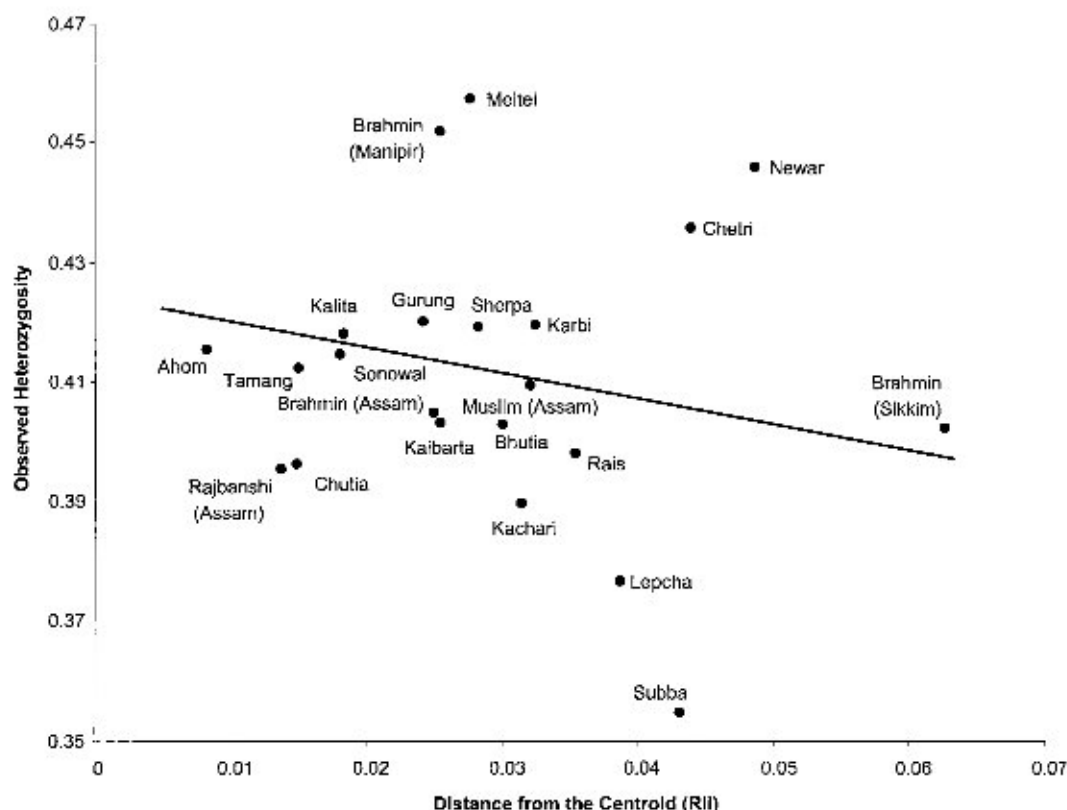


Fig. 4. Regression plot of the average heterozygosity vs. distance of the 22 populations to the centroid.

heterogeneity within a population is negatively correlated with its 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 higher than average heterozygosity than predicted by the regression model and lie above the theoretical regression line. Conversely, populations experiencing isolation and less than average gene flow from outside show lower values of heterozygosity and lie below the regression line. A regression plot of average heterozygosity versus distance of each of the populations to the centroid is given in Figure 4, along with the theoretical regression line. In the regression plot, the Meiteis and Brahmins of Manipur along with Newar and Chetri of Sikkim are outliers above the theoretical regression line. The position of Brahmins who were migrants from Bengal and Meitei of Manipur above the regression line may be a reflection of gene flow between them, as discussed above. Similarly, the posi-

tion of Newar as an outlier may account for its ambiguous ethnic position of being a mixed population. On the other hand, Subba and Lepchas of Sikkim are extreme outliers below the regression line and are probably affected by genetic drift due to their isolation. The remaining populations are placed either on the line or in close proximity.

DISCUSSION

The pattern of clustering of populations as depicted in the NJ tree as well as in the MDS plot is consistent with the broad ethnic and linguistic affiliations of the groups within which the geographic affinities are subtly reflected. For example, the Mongoloid populations of Assam (Assam-Burmese) and Sikkim (Tibeto-Himalayan) form two distinct sub-clusters, albeit within a major cluster. However, the Caucasoid Indo-European groups of Assam and Sikkim, with a relatively more homogenous background, tend to form a

single cluster. This may suggest a diverse origin and/or considerable differentiation of the Mongoloid groups from the two geographic areas within the northeastern region. These findings may confirm the hypotheses that Mongoloids have entered northeastern India at different points of time by different routes and therefore might represent different parental stocks. On the other hand, the Caucasoid populations, except for the Brahmins and Chetri of Sikkim, have migrated from the western route of northeastern India. As far as the Brahmins and Chetri of Sikkim are concerned, they came from Northern India to Nepal and then to northeastern India around the 19th century.

