Z. Morph. Anthrop.	77	3	287–296	Stuttgart, April 1989

Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta, India; Dept. of Human Biology / Physical Anthropology, University of Bremen, W. Germany; and Center for Demography and Population Genetics, University of Texas, Houston, USA

Genetic heterogeneity and population structure in eastern India: Red cell enzyme variability in ten Assamese populations

By B. N. Mukherjee, K. C. Malhotra, M. Roy, S. Banerjee, H. Walter and R. Chakraborty

With 2 figures and 4 tables in the text

Summary: This paper is a part of the genetic study of the people of Assam (eastern India), initiated by the Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta, India, and the Dept. of Human Biology / Physical Anthropology, University of Bremen, W. Germany. The results of 1. allele distribution of five red cell enzyme polymorphisms in ten Assamese populations, 2. heterogeneity of allele frequencies and extent of gene differentiation among these populations, and 3. standard genetic distances are presented here. A total of 1024 blood samples was screened for aP, E D, AK, ADA and LDH enzyme systems for Brahmins, Kalitas, Kaibartas, Rajbanshis, Muslims, Ahoms, Chutiyas, Kacharis, Karbis (Mikirs) and Sonowals, of which the latter three are tribes. The gene diversity (F_{ST}) is smallest (0.0035) for p^a and highest (0.1604) for Hb^F. The total F_{ST} value (0.0399 ± 0.0141) appears to be statistically significant. From distance analysis two major clusters with sub-clusters in each are visible, which are in conformity with the ethnohistory of these populations.

Zusammenfassung: Diese Untersuchung ist Teil eines populationsgenetischen Forschungsprojektes an der Bevölkerung von Assam (Ostindien), das gemeinsam von der Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta, Indien, sowie dem Dept. f. Humanbiologie/Anthropologie der Universität Bremen, Bundesrepublik Deutschland, initiiert und durchgeführt wurde. Hier werden 1. die Allelenverteilungen für fünf polymorphe Enzymsysteme in zehn Populationen, 2. die Heterogenität der Allelenfrequenzen und das Ausmaß der genetischen Differenzierung zwischen diesen Populationen sowie 3. die genetischen Abstände zwischen ihnen vorgestellt und diskutiert. Insgesamt konnten 1024 Individuen bezüglich der polymorphen Systeme aP, EsD, AK, ADA und LDH typisiert werden. Die untersuchten Populationen sind Brahmins, Kalitas, Kaibartas, Rajbanshis, Muslims, Ahoms, Chutiyas, Kacharis, Karbis (Mikirs) und Sonowals; bei den drei letzteren handelt es sich um Stammesbevölkerungen. Die gene diversity (F_{ST}) ist für p^a am kleinsten (0.0035), für Hb^F am größten (0.1604). Der Gesamt-F_{ST}-Wert (0.0399 ± 0.0141) dürfte statistisch signifikant sein. Die genetische Abstandsanalyse läßt zwei Hauptcluster erkennen, wobei jedes Subcluster enthält. Diese Cluster stehen mit der Bevölkerungsgeschichte der untersuchten Populationen im Einklang.

Introduction

Studies on the distribution of the marker alleles have significantly contributed to the understanding of both micro- and macroevolutionary processes of man. The Indian subcontinent with its large number of caste, religious, linguistic (about 30 000) and tribal

(about 427) groups provides an unique situation for the study of ongoing biological differentiation. In this paper we report the results of 1. the distribution of five red cell enzyme systems in ten Assamese populations belonging to Mongoloid and Caucasoid origin, followed by intergroup comparisons, and 2. the analysis of heterogeneity of allele frequencies as well as extent of gene differentiation among the population groups based on eight enzyme and protein marker systems; some of them were reported earlier. Standard genetic distances have been computed to examine the relationship between the ten Assamese populations under study.

