

## Genetic relationships among some tribal groups inhabiting the north-eastern, eastern and sub-Himalayan regions of India

C. S. CHAKRABARTI<sup>1</sup>, M. ROY<sup>2</sup>, N. K. SENGUPTA<sup>3</sup>, R. LALTHANTLUANGA<sup>4</sup>  
AND P. P. MAJUMDER<sup>2</sup>

<sup>1</sup>*Department of Zoology, University of Burdwan, Burdwan, India*

<sup>2</sup>*Anthropology & Human Genetics Unit, Indian Statistical Institute, Kolkata, India*

<sup>3</sup>*Juthika Nursing Home, Chaibasa, India*

<sup>4</sup>*Department of Biochemistry, North-Eastern Hill University, Shillong, India*

(Received 5.3.02. Accepted 6.6.02)

### SUMMARY

The origins and genomic affinities of various tribal populations of India are of considerable contemporary interest. In this study, we have investigated relationships among five tribal groups inhabiting the north-eastern, eastern and sub-Himalayan regions of India. DNA samples have been analysed in respect of 25 polymorphic loci, based on which genetic affinities have been estimated. The interesting findings of this study are (i) the Tibeto-Burman speaking, morphologically Mongoloid, tribal groups of India are not genetically very homogeneous, and (ii) the Tharu, a group inhabiting the sub-Himalayan region, may indeed have undergone considerable admixture as has been postulated by some anthropologists.

### INTRODUCTION

People of the Indian subcontinent represent a great diversity of morphological, genetic, cultural and linguistic features (Majumder, 1998). To the characteristics of the indigenous populations of India many exogenous features have been added, due to immigrations that have taken place in historical times. As a result, the present Indian population comprises indigenous tribes with traditional lifestyles, along with many non-tribal groups. The number of tribal groups has been estimated to be about 450 (Singh, 1992). Some of the tribal groups are numerically large and found in various parts of India, while some groups are very small and inhabit restricted geographical territories. The tribal groups of India speak dialects that belong to three major language families – Austro-Asiatic (a branch of Austric),

Dravidian and Tibeto-Burman (a branch of Sino-Tibetan). The Dravidian speaking groups are primarily confined to the southern region of India, while the Tibeto-Burman speaking tribes are restricted to the northwestern region of India. In this study, we have attempted to estimate genomic relationships among three morphologically Mongoloid, Sino-Indian speaking tribal groups (Toto, Mizo, Tharu), and a morphologically proto-Australoid, Austro-Asiatic speaking tribal group (Ho). The Ho are included to serve as an anthropological and linguistic, and possibly genetic, ‘outgroup’ relative to the three other groups. While our expectation was that the Sino-Indian speaking, Mongoloid groups would be quite similar and distinct from the Austro-Asiatic speaking, proto-Australoid group, our study has yielded results contrary to these expectations.

Correspondence: Professor Partha P. Majumder, Anthropology & Human Genetics Unit, Indian Statistical Institute, B. T. Road, Kolkata 700 108, India. Fax: +91-33-577 3049 or +91-33-577 6680.

E-mail: ppm@isical.ac.in

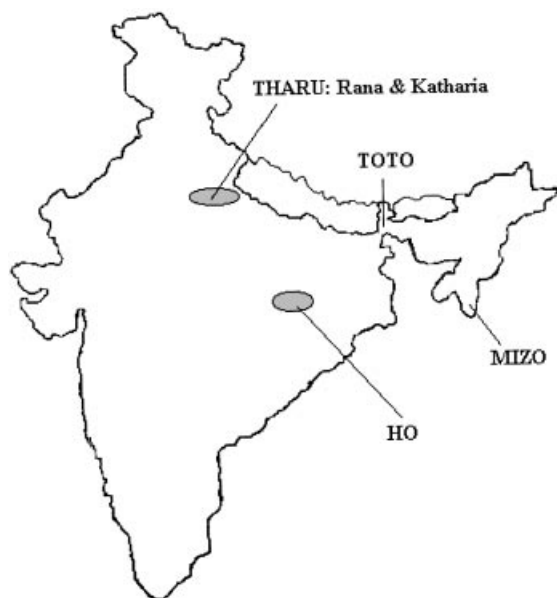


Fig. 1. Map of India showing the locations of sampling of the populations under study.

