

# Status of Austro-Asiatic groups in the peopling of India: An exploratory study based on the available prehistoric, linguistic and biological evidences

VIKRANT KUMAR and B MOHAN REDDY\*

*Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata 700 108, India*

*\*Corresponding author (Fax, 91-33-2577-6680; Email, bmr@isical.ac.in)*

Among the most contentious currently debated issues is about the people who had settled first in the Indian subcontinent. It has been suggested that the communities affiliated to the Austro-Asiatic linguistic family are perhaps the first to settle in India and the palaeoanthropological evidences suggest the earliest settlement probably around 60,000 years BP. Recent speculations, based on both traditional genetic markers and DNA markers, seem to corroborate the aforesaid view. However, these studies are inadequate both in terms of the representation of the constituent groups within this broad linguistic category as well as the number of samples that represent each of them. We strongly feel that, before making any formidable conclusions on the peopling of India and/or the history of settlement, it is necessary to ascertain that the Austro-Asiatic speakers, represented by over 30 different tribal groups, either genetically constitute a homogenous single entity or are a heterogeneous conglomeration, derived from different sources. As a first step towards this we tried to collate and analyse the existing information – geographic, ethno-historic, cultural and biological.

The results of the analyses of anthropometric and genetic marker data indicate that the Austro-Asiatic groups, particularly the Mundari speakers, with certain exceptions, show greater homogeneity among them when compared to the other linguistic groups, although certain groups show as outliers. However, traditional genetic markers show lower within population heterozygosity compared to Dravidian and other Indian populations. This is contrary to what has been claimed in case of certain DNA markers. Given that relatively greater heterozygosity among the Austro-Asiatic populations has been taken as one of the important evidences supporting greater antiquity of these populations one should await results of detailed DNA studies being currently undertaken by us, involving a number of Austro-Asiatic and other ethnic populations of India to resolve the issue unequivocally.

[Kumar V and Reddy B M 2003 Status of Austro-Asiatic groups in the peopling of India: An exploratory study based on the available prehistoric, linguistic and biological evidences; *J. Biosci.* **28** 507–522]

## 1. Introduction

Both the anthropological and archaeological evidences suggest that from time immemorial people belonging to different ethnic, cultural and linguistic groups have entered India and contributed to the present day population of the region. The languages of India have been broadly classified into four major families – Austro-Asiatic, Dravidian, Indo-European and Sino-Tibetan/Tibeto-Burman. While Dravidian languages are limited to the southern and central Indian regions, Tibeto-Burman languages are

most common in East-Bengal and in the eastern Himalayan region. Whereas in western, northern and parts of eastern India Indo-European languages are chiefly spoken. The two main branches that represent the Austro-Asiatic linguistic family in India are: (i) the *Mundari*, spoken mostly in and around Chota-Nagpur plateau and (ii) the *Mon-Khmer* spoken in northeast and in the Andaman and Nicobar islands (figure 2).

The Mundari branch is further trifurcated into Western, Southern and Northern Mundari groups. The western group is divided into two linguistic families, the Southern

**Keywords.** Anthropometry; australoid; culture; gene flow; genetic markers; geography; language; migration; Mon-Khmer; Mundari

and Northern Munda. Whereas the Southern Mundari group is further divided into Gorum, Saora, Remo, Gutub, Gata, Kharia, Juang and Korku languages, Kherwerian group of languages represents Northern Mundari, encompassing Santhali, Mundari, Ho and Korwa. About 25 tribal communities, belonging to the central and eastern region of India represent these languages. The Mon-Khmer branch in India includes three groups – the Khasi tribe with a number of subdivisions confined to the Khasi and Jaintia hills of Meghalaya and Nicobarese and Shompen of the Nicobar islands.

In the present paper, we shall explore if the available data are adequate to test the veracity of anthropological hypothesis that the groups affiliated to the Austro-Asiatic linguistic family are the oldest in India and perhaps constitute the basic substratum of the Indian population. Towards this, we shall critically review the prehistoric,

cultural and ethno-historic evidences and attempt comprehensive analyses of the published data on anthropometric measurements and genetic markers of about 30 populations, 14 of them being Austro-Asiatic, and put the results in the proper perspective so as to gauge the status of Austro-Asiatic tribes in the peopling of India.