The most important feature that can be discerned from the foregoing analysis is the possibility that the change in sociocultural aspects of a group may affect its biological makeup. Bose (1941), in his study of the Hindu method of tribal absorption, has elucidated that tribes over a period of time assume the characteristics and status of castes and this process of transformation of a tribe into a caste results in a Tribe-Caste continuum. A few such cases he cited are Bhumij, Kharia, Bauris, and RajGonds. This sociological concept of a Tribe-Caste continuum postulates that one end of the continuum is formed by caste populations, while the tribal populations constitute the other end. In between are the populations who were once tribes but gradually adopted the attributes of the caste population and ultimately became absorbed as an integral part of a caste system, albeit at the lowest rung of caste hierarchy. However, they do maintain a separate group identity and over a period of time become distinct castes or Jaties. Risley (1981) discerned four processes by which a tribe is converted into a caste and gradually becomes a distinct caste group. He described the case of Rajbanshi, which claims to be Kshatriya, although a majority of them are Koch. Similar processes were reported from other groups such as the Dimasa (a Kachari group) of Tripura, Jantia of Jaintipur, and Koch of Cooch Behar (K. Mukherjee, pers. commun.). The dwindling of Kachari groups from a large number of about 30 to 10 at present is said to be a result of this process. These studies suggest that the process of a Tribe-Caste continuum has an affect on the sociological dimensions of the tribe, but the probable biological consequences of this process have not been explored. Our results suggest that there are

strong indications of a biological effect of this process. The process of assimilation of some of the tribes in the caste hierarchy is probably reflected in the broad constellation of populations in the multidimensional plot (Fig. 3); on the one end it is the constellation of populations subscribing to the caste system, while the other end of the plot is the constellation of tribal groups. In between lie groups like the Rajbanshi, Chutiya, and Ahom, which were supposed to have undergone the process discussed above. For example, the Chutiya and Ahom tribes, which were originally affiliated with the Tibeto-Chinese linguistic family, are now semi-Hinduized caste groups speaking Indo-European languages. The ethnic status of Rajbanshi is not clear, and different opinions are advanced regarding this (Dalton, 1872; Risley, 1915; Waddel, 1975; Das et al., 1987). Rajbanshis are hypothesized to be a mixture of different tribal groups (Rabhas, Tiwas, Kacharis, Garos, Karbis, etc.) that were converted to Hinduism and in the process became admixed with certain Caucasoid caste populations (Das et al., 1987). The Ahom and the Rajbanshi, who were originally tribal groups, held a dominant position in terms of economy as well as power. Both of these groups tried to perpetuate their domination and obtain social sanction for it by becoming an upper caste group. In this process, they might have tried to emulate the cultural and social ideologies of the neighboring upper caste groups and developed marital ties with them. The position of the Rajbanshi and Ahom in the plot is consistent with this conjecture. It may be pertinent to note that Das (1981) suggests that Rajbanshi is a fit example of a Tribe-Caste continuum not only in their sociocultural aspects but also in their biological makeup. Therefore, one may infer that what appears to be a purely sociological phenomenon of a Tribe-Caste continuum may well be reflected in their genetic structure.

LITERATURE CITED

- Bhasin MK, Walter H. 2001. Genetics of castes and tribes of India. Delhi: Kamla-Raj Enterprises.
- Bhasin MK, Walter H, Chahal SMS, Bhardwaj V, Sudhakar K, Danker-Hopfe H, Dannewitz A, Singh IP, Shil AP, Sharma MB, Wadhavan D. 1986. Biology of the people of Sikkim, India. 1. Studies on the variability of genetic markers. *Z Morph Anthropol* 77:49-86.
- Bhasin MK, Walter H, Danker-Hopfe H. 1992. The distribution of genetical, morphological and behavioural traits among the peoples of Indian origin. Delhi: Kamla-Raj Enterprises.