The Tharu is a numerically large (about 30000 individuals) tribe of Uttar Pradesh whose origins are largely unknown. Their name may have been derived from 'Terai', a region in the western Himalayan foothills, or from 'Thar' meaning a 'forestman', or from 'Thar', a desert in western India. The Tharus claim to be descendants of the Rajputs, a dominant caste population of northern India who are supposedly ancestors of many of the present-day Hinduised sections of the north Indian ethnic groups. Some anthropologists (Majumdar, 1944) have claimed that they have a mixed ancestry. Because they have Mongoloid morphological features, it has been suggested that they may have been derived from admixture between the proto-Australoid Indian tribes and some Mongoloid tribal group of Nepal. They are primarily agriculturists, although some are fishermen and some continue to be hunter-gatherers. They are divided into five main, largely exogamous, subgroups. We have sampled individuals from two of these subgroups – Rana and Katharia.

The Totos are a very small, secluded and primitive tribe comprising less than 1000 individuals, confined to a particular village called Totopara (Sanyal, 1973). Totopara is situated

228–282 m above sea level, in the district of Jalpaiguri in West Bengal, India, covering an area of approximately 2000 acres (Sanyal, 1973; Sarkar, 1993; Sinha & Pal, 1984). The Totos are morphologically, linguistically and culturally very different from the tribes and communities living around them, such as the Koch, Rajbanshi, Mech, Garo, Tephy, Panikoch and Bhutia (Sanyal, 1973). Gates (1963) considered them as a distinct isolated tribe having Mongoloid features. The Toto society is male dominated though women prefer husbands younger to them (Sinha, 1988). The Totos are an endogamous group comprising 13 exogamous clans (Sarkar, 1993). Inter-clan marriage is prohibited; monogamy is the normal practice (Sinha, 1988), and cross-cousin marriages, both maternal and paternal, are common (Debnath, 1982). Although there are many views regarding the origin of the Toto, the dominant view is that the Totos are a fraction of the Bhutia community, with Tibetan ancestry.

The people who live in the north-east Indian state of Mizoram designate themselves as the Mizo ('the highlanders'). The state of Mizoram shares borders with Bangladesh and Myanmar. The origin of the Mizo tribes remains uncertain, but it is known that they entered from the direction of Myanmar in large numbers. The word Mizo appears to be a blanket term and probably comprised many tribes, such as the Lushai, Punte, Poi, Roite, Darlong, Thado and Hmar. Some of these tribes are no longer separately identifiable (e.g. Darlong and Roite). However, the languages spoken by these groups, which belong to the Kuki-Chin subfamily, are dialectically and structurally so similar that this points to their belonging to a common ancestral stock (Goswami, 1979). They possess Mongoloid morphological features. Agriculture is the principal occupation of the Mizos. Most Mizos are followers of Christianity; a significant fraction also follow Buddhism. The total population size is about 450000.

The Hos are an Austro-Asiatic speaking tribal group, predominantly inhabiting the east Indian region. Many anthropologists believe that the numerically large Austro-Asiatic tribes of this

Table 1. Sample sizes, genotype and allele frequencies at 25 loci among tribal populations of India