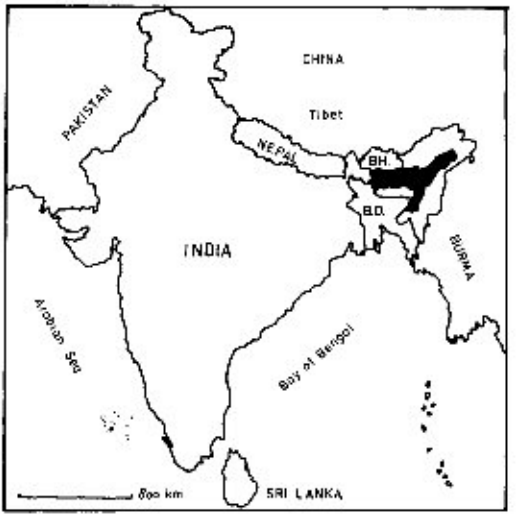

Fig. 1. Location of Assam (black). BH = Bhutan, B. D. = Bangla Desh.

Materials and methods

The state of Assam is situated in the north eastern part of India (Fig. 1). Assam consists of a large number of caste and tribal groups with various racial affiliations. Assam is a meeting place of two broad racial groups of mankind, namely *Caucasoids* and *Mongoloids;* even the *Australoid* stock is tracable (Das et al. 1987; Walter et al. 1987). Though, it appears that in the Assamese populations a lot of admixture has taken place, but generally speaking, on the basis of anthropometrical and morphological traits, the Hindu caste groups and Muslims can be considered as of Caucasoid origin, while the tribes come mostly from the Mongoloid stock. Keeping in view this interesting racial composition of the people of Assam, a survey of a number of biochemical and serological genetic markers was conducted in 1984 on ten population groups of the state. The populations were selected in such a way that they represent populations of both Caucasoid and Mongoloid as well as of known mixed origin. The populations studied are: *1. Brahmins, 2. Kalitas, 3. Kaibartas, 4. Rajbanshis* (caste group which belongs to the Indo-European language family); *5. Muslims* (a religious group which belongs to the same language family). All these groups can be considered as Caucasoids or at least to have a strong Caucasoid component. *6. Ahoms* and *7. Chutiyas,* which are very old populations of

Assam; both belong linguistically to the Tibeto-Chinese language family. Ahoms are originally a Shan or Thai people. Both the populations are Mongoloid in their physical appearance and Hindu by religion. *8. Kacharis, 9. Karbis (Mikirs)* and *10. Sonowals;* these three groups are tribes and have a lot of Mongoloid admixture. They belong, too, to the Tibeto-Chinese language family. For further ethnohistory of these poulations we refer to DAS (1987) and WALTER et al. (1986).

The enzym system studied are: *1. acid phosphatase (aP), 2. adenylate kinase (AK), 3. esterase D (EsD), 4. adenosinedeaminase (ADA),* and *5. lactate dehydrogenase (LDH).* Screening of these markers was done by starch gel electrophoresis following the techniques described by HARRIS & HOPKINSON (1977). 1024 blood samples from individuals of both sexes were collected from different regions of the Brahmaputra valley of Assam (Fig. 1). The Ahom, Chutiya and Sonowal samples were drawn from Dibrugarh in the eastern part of Upper Assam; the rest of the populations was surveyed from the area around Guwahati in the western part of lower Assam. The numbers of blood samples collected by finger pricking from each group are as follows: Brahmins (100), Kalitas (105), Kaibartas (100), Ahoms (120), Rajbanshis (105), Muslims (105), Kacharis (113), Chutiyas (58), Karbis (111), and Sonowals (107). Immediately after collection, the blood samples were kept in the refrigerator till dispatched by air to the Indian Statistical Institute, Calcutta, where all the tests were done.

Results and discussion

Table 1 and 2 show the phenotypes with X^2-values and allele frequencies of five enzyme systems, respectively. The Hardy-Weinberg test for equilibrium shows a good agreement between observed and expected values except for Kaibartas and Muslims in the ADA system; this may be due to small sample sizes.

Acid phosphatase

Three common phenotypes of aP namely A, AB and B are present in all the groups. The most frequent phenotype is aP B, which varies between 53.0 % in Kaibartas and 65.0 % in Brahmins. This result can be compared well with similar populations in West Bengal, where the frequencies of aP B in the Brahmins and Kaibartas are about 53.0 % and 54.0 %, respectively (MUKHERJEE et al., 1987). The phenotype aP A ranges between 4.16 % (Ahoms) and 13.33 % (Kalitas). The Rajbanshis show about 6.0 % aP A, which is very similar to the Rajbanshis from West Bengal, who show an incidence of 7.0 %. The p^a allele is generally low in most of the Indian populations and varies between 0.200 and 0.400. In this study it is lowest in the Brahmins (0.205) and highest in the Kaibartas (0.300).