- Bose NK. 1941. The Hindu method of tribal absorption. *Sci Culture* 7:188-194.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton, NJ: Princeton University Press.
- Chahal SMS, Papiha SS, Roberts DF, Singh IP. 1982. Serological and biochemical variation in the Gaddi tribe of Himachal Pradesh, India. *Z Morph Anthrop* 73:197-208.
- Chakraborty R, Chakravarti B, Malhotra KC. 1977. Variation in the allele frequencies among caste groups of the Dhangar of Maharashtra, India: an analysis with Wright's F_{ST} statistic. *Ann Hum Biol* 4:275-280.
- Chakraborty R, Walter H, Mukherjee BN, Malhotra KC, Sauber P, Banerjee S, Roy M. 1986. Gene differentiation among 10 endogamous groups of West Bengal, India. *Am J Phys Anthropol* 71:295-309.
- Chakravarti MR, Mukherjee DP. 1963. Dermatoglyphic study of tribes and castes of Nagaland and Manipur states. *Bull Anthropol Survey India* 11:233-234.
- Chatterjee SK. 1950. Kirata-Jana-Krit (the Indo-Mongoloids: their contribution to the history of the culture of India). *J R Soc Bengal* 16:2.
- Dalton ET. 1872. The descriptive ethnology of Bengal. Calcutta: Govt. Printing.
- Dani AH. 1960. Prehistory and protohistory of Eastern India. Calcutta: Firma KL Mukhopadhyay.
- Das BM. 1971. Medullary structure of head hair in some Caucasoid and Mongoloid populations of Assam India. *Z Morph Anthropol* 53:102-109.
- Das BM. 1973. Caste and head hair. *Indian Anthropol* 3:135-138.
- Das BM. 1977. Inheritance of cerumen type and its importance in population study. *Man India* 57:305-312.
- Das BM. 1979. Physical variation in three Assamese caste. *Anthrop Anz* 37:204-210.
- Das BM. 1981. Microevolution. New Delhi: Concept Publishing House.
- Das BM, Das R. 1981. Intra- and intertribal variation in palmar flexion crease: A population study. *Indian J Phys Anthropol Hum Genet* 7:173-178.
- Das BM, Deka R, Das R. 1980a. Haemoglobin E in six populations of Assam. *J Indian Anthropol Soc* 15:153-156.
- Das BM, Deka R, Das R. 1980b. Haemoglobin E in North-East India. A review. *Indian J Phys Anthropol Hum Genet* 6:55-82.
- Das BM, Walter H, Gilbert K, Lindenberg P, Malhotra KC, Mukherjee BN, Deka R, Chakraborty R. 1987. Genetic variation of five blood polymorphisms in ten populations of Assam India. *Int J Anthropol* 2:325-340.
- Das K, Malhotra KC, Mukherjee BN, Walter H, Majumdar PP, Papiha SS. 1996. Population structure and genetic differentiation among 16 tribal populations of Central India. *Hum Biol* 68:679-705.
- Deka R, Reddy AP, Mukherjee BN, Das BM, Banerjee S, Roy M, Dey B, Malhotra KC, Walter H. 1988. Hemoglobin E distribution in ten endogamous population groups of Assam India. *Hum Hered* 38:261-266.
- Dutta R, Reddy BM, Chattopadhyay P, Kashyap VK, Sun G, Deka R. 2002. Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum Biol* 74:33-49.
- Dutta R, Kashyap VK, Trivedi R, Vasulu TS. 2003. Genomic diversity in North-East, India: Genetic relationship among 5 ethnic populations from Manipur based in VNTR Polymorphism. *Int J Hum Genet* 3:21-27.
- Harpending HC, Ward R. 1982. Chemical systematics and human evolution. In: Nitecki M, editor. *Biochemical aspects of evolutionary biology*. Chicago: University Press. p 213-256.
- Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244-5.
- Lanchbury JS, Agarwal SS, Papiha SS. 1996. Genetic differentiation and population structure of some occupational caste groups in Uttar Pradesh, India. *Hum Biol* 68:655-678.
- Langstieh BT, Reddy BM, Thangaraj K, Singh L. 2003. Genetic relationships among the tribes of Meghalaya, India: matriliney and reduced microsatellite diversity. *Hum Biol* (in press).
- Majumdar DN. 1980. North east India—a profile. In: Sharma TC, Majumdar DN, editors. *Eastern Himalayas: a study on anthropology and tribalism*. N. Delhi: Cosmo.
- Mukherjee BN, Malhotra KC, Roy M, Banerjee S, Walter H, Chakraborty R. 1989. Genetic heterogeneity and population structure in eastern India: red cell enzyme variability in ten Assamese populations. *Z Morph Anthropol* 77:287-296.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* 19:153-170.
- Nepali GS. 1965. The Newar: an ethno-sociological study of a Himalayan community. Bombay: United Asia.
- Ota T. 1993. Dispan: genetic distance and phylogenetic analysis. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania.
- Papiha SS, Mukherjee BN, Chahal SMS, Malhotra KC, Roberts DF. 1982. Genetic heterogeneity and population structure in northwest India. *Ann Hum Biol* 9:235-251.
- Papiha SS, White I, Roberts DF. 1983. Some genetic implications of isoelectric focusing of human red cell phosphoglucomutase (PGM₁) and serum protein group specific components (Gc): genetic diversity in the populations of Himachal Pradesh, India. *Hum Genet* 63:67-72.
- Papiha SS, Chahal SMS, Roberts DF, Murty KJ, Gupta RL, Sidhu LS. 1984. Genetic differentiation and population structure in Kinnair district, Himachal Pradesh, India. *Hum Biol* 56:231-253.
- Papiha SS, Chahal SMS, Mastana SS. 1996. Variability of genetic markers in Himachal Pradesh, India: variation among the subpopulations. *Hum Biol* 68:629-654.
- Pemberton RB. 1835. Report on the Eastern Frontier of British India. Guwahati, India: Department of Historical and Antiquarian studies.
- Phookan MN. 1974. The Kachari; a study of intra-tribal relations in respect to anthropometry, ABO blood groups and taste sensitivity. Ph.D. Thesis, Dibrugarh University, Dibrugarh, Assam.
- Rapson EJ. 1955. People and languages. In: Rapson EJ, editor. *Cambridge history of India, vol. 1. Ancient India*. Delhi: S Chand. p 33-57.
- Reddy BM, Dutta R, Langstieh BT, Kashyap VK. 2001a. Diversity at the three tetrameric STR loci in a substructured Golla caste population of Southern Andhra Pradesh in comparison to other Indian populations. *Int J Hum Genet* 1:101-108.
- Reddy BM, Sun G, Luis JR, Crawford MH, Hemam NS, Deka R. 2001b. Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of Southern Andhra Pradesh, India. *Hum Biol* 73:175-190.