Locus	Katharia Tharu	Rana Tharu	Toto	Mizo	Ho
<i>Alu</i> MtNuc					
<i>n</i>	36	53	30	27	53
+/+	9	20	15	5	10
+/-	11	13	11	13	28
-/-	16	20	4	9	15
$p \pm$ S.E.	$0.403 \pm 0.058$	$0.500 \pm 0.048$	$0.683 \pm 0.060$	$0.426 \pm 0.067$	$0.453 \pm 0.048$
$\chi^2$	4.793*	13.755*	0.700	0.006	0.232
<i>Alu</i> PV92					
<i>n</i>	36	53	30	29	54
+/+	19	36	26	21	17
+/-	12	15	4	8	27
-/-	5	2	0	0	10
$p \pm$ S.E.	$0.694 \pm 0.0548$	$0.821 \pm 0.037$	$0.933 \pm 0.032$	$0.862 \pm 0.045$	$0.565 \pm 0.048$
$\chi^2$	1.657	0.077	0.153	0.742	0.016
<i>Alu</i> FX3B					
<i>n</i>	36	53	29	27	53
+/+	22	35	24	10	35
+/-	13	17	5	15	15
-/-	1	1	0	2	3
$p \pm$ S.E.	$0.792 \pm 0.0478$	$0.821 \pm 0.037$	$0.914 \pm 0.037$	$0.648 \pm 0.065$	$0.802 \pm 0.039$
$\chi^2$	0.323	0.431	0.258	1.284	0.633
<i>Alu</i> D1					
<i>n</i>	36	53	30	28	52
+/+	12	19	8	7	18
+/-	6	9	4	2	11
-/-	18	25	18	19	23
$p \pm$ S.E.	$0.417 \pm 0.058$	$0.443 \pm 0.048$	$0.333 \pm 0.061$	$0.286 \pm 0.060$	$0.452 \pm 0.049$
$\chi^2$	15.546*	22.806*	14.700*	19.058*	17.072*
<i>Alu</i> APO					
<i>n</i>	36	53	30	29	53
+/+	25	26	23	22	37
+/-	11	25	5	5	16
-/-	0	2	2	2	0
$p \pm$ S.E.	$0.847 \pm 0.042$	$0.726 \pm 0.043$	$0.850 \pm 0.046$	$0.845 \pm 0.048$	$0.849 \pm 0.035$
$\chi^2$	1.171	1.848	3.600	3.400	1.675
<i>Alu</i> ACE					
<i>n</i>	36	53	30	29	54
+/+	12	21	9	10	30
+/-	17	24	16	15	19
-/-	7	8	5	4	5
$p \pm$ S.E.	$0.569 \pm 0.058$	$0.623 \pm 0.047$	$0.567 \pm 0.064$	$0.603 \pm 0.064$	$0.731 \pm 0.043$
$\chi^2$	0.049	0.070	0.222	0.189	0.588
<i>Alu</i> CD4					
<i>n</i>	36	54	30	26	54
+/+	35	52	30	26	54
+/-	1	2	0	0	0
-/-	0	0	0	0	0
$p \pm$ S.E.	$0.986 \pm 0.014$	$0.981 \pm 0.013$	1.000	1.000	1.000
$\chi^2$	0.007	0.019	—	—	—
<i>Alu</i> PLAT					
<i>n</i>	36	53	30	29	54
+/+	24	32	0	10	26
+/-	11	20	2	14	25
-/-	1	1	28	5	3
$p \pm$ S.E.	$0.819 \pm 0.045$	$0.792 \pm 0.039$	$0.033 \pm 0.023$	$0.586 \pm 0.065$	$0.713 \pm 0.044$
$\chi^2$	0.038	1.148	0.036	0.001	0.928
<i>ESR1</i>					
<i>n</i>	37	54	30	29	54
+/+	15	17	4	13	26

Table 1 (*cont.*)