Esterase D

The EsD exhibits three common phenotypes in most of the world populations. In this study type EsD 1 ranges from 35.5 % (Sonowals) to 67.7 % (Brahmins). EsD 2 is maximum in Kaibartas (17.12 %) and minimum in Brahmins (6.06 %). The EsD^1 frequency is

Table 1. Distribution of phenotypes of four enzyme systems in ten Assamese populations.

	Acid phosphatase					Esterase D				
	No.	A	AB	B	χ^2	No.	1	2–1	2	χ^2
Brahmins	100	6.00	29	65		99	67	26	6	
		6.00	29.00	65.00	1.216		67.68	26.26	6.06	2.326
Kalitas	105	14	35	56		105	57	38	10	
		13.33	33.33	53.33	4.471		54.28	36.19	9.53	0.844
Kaibartas	100	5	42	53		101	59	32	10	
		5.00	42.00	53.00	0.837		58.42	31.68	9.90	2.963
Ahoms	120	5	43	72		110	50	50	10	
		4.16	35.83	60.00	0.204		45.45	45.45	9.10	0.249
Rajbanshis	105	6	43	56		103	51	42	10	
		5.71	40.95	53.34	0.368		49.51	40.78	9.71	0.098
Muslims	105	10	36	59		104	53	41	10	
		9.52	34.28	56.20	1.598		50.96	39.42	9.62	0.249
Kacharis	113	9	39	65		112	61	36	15	
		7.96	34.51	57.53	0.817		54.46	32.15	13.39	5.756
Chutiyas	58	5	23	30		64	38	21	5	
		8.62	39.65	51.73	0.039		59.38	32.81	7.81	0.720
Karbis	111	9	37	65		111	41	51	19	
		8.11	33.33	58.56	1.241		36.94	45.94	17.12	0.210
Sonowals	107	10	36	61		107	38	56	13	
		9.35	33.64	57.01	1.789		35.51	52.34	12.15	1.229

highest in Brahmins (0.808) and lowest in Karbis (0.599). The present EsD allele frequencies among Brahmins, Kaibartas and Rajbanshis are comparable with that of their counterparts in West Bengal. The EsD^1 allele in Brahmins, Kaibartas and Rajbanshis of West Bengal exhibits about 0.78, 0.79 and 0.70, respectively, against 0.80, 0.74, and 0.72 in the corresponding populations of Assam, which gives an indication of their common ancestral stock. In Indian populations the EsD^2 allele depicts a wide range between 0.17 and 0.50, which covers the values observed in the populations under study.

Adenylate Kinase

In this study most AK 1 is found in all the populations, being maximum (96.3 %) in Karbis and minimum in Muslim (82.3 %). One case of AK 2 in each of Brahmins, Kalitas, Kaibartas and Sonowals have also been observed, which supports the rather low incidence of this AK phenotype in Indian populations. Consequently the AK^2 allele is also low in these populations, being 0.018 in Karbis and 0.089 in Kaibartas.

Adenosine deaminase

The ADA system is not yet extensively studied in Indian populations. Due to small sample size, some groups are not taken into consideration for discussion. Here only references are made for Brahmins, Ahoms, Chutiyas, and Sonowals. The distribution of