Although cultural evidences for the existence and activities of the hominid populations are abundant throughout the Indian subcontinent during the middle and late Pleistocene, negligible fossil records of any Paleolithic man have been found except for the sites in Sri Lanka. While thermo-luminescence dates of 74,000–64,000 BP and 28,000 BP were assigned for samples from the two sites near Bundala, Hambantota district (Singhvi *et al* 1986), skeletal remains from Batadomba Lena and Beli Lena Kitulgala (the basal layer) in Sri Lanka are dated, respectively, to 28,000 BP and 24,500 BP. The archaeo-



**Figure 1.** Schematic representation of the two alternative hypotheses suggesting possible routes of earliest migration of people into the Indian subcontinent.

logical evidences from Bundala-Patirajawela further indicate settlement prior to 75,000 to 65,000 BP (Deraniyagala 1985, 1986). In the light of above evidences it may be suggested that the anatomically modern humans were present in the Indian subcontinent before the last interglacial period.

The studies of specimens from Fa Hien, Batadomba Lena and Beli Lena Kitulgala sites in Sri Lanka and the extant aboriginal people, the Veddas/Veddoid, indicate that they share a number of similar biological features of the skeleton and dentition. A few other late-Pleistocene anatomically modern *Homo sapiens* skeletal remains have been discovered from Darr-I-Kur in Afghanistan and Hathnora in India. Further, the Veddoid skeletal types have been found in Iran and Mesopotamia (Sergent 1997). These evidences may suggest that the Veddoid/ Australoid types were probably the earliest anatomically modern humans to arrive in the Indian sub-continent around 60,000–70,000 years ago and then went to Sri Lanka. In India, extensive fossil evidences come from the Mesolithic sites (especially of the Gangetic plain), dated to Holocene period. These hominids retained large body size, robustness and megadonty that are characteristics of terminal Pleistocene populations from other parts of Europe, Africa and Asia. This constellation of features survives among certain hunting-gathering populations of both prehistoric and modern times (Kennedy 1984), who were probably affiliated to Austro-Asiatic linguistic family such as Hos, Santhals and Korkus (Gordon 1958). Consistent with the foregoing observations, is the hypothesis that the people speaking Austro-Asiatic languages were probably the earliest to arrive in India. Gadgil *et al* (1998) state that they have compiled words for universally used nouns in Austro-Asiatic, Dravidian, Indo-European and Sino-Tibetan languages and although a more objective analysis of the extent of such variation is yet to be done, it appears that Austro-Asiatic languages show the greatest and Indo-European the least divergence. An interesting corollary to this is that among the Austro-Asiatic speakers, one finds all of them to be not only tribes but also practicing relatively more primitive technologies.

Very few genetic studies have been undertaken which focus on the issue of peopling of India. In one of the recent papers, based on mtDNA HVS1 sequence and RFLP data, Majumder (2001) concludes that Austro-Asiatic tribes show a higher diversity than the Dravidian tribes, implying their greater antiquity. However, this inference is based on inadequate data, both in terms of populations representing the two ethnic group/linguistic categories and also in terms of sample sizes representing these groups. Therefore, the spectrum of genomic diversity within each of these categories was not adequately reflected, hence the conclusions based on this study can at best be tentative.

The routes by which the first migrants entered India are also not clearly understood. Basically, there are two major schools of thought on this and figure 1 depicts the plausible scenarios of the earliest migration into India. According to some scholars (Ballinger *et al* 1992; Gadgil *et al* 1998; Diamond 1997) the Austrics had their origin in China, entered India through northeast corridor and then passed onto islands beyond. A strong support for this theory comes from the fact that almost all the Austro-Asiatic tribes are located in eastern and north-eastern-central India. Archaeological evidences provided by Lal (1956) Mohapatra (1975, 1985) and Gupta (1979) support this theory. The second principal component, derived from the analysis of 69 genes from 42 populations of Asia, which explains 17.7% of the variation is also compatible with fanning out of people from southeast Asia and China into India (Cavalli-Sforza *et al* 1994).