Locus	Katharia Tharu	Rana Tharu	Toto	Mizo	Ho
+/-	16	30	15	16	22
-/-	6	7	11	0	6
$p \pm$ S.E.	$0.622 \pm 0.056$	$0.593 \pm 0.047$	$0.383 \pm 0.063$	$0.724 \pm 0.059$	$0.685 \pm 0.045$
$\chi^2$	0.241	1.224	0.099	4.209*	0.167
<b>Nat</b>					
<i>n</i>	37	54	30	26	54
+/+	17	27	28	26	30
+/-	20	24	2	0	20
-/-	0	3	0	0	4
$p \pm$ S.E.	$0.730 \pm 0.052$	$0.722 \pm 0.043$	$0.967 \pm 0.023$	1.000	$0.741 \pm 0.042$
$\chi^2$	5.075*	0.626	0.036	—	0.069
<b>PPIG</b>					
<i>n</i>	37	52	30	26	54
+/+	5	10	7	3	13
+/-	15	28	12	10	25
-/-	17	14	11	13	16
$p \pm$ S.E.	$0.338 \pm 0.055$	$0.462 \pm 0.049$	$0.433 \pm 0.064$	$0.308 \pm 0.064$	$0.472 \pm 0.048$
$\chi^2$	0.326	0.361	1.033	0.246	0.274
<b>PSCR</b>					
<i>n</i>	37	54	30	26	53
+/+	3	1	0	0	4
+/-	9	17	3	0	20
-/-	25	36	27	26	29
$p \pm$ S.E.	$0.203 \pm 0.047$	$0.176 \pm 0.037$	$0.050 \pm 0.028$	1.000	$0.264 \pm 0.043$
$\chi^2$	2.266	0.397	0.083	—	0.046
<b>T2</b>					
<i>n</i>	37	54	30	26	53
+/+	12	9	3	3	23
+/-	14	26	18	18	21
-/-	11	19	9	5	9
$p \pm$ S.E.	$0.514 \pm 0.058$	$0.407 \pm 0.047$	$0.400 \pm 0.063$	$0.462 \pm 0.069$	$0.632 \pm 0.047$
$\chi^2$	2.179	0.000	1.875	4.013*	1.163
<b>LPL</b>					
<i>n</i>	36	54	30	26	54
+/+	14	12	13	10	18
+/-	19	34	16	12	26
-/-	3	8	1	4	10
$p \pm$ S.E.	$0.653 \pm 0.056$	$0.537 \pm 0.048$	$0.700 \pm 0.059$	$0.615 \pm 0.067$	$0.574 \pm 0.048$
$\chi^2$	0.971	3.827	2.184	0.016	0.013
<b>ALB</b>					
<i>n</i>	37	54	30	29	54
+/+	5	14	11	9	6
+/-	11	26	16	10	24
-/-	21	14	3	10	24
$p \pm$ S.E.	$0.284 \pm 0.052$	$0.500 \pm 0.048$	$0.633 \pm 0.062$	$0.483 \pm 0.066$	$0.333 \pm 0.045$
$\chi^2$	2.670	0.074	0.660	2.778	0.000
<b>ALAD-MspI</b>					
<i>n</i>	37	54	30	28	54
+/+	4	3	0	0	0
+/-	10	15	2	4	12
-/-	23	36	28	24	42
$p \pm$ S.E.	$0.243 \pm 0.050$	$0.194 \pm 0.038$	$0.033 \pm 0.023$	$0.071 \pm 0.034$	$0.111 \pm 0.030$
$\chi^2$	2.615	0.693	0.036	0.166	0.844
<b>ALAD-RsaI</b>					
<i>n</i>	35	54	30	28	54
+/+	12	20	12	3	6
+/-	16	25	14	6	22
-/-	7	9	4	19	26
$p \pm$ S.E.	$0.571 \pm 0.059$	$0.602 \pm 0.047$	$0.633 \pm 0.062$	$0.214 \pm 0.055$	$0.315 \pm 0.045$
$\chi^2$	0.156	0.062	0.001	3.702	0.167

Table 1 (cont.)