- Relethford JH. 2003. Mantel 3.1: a program for comparison of two distance matrices. Department of Anthropology, SUNY College at Oneonta, Oneonta, New York.
- Risley HH. 1915. The people of India. Calcutta: Thacker Spink.
- Risley HH. 1981. The tribes and castes of Bengal. Calcutta: Firma KL Mukhopadhyay.
- Roychoudhury AK. 1974. Gene differentiation among caste and linguistic populations of India. *Hum Hered* 24:317-322.
- Roychoudhury AK. 1992. Genetic relationships of the populations in eastern India. *Ann Hum Biol* 19:489-501.
- Saitou N, Nei M. 1987. The neighbor joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Singh KB. 1963. Manipuri vaisnavism. *Soc Bull Bombay* 12:2.
- Singh NR, Malhotra KC. 1970. Ear lobe attachment among the Brahmin and Meiteis of Manipur. *India Bull Deccan College Res Inst* 29:1-4.
- Singh NR, Malhotra KC. 1971. Hand clasping and arm folding among the Manipuris, India. *Hum Hered* 21:203-207.
- Singh IP, Bhasin MK, Walter H, Hilling M, Bhasin V, Singh R. 1982. Genetic studies of Pangwalas, Transhumant and settled Gaddis. 2. Serum protein and red cell enzyme polymorphisms. *Z Morph Anthropol* 73:175-195.
- Singh KS, Mukherjee BN, Walter H, Lindenberg P, Gilbert K, Dannewitz A, Malhotra KC, Banerjee S, Roy M, Dey B. 1986. Genetic marker among Meiteis and Brahmins of Manipur, India. *Hum Hered* 36:177-187.
- Sirajuddin SM, Duggirala R, Crawford MH. 1994. Population structure of Chenchu and other South Indian tribal groups: relationships between genetic, anthropometric, dermatoglyphic, geographic and linguistic distances. *Hum Biol* 66:865-884.
- Takezaki N, Nei M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389-399.
- Thanhlira R. 1969. The tribes of North cachar hills. In: Barkataki S, editor. *Tribe of Assam*. N. Delhi: National Book Trust.
- Waddell LA. 1975. The tribes of the Brahmaputra Valley. A contribution on their physical type and affinities. N. Delhi: Sanskaran Prakashak.
- Walter H, Mukherjee BN, Gilbert K, Lindenberg P, Dannewitz A, Malhotra KC, Das BM, Deka R. 1986. Investigations on the variability of haptoglobin, transferrin and Gc polymorphisms in Assam, India. *Hum Hered* 36:388-396.