	Adenosine deaminase					Adenylate kinase				
	No.	1	2–1	2	χ^2	No	1	2–1	2	χ^2
Brahmins	55	21	27	7		106	89	16	1	
		38.18	49.10	12.72	0.137		83.96	15.09	0.95	0.087
Kalitas	24	11	12	1		103	87	15	1	
		45.83	50.00	4.17	1.059		84.47	14.56	0.97	0.151
Kaibartas	34	6	25	3		101	84	16	1	
		17.64	73.52	8.82	7.903		83.17	15.84	0.99	0.059
Ahoms	110	71	35	4		97	89	8	0	
		64.55	31.82	3.63	0.015		91.75	8.25	0.00	0.179
Rajbanshis	6	1	5	0	—	103	93	10	0	
		—	—	—			90.29	9.71	0.00	0.268
Muslims	34	0	34	0		34	28	6	0	
		—	—	—			82.35	17.65	0.00	0.318
Kacharis	2	0	2	0		96	92	4	0	
		—	—	—	—		95.83	4.17	0.00	0.043
Chutiyas	49	30	18	1		57	54	3	0	
		61.23	36.73	2.04	0.838		94.74	5.26	0.00	0.042
Karbis						109	105	4	0	
	—	—	—	—	—		96.33	3.67	0.00	0.038
Sonowals	83	43	34	6		106	96	9	1	
		51.81	40.96	7.23	0.042		90.57	8.49	0.94	1.991

ADA 1 has a wide range of variation: Brahmins have the lowest (38.2 %) and Ahoms have the highest (64.5 %) frequency. The ADA^2 allele varies widely between 0.073 ad 0.214 in most of the Indian populations; the present findings fall within this range.

Lactate dehydrogenase

LDH is monomorphic in most of the world populations, but in Indian populations the LDH Cal-1 variant appears to be polymorphic (Mukherjee & Reddy 1983). In the present study only one solitary case of LDH Delhi-1 variant has been detected in the Chutiyas (LDH^{Del-1} allele frequency is 0.0082). The rest show the normal LDH type. In India, hardly any population with strong Mongoloid affinities have been screened for LDH variants, except the present Assamese groups. It is interesting to note that no LDH Cal-1 type has been observed in these populations.

Analysis of genetic heterogeneity and distance measure

We have already published the sample sizes, phenotype and gene frequencies and results of distance analysis based on three serum protein and the haemoglobin loci for these ten populations (Walter et al. 1986; Deka et al. 1988). In this paper we have primarily reported the results of five enzyme systems. However, reference will also be made to the data on serum protein and Hb loci. This is done because certain types of analysis (e. g.,

Table 2. Allele frequencies of four enzyme systems in ten Assamese populations.

System	Allele	Brahmins	Kalitas	Kaibartas	Ahoms	Rajbanshis	Muslims	Kacharis	Chutiyas	Karbis	Sonowals
aP	p^a	0.2050	0.3000	0.2600	0.2208	0.2619	0.2667	0.2522	0.2845	0.2477	0.2617
	p^b	0.7950	0.7000	0.7400	0.7792	0.7381	0.7333	0.7478	0.7155	0.7523	0.7383
EsD	EsD^1	0.8081	0.7238	0.7426	0.6818	0.6990	0.7067	0.7054	0.7578	0.5991	0.6168
	EsD^2	0.1919	0.2762	0.2574	0.3182	0.3010	0.2933	0.2946	0.2422	0.4009	0.3832
ADA	ADA^1	0.6273	0.7083	0.5441	0.8045	0.5833	0.5000	0.5000	0.7959	–	0.7229
	ADA^2	0.3727	0.2917	0.4559	0.1955	0.4167	0.5000	0.5000	0.2441	–	0.2771
AK	AK^1	0.9151	0.9175	0.9109	0.9588	0.9515	0.9118	0.9792	0.9737	0.9817	0.9481
	AK^2	0.0849	0.0825	0.0891	0.0412	0.0485	0.0282	0.0208	0.0263	0.0183	0.0519

Table 3. Analysis of heterogeneity in ten Assamese populations based on enzyme and serum protein data.