Locus	Katharia Tharu	Rana Tharu	Toto	Mizo	Ho
<i>HB Ψβ-HincII</i>					
<i>n</i>	37	54	30	26	50
+ / +	7	9	5	0	2
+ / -	19	23	8	2	24
- / -	11	22	17	24	24
<i>p</i> ± S.E.	0.446 ± 0.058	0.380 ± 0.046	0.300 ± 0.059	0.038 ± 0.027	0.280 ± 0.045
$\chi^2$	0.057	0.495	3.998*	0.042	1.814
<i>HB 3' Ψβ-ΨβHincII</i>					
<i>n</i>	37	54	30	27	54
+ / +	10	10	6	2	1
+ / -	15	17	9	8	38
- / -	12	27	15	17	15
<i>p</i> ± S.E.	0.473 ± 0.058	0.343 ± 0.046	0.350 ± 0.062	0.222 ± 0.057	0.370 ± 0.046
$\chi^2$	1.291	4.896*	3.481	0.551	13.981*
<i>HB 5' β-Hinf I</i>					
<i>n</i>	37	54	30	28	53
+ / +	16	34	30	26	40
+ / -	15	17	0	1	12
- / -	6	3	0	1	1
<i>p</i> ± S.E.	0.635 ± 0.056	0.787 ± 0.039	1.000	0.946 ± 0.030	0.868 ± 0.033
$\chi^2$	0.581	0.200	—	11.750*	0.008
<i>HOXB4-MspI</i>					
<i>n</i>	37	54	30	25	54
+ / +	8	6	6	5	21
+ / -	16	23	14	9	27
- / -	13	25	10	11	6
<i>p</i> ± S.E.	0.432 ± 0.057	0.324 ± 0.045	0.433 ± 0.064	0.380 ± 0.069	0.639 ± 0.046
$\chi^2$	0.524	0.042	0.074	1.392	0.378
<i>DRD2-TaqIA</i>					
<i>n</i>	37	54	30	28	52
+ / +	14	14	3	13	7
+ / -	19	29	16	15	31
- / -	4	11	11	0	14
<i>p</i> ± S.E.	0.635 ± 0.05596	0.528 ± 0.04804	0.367 ± 0.062	0.732 ± 0.059	0.433 ± 0.049
$\chi^2$	0.431	0.323	0.660	3.748	2.388
<i>DRD2-TaqIB</i>					
<i>n</i>	37	54	30	26	48
+ / +	17	20	4	13	10
+ / -	13	21	15	13	28
- / -	7	13	11	0	10
<i>p</i> ± S.E.	0.635 ± 0.05596	0.565 ± 0.04771	0.383 ± 0.063	0.750 ± 0.060	0.500 ± 0.051
$\chi^2$	2.165	2.357	0.099	2.889	1.333
<i>DRD2-TaqID</i>					
<i>n</i>	37	54	30	29	54
+ / +	20	24	11	28	37
+ / -	13	22	17	1	16
- / -	4	8	2	0	1
<i>p</i> ± S.E.	0.716 ± 0.05241	0.648 ± 0.04595	0.650 ± 0.062	0.983 ± 0.017	0.833 ± 0.036
$\chi^2$	0.681	0.616	1.807	0.009	0.240
<i>ADH1β-RsaI</i>					
<i>n</i>	35	54	30	26	54
+ / +	11	16	12	7	9
+ / -	8	24	11	6	25
- / -	16	14	7	13	20
<i>p</i> ± S.E.	0.429 ± 0.05915	0.519 ± 0.04808	0.583 ± 0.064	0.385 ± 0.067	0.398 ± 0.047
$\chi^2$	9.956*	0.652	1.811	6.829*	0.062

\* Significant at the 5% level.

region were once a single tribe. Subsequently, in the course of migration to different areas, they assumed different names. The Austro-Asiatic speaking tribes are supposedly the most ancient inhabitants of India. The Hos have been living in isolation for a long time (Roy Chaudhury, 1958), in spite of the fact that they number over 400 000 individuals. Most are ancestor worshippers, although some have adopted Hinduism or Christianity. Their society comprises a number of exogamous clans; cross-cousin marriages are not a taboo.

#### MATERIALS AND METHODS

Blood samples (5–10 ml in EDTA) were collected from individuals with informed consent. All sampled individuals were unrelated at least to the first cousin level. The locations of sampling are indicated in Figure 1. Blood samples were transported in ice to the laboratory of the Anthropology and Human Genetics Unit, Indian Statistical Institute. High molecular weight DNA was isolated from the blood samples by the salting-out procedure (Miller *et al.* 1988). Each DNA sample was analysed for polymorphisms at 25 loci of which 8 were insertion/deletion polymorphisms (IDPs), and the remaining 17 were RFLPs. The names and GDB accession numbers or ALFRED (<http://alfred.med.yale.edu>) UID of the RSP loci are: *ESRI* (GDB: 185229); *Nat* (GDB: 187676); *PPIG* (CYP1A, GDB: 9956062)-*MspI*; D21S13E PSCR, (GDB: 182305); T2 (GDB: 196856); *LPL* (GDB: 285016); *ALB* (GDB: 178648); *ALAD-MspI* (GDB: 155925); *ALAD-RsaI* (GDB: 155924); HBBP1-*HincII* (GDB: 56084); HBBP1 3'-*HincII*; HBB 5'-*HinfI*; *HOXB4-MspI* (UID: SI0001670); *DRD2* (UID: SI000191L)-*TaqIB*, *TaqID*, *Taq1A*; *ADH1B-RsaI* (UID: SI000002C). The names of the IDPs are given in Table 1. Primers and protocols used for screening of the IDPs were as given in Majumder *et al.* (1999a) and Tishkoff *et al.* (1996), and those for RSPs were as given in Jorde *et al.* (1995), Majumder *et al.* (1999b) and from K. Kidd (personal communication). All

DNA isolations and analyses were performed in the laboratory of the Indian Statistical Institute, Kolkata.

Allele frequencies at each of these biallelic loci were estimated for each population by the maximum likelihood method.  $\chi^2$  tests of significance between the observed genotype frequencies and those expected under Hardy–Weinberg equilibrium were performed. For the three sets of linked loci in and around the *ALAD*, *DRD2* and *HB* genes, maximum likelihood estimates of haplotype frequencies were estimated by the EM algorithm, using the HAPLOFREQ package (Majumder & Majumder, 2000). For analyses of genomic diversities, differentiation and distances, estimated haplotype frequencies at these three sets of linked loci were used. Observed gene diversities and the coefficient of gene differentiation,  $G_{ST}$ , were estimated (Nei, 1973). Genetic distances between populations were estimated using the  $D_A$  distance measure (Nei, 1987). An unrooted neighbour-joining tree (Saitou & Nei, 1987) was constructed to identify affinities among the tribal populations.

#### RESULTS AND DISCUSSION

Sample sizes and the +allele (Insertion allele for the IDP loci and presence of the restriction site for the RFLP loci) frequencies are given in Table 1. All the loci except *Alu* CD4 showed high degrees of polymorphism across most populations. All populations at most loci show statistically non-significant differences of observed genotype frequencies and those expected under Hardy–Weinberg equilibrium (Table 1). However, all populations significantly deviated from the Hardy–Weinberg equilibrium at the *Alu* D1 locus. At this locus, there is a significant excess of homozygotes over the expected. There are striking differences in allele frequencies at the D2I5I3E (PSCR) locus; the Mizos are monomorphic at this locus. Among the 25 loci, there are three sets of linked loci in and around the *ALAD*, *DRD2* and *HB* genes (gene clusters). Estimated haplotype frequencies for these sets of

Table 2. Haplotype frequencies at linked sets of loci in and around three genes in tribal populations of India