Locus/Allele		No. of population	χ^2 **	df	Prob.	F_{ST}
aP:	p^a	10	7.24	9	0.612	0.0035
EsD:	EsD 1	10	33.63	9	0.0001	0.0166
LDH:	LDH Normal	10	12.39	9	0.192	0.0077
Hp:	Hp-1	10	11.71	9	0.230	0.0059
Tf:	$Tf\text{-}C^1$	10	19.77	9	0.019	0.0180
	$Tf\text{-}C^2$	10	22.87	9	0.007	0.0209
	$Tf\text{-}C^3$	10	9.68	9	0.377	0.0088
	Tf-D*	10	13.60	9	0.137	0.0124
Gc:	$Gc\text{-}1^F$	10	81.21	9	$< 10^{-4}$	0.0416
	$Gc\text{-}1^S$	10	96.10	9	$< 10^{-4}$	0.0492
	Gc-2	10	31.55	9	0.0002	0.0162
	$Gc\text{-}1^{A8}$*	10	24.29	9	0.004	0.0124
Hb:	Hb-A	10	323.96	9	$< 10^{-4}$	0.1594
	Hb-D	10	325.99	9	$< 10^{-4}$	0.1604
	Hb-S*	10	8.76	9	0.460	0.0043
ADA:	ADA-1	6	25.97	5	0.0001	0.0366
AK:	AK-1	10	25.64	9	0.002	0.0141
Pooled			1027.711	122	$< 10^{-6}$	0.0399 ± 0.0141

* Excluded in the total; ** χ^2-value for testing homogeneity of gene frequencies

gene diversity analysis etc.) was not performed in our earlier publications on these markers. Further, since for the purpose of reconstructing relationship among populations based on genetic data it is desirable to include as many loci as possible, and therefore we have used the serum protein and Hb data published earlier to construct a dendrogram of the Assamese populations.

Table 3 provides the summary statistics in the form of χ^2-values as well as the F_{ST} values for each enzyme, serum protein and haemoglobin marker. A significant heterogeneity exists for all alleles in the individual systems, except p^a, LDH^N, Hp^1, Tf^{C3}, Tf^D and Hb^S alleles.

For the total of all independent alleles the χ^2-value is 1027.11 with 122 d.f. ($p < 10^{-6}$), which shows that these populations are well differentiated in respect to the nine genetic loci considered here. For measuring the extent of gene diversity the F_{ST} computations were conducted using all the ten populations together. The F_{ST}-values show that the gene diversity is the smallest (0.0035) for the p^a and highest (0.1604) for the Hb^F allele. The degree of overall genetic differentiation as measured by the total F_{ST} value (0.0399 ± 0.0141) seems to be statistically significant (since the actual value of F_{ST} is about 3 times of its standard error). The estimation of standard genetic distance (Nei 1972) among all pairs of

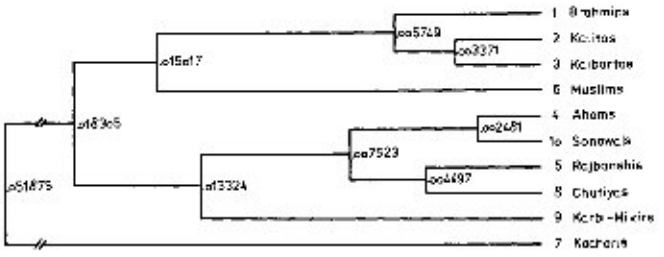

Fig. 2. Enzyme and serum protein tree in ten Assamese populations.

ten populations and their standard errors have been made to ascertain the relative position of these populations on the basis of these allele frequencies. Table 4 shows the average heterozygosities in these population (presented in the diagonal elements). The distance computations have been made only for eight loci to all populations; ADA has been excluded. Clustering of the populations on the basis of the distance values is shown in the dendrogram (Fig. 2). In aggregate we observe two major clusters (besides one single point cluster: Kacharis). *Cluster* 1 consists of Brahmins, Kalitas, Kaibartas and Muslims, who belong to the Caucasoid group according to traditional classification, while *Cluster* 2 includes Ahoms, Sonowals, Rajbanshis, Chutiyas and Karbis, who have strong Mongoloid affinities. Within the major clusters a few subclusters are also evident. In Cluster 1, there is one sub-cluster consisting of Brahmins, Kalitas and Kaibartas, which separate out from the Muslims. This means that the Muslims are somewhat different from the other groups in this cluster. Within the Cluster 2, two sub-clusters are visible, one having Ahoms and Sonowals and the other consisting of Rajbanshis and Chutiyas. The Karachis fall apart though they belong to the Mongoloid group, perhaps because of the very high Hb^E allele frequency (about 0.65). On the whole the pattern of clustering observed here is in conformity with the traditional classification of the populations of Assam based on anthropometric and linguistic evidence.