Locus	Haplotype	Katharia Tharu	Rana Tharu	Toto	Mizo	Ho
<i>ALAD</i>	+/+	0.145	0.010	0.000	0.000	0.000
	+/-	0.313	0.188	0.033	0.111	0.071
	-/+	0.147	0.235	0.633	0.315	0.214
	-/-	0.395	0.567	0.333	0.574	0.714
<i>DRD2</i>	+/+/+	0.351	0.293	0.201	0.211	0.731
	+/+/-	0.079	0.204	0.164	0.163	0.000
	+/-/+	0.160	0.021	0.000	0.039	0.000
	+/-/-	0.035	0.011	0.000	0.000	0.000
	-/+/+	0.143	0.102	0.017	0.126	0.019
	-/+/-	0.052	0.014	0.000	0.000	0.000
	-/-/+	0.180	0.273	0.432	0.461	0.231
	-/-/-	0.000	0.082	0.185	0.000	0.019
<i>HB</i>	+/+/+	0.216	0.084	0.300	0.130	0.040
	+/+/-	0.131	0.045	0.000	0.105	0.000
	+/-/+	0.041	0.152	0.000	0.026	0.000
	+/-/-	0.071	0.068	0.000	0.025	0.000
	-/+/+	0.115	0.220	0.050	0.132	0.180
	-/+/-	0.024	0.000	0.000	0.000	0.000
	-/-/+	0.239	0.346	0.650	0.569	0.760
	-/-/-	0.163	0.085	0.000	0.013	0.020

Table 3. Gene diversities (on diagonal) and genetic distances between tribal populations of India

	Katharia Tharu	Rana Tharu	Toto	Mizo	Ho
Katharia Tharu	$0.467 \pm 0.041$				
Rana Tharu	0.0164	$0.462 \pm 0.038$			
Toto	0.0773	0.0517	$0.368 \pm 0.046$		
Mizo	0.0213	0.0172	0.0536	$0.448 \pm 0.032$	
Ho	0.0736	0.0634	0.0820	0.0576	$0.379 \pm 0.040$

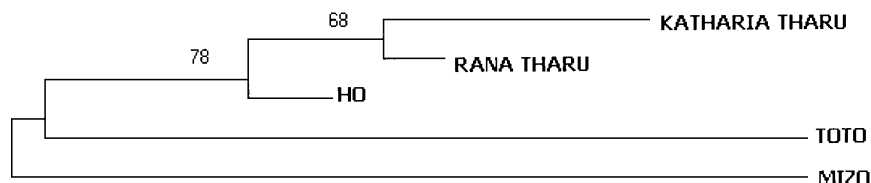


Fig. 2. Neighbour-joining tree depicting genomic relationships among tribal populations of India.

linked loci are given in Table 2. It is seen from this Table that there is a large variation in estimated frequencies of haplotypes across populations. The differences in haplotype frequencies are not only significantly different at the 5% level among the 5 groups, but are also significantly different between the two Tharu subgroups. However, the two Tharu subgroups share similarities which are distinctive from the remaining three groups. The *ALAD* +/+ haplotype is only present among the two Tharu subgroups, but not in the other populations. Similarly, the *DRD2* +/+/- and the -/+/-

haplotypes are not present among Totos, Mizos and Hos, but are present among Rana and Katharia Tharus. The gene diversity for each of the study populations is given in Table 3. Gene diversity values range from 0.368 (Toto) to 0.467 (Katharia Tharu). It is not surprising that the Toto should have the lowest gene diversity, in view of the small population size compared to the other populations studied. It is also not surprising that the two subgroups of the Tharu have remarkably similar values of gene diversity.

Genetic distances calculated on the basis of allele and haplotype frequencies at the 20 loci are

provided in Table 3. A neighbour-joining tree (Fig. 2) was constructed. Bootstrapping was done using 1000 replications. It is seen from the reconstructed genomic affinities that the two subgroups of the Tharu are close to each other. Even though most Tibeto-Burman speakers of northeast India are said to have originated from closely-related ancestral groups of southern China or southeast Asia, the Totos and the Mizos do not seem to be genetically close. Surprisingly, the Tharu subgroups do not cluster either with the Toto or the Mizo; instead the Tharu subgroups are closer to the Ho. The bootstrapped estimates of the confidence associated with these branches are quite high (Figure 2). This confirms the earlier observations of anthropologists (Majumdar, 1944) that the Tharu may be a mixed population. Indeed, the speculation that the Tharu may have been derived from admixture between a proto-Australoid tribal population of India and a Mongoloid tribal population of Nepal seems to be consistent with the findings of the present study. We have been unable to test this hypothesis of admixture formally and to obtain statistical estimates of admixture proportions because comparable genotype data from Mongoloid tribal populations of Nepal are not currently available. It would be interesting to attempt this in the future.

This study was supported by grants from the Department of Biotechnology, Government of India, and the Indian Statistical Institute. We are grateful to Dr Ken Kidd and Dr Lynn Jorde for sharing some laboratory protocols with us and to all members of the laboratory of the Anthropology & Human Genetics Unit, Indian Statistical Institute, for help and co-operation throughout the course of this study.

#### REFERENCES

- Debnath, S. K. (1982). Courtship and ceremonial aspect of marriage among Totos: a tribal isolated of Indo-Bhutan Border. *Vanyajati*, XXX: 49.
- Gates, R. R. (1963). The Totos (Cited by Sanyal, C. C. 1973. In *Meches and the Totos: two sub-Himalayan Tribes of the North Bengal*. Darjeeling: North Bengal University Press.
- Goswami, B. B. (1979). *Mizo Unrest: A Study of Politicization of Culture*. Jaipur: Aalekh Publishers.
- Jorde, L. B., Bamshad, M. J., Watkins, W. S., Zenger, R., Fraley, A. E., Krakowiak, P. A., Carpenter, K. D., Soodyall, H., Jenkins, T. & Rogers, A. R. (1995). Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57**, 523–538.
- Majumdar, D. N. (1944). *The Fortunes of Primitive Tribes*. Lucknow: Universal Publishers.
- Majumder, P. P. (1998). People of India: biological diversity and affinities. *Evol. Anthropol.* **6**, 100–110.
- Majumder, P. P., Roy, B., Banerjee, S., Chakraborty, M., Dey, B., Mukherjee, N., Roy, M., Guha Thakurta, P. & Sil, S. K. (1999a). Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur. J. Hum. Genet.* **7**, 435–446.
- Majumder, P. P., Roy, B., Balgir, R. S. & Dash, B.P. (1999b). Polymorphisms in the beta-globin gene cluster in some ethnic populations of India and their implications on disease. In *Molecular Intervention in Disease* (eds S. Gupta & O. P. Sood), pp. 75–83. New Delhi: Ranbaxy Science Foundation.
- Majumder, P. P. & Majumder, P. (2000). HAPLOFREQ: A computer program for maximum likelihood estimation of haplotype frequencies in a population from genotype data on unrelated individuals via the EM algorithm. Technical Report, Calcutta: Indian Statistical Institute.
- Miller, S. A., Dykes, D. D. & Polsky, H. F. (1998). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl. Acids Res.* **16**, 1215.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**, 3321–3323.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Roy Chaudhury, P. C. (1958). *Bihar District Gazetteers: Singhbhum*. Patna: Secretariat Press.
- Saitou, N. & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Sanyal, C. C. (1973). *The Meches and the Totos: Two Sub-Himalayan Tribes of North Bengal*. Darjeeling: North Bengal University Press.
- Sarkar, A. (1993). *Toto: Society and Change* (A sub-Himalayan Tribe of West Bengal). Calcutta: Firma KLM Private Ltd.
- Singh, K. S. (1992). *People of India*. Delhi: Oxford India Press.
- Sinha, D. (1988). Ethno-zoological studies of the sub-Himalayan Tribal area of North Bengal with emphasis on Ecological and Bio-energetics of the Totos. Ph.D. Thesis, Siliguri: North Bengal University.
- Sinha, D. & Pal, B. C. (1984). A study of Eco-sociological change among an underdeveloped Tribes: The Totos of North Bengal. *Bulletin of Cultural Research Institute*, Vol. XVI, No. 1–4.
- Tishkoff, S. A., Ruano, G., Kidd, J. R. & Kidd, K. K. (1996). Distribution and frequency of a polymorphic *Alu* insertion at the plasminogen activator locus in humans. *Hum. Genet.* **97**, 759–764.