Acknowledgement

This study was supported by a grant from the Deutsche Forschungsgemeinschaft (Wa 122/31-1), for which the authors are thankful.

References

Das, B. M. (1987): The peoples of Assam. — Gian Publishing House, Delhi.

Das, B. M., Walter, H., Mukherjee, B. N., Gilbert, K., Lindenberg, P., Deka, R., Malhotra, K. C. & Chakraborty, R. (1987): Genetic variation of five blood group polymorphisms in ten populations of Assam, India. — Int. J. Anthrop. 2, 325—340.

Deka, R., Reddy, A. P., Mukherjee, B. N., Das, B. M., Banerjee, S., Roy, M., Dey, B., Malhotra, K. C. & Walter, H. (1988): Haemoglobin E distribution in ten endogamous populations of Assam, India. — Hum. Hered. 38, 261—266.

Harris, H. & Hopkinson, D. A. (1977): Handbook of enzyme electrophoresis in human genetics. — North-Holland Publishing Company, Amsterdam.

Table 4. Genetic distance and average heterozygosity (diagonal) based on eight enzymes and serum protein markers in ten Assamese populations.

	Brahmins (1)	Kalitas (2)	Kaibartas (3)	Ahoms (4)	Rajbanshis (5)	Muslims (6)	Kacharis (7)	Chuti (8)	Karbis (Mikirs) (9)	Sonowals (10)
Brahmins (1)	0.296 ±0.072	.0060 ±.0021	.0055 ±.0019	.0289 ±.0216	.0106 ±.0049	.0174 ±.0120	.0880 ±.0688	.0166 ±.0099	.0266 ±.0142	.0335 ±.0211
Kalitas (2)	–	0.323 ±0.072	.0034 ±.0022	.0194 ±.0152	.0061 ±.0036	.0093 ±.0091	.0672 ±.0560	.0081 ±.0059	.0187 ±.0115	.0227 ±.0142
Kaibartas (3)	–	–	0.312 ±0.065	.0187 ±.0133	.0039 ±.0019	.0184 ±.0151	.0747 ±.0538	.0101 ±.0054	.0254 ±.0197	.0225 ±.0131
Ahoms (4)	–	–	–	0.352 ±0.078	.0076 ±.0057	.0246 ±.0170	.0219 ±.0135	.0044 ±.0022	.0143 ±.0090	.0025 ±.0010
Rajbanshis (5)	–	–	–	–	0.333 ±0.072	.0177 ±.0109	.0533 ±.0366	.0045 ±.0022	.0128 ±.0108	.0100 ±.0056
Muslims (6)	–	–	–	–	–	0.325 ±0.073	.0585 ±.0581	.0136 ±.0077	.0138 ±.0061	.0246 ±.0158
Kacharis (7)	–	–	–	–	–	–	0.344 ±0.078	.0320 ±.0234	.0474 ±.0356	.0237 ±.0145
Chutiyas (8)	–	–	–	–	–	–	–	0.338 ±0.077	.0139 ±.0088	.0080 ±.0039
Karbis (Mikirs) (9)	–	–	–	–	–	–	–	–	.0340 ±.0074	.0123 ±.0085
Sonowals (10)	–	–	–	–	–	–	–	–	–	0.366 ±0.076

Mukherjee, B. N. & Reddy, A. P. (1983): Wide spread distribution of LDH variants in various ethnic groups of India. — SAP 5, 1—12.

Mukherjee, B. N., Walter, H., Malhotra, B. C., Chakraborty, R., Sauber, P., Banerjee, S. & Roy, M. (1987): Population genetic study in ten endogamous groups of West Bengal, India. — Anthrop. Anz. 45, 239—254.

Nei, M. (1972): Genetic distance between populations. — Amer. Nat. 106, 283—292.

Walter, H., Mukherjee, B. N., Gilbert, K., Lindenberg, P., Dannewitz, A., Malhotra, K. C., Das, B. M. & Deka, R. (1986): Investigations on the variability of haptoglobin, transferrin and Gc polymorphisms in Assam, India. — Hum. Hered. 36, 388—396.

Address for correspondence:

Prof. Dr. B. N. Mukherjee, Anthropometry and Human Genetics Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Calcutta — 700035, India.