A more recent view is that the Austro-Asiatic speakers were another wave of migration from Africa to India and then to southeast Asia (Nei and Ota 1991; Chu *et al* 1998; Su *et al* 1999; Majumder 2001). The skull and the general anatomy of the fossil specimen, near Panchmari, in the village of Langhnaj in Gujrat, India, were similar to the specimens found in the Northeast Africa (Kennedy 2000). Nei and Ota (1991) also suggests that Mundari groups (branch of Austro-Asiatic linguistic family) in India, which are known to have dark skin, constitute perhaps another wave of migration from Africa. The Australoid-looking skeletons have also been found in Iran and Mesopotamia (Sergent 1997). All these related evidences may suggest that the probable migratory route of Austro-Asiatic people was from Africa to the Indian sub-continent and then to Indonesia and Australia.

An early wave of migration into India, actually from Africa through India, to southeast Asia has also been proposed using nuclear DNA microsatellite markers (Chu *et al* 1998) and Y-chromosomal DNA markers (Su *et al* 1999). This view is reinforced by the fact that the 9bp deletion, which was hypothesized to have arisen in Central China and radiated out from this region to southeast Asia (Ballinger *et al* 1992), is absent in most Indian populations and present in low frequency in southeast Asia (Majumder 2001). However, this inference seems to have limited value as many recent studies indicate that a number of south Indian caste and tribal populations showed 9bp deletion in variable frequency ranging up to 50% (Watkins *et al* 1999; Clark *et al* 2000; Reddy B M, Naidu V M, Madhavi V K, Thangaraj K and Singh L, unpublished results). On the basis of high frequency of haplogroup M, Quintana-Murci *et al* (1999) have proposed that this haplogroup has originated in East Africa approximately 60,000 years BP and that it was carried into India through an East African exit route by an early dispersal event of modern humans out of Africa. Concur-

rent to this, the first principal component based on 69 genes from 42 populations of Asia (Cavalli-sforza *et al* 1994), which explains about 35% of the total variation in the gene frequency, suggests movement of people radially fanning from Middle East.

The fact that Mundari speaking tribes inhabit mostly eastern, central and north-eastern region of India has been considered as an indirect evidence for the first theory of migration. This can be, however countered by hypothesizing the assimilation of some of the Austro-Asiatic tribes by later migrants; the disinterest of the later migrants in accessing the difficult terrain of hills and forests of central and eastern India had probably resulted in the survival of the Mundari and the other Austro-Asiatic groups inhabiting such tracts. Further, Sergent (1997) noted that the Austro-Asiatic groups were once predominant in a continuous belt, from central India to Vietnam,

but later many of the Mundari speaking tribes might have been absorbed by the Indo-European populations, leaving the peripheral ones in the inaccessible hills untouched. This explanation is consistent with the fact that east-Asia marker of ancient origin i.e. to say the concomitant presence of two sites Ddel at 10,394 and Alul at 10,397 in Austrics, Melanesians and native Americans has also been found in more than 50% of Indians with Caucasoid characteristics, not only in nuclear DNA but also in mtDNA (Passarino *et al* 1996). This explanation is also reflected in the study conducted by Roychoudhury *et al* (2000) that haplogroup M is ubiquitous in India although tribal populations have higher frequencies than the caste populations, most probably due to the process of assimilation.

It may be pertinent at this juncture to consider that the Austro-Asiatic speakers in India are represented by almost 30 different tribal groups, covering a wide geographic



Source: Modified from Gadgil *et al.* 1998

**Figure 2.** Map of India, showing geographic distribution of Austro-Asiatic tribes in India and the names of the populations analysed.

area. Although having broad linguistic and cultural similarity, they do represent two different subgroups of languages – Mundari and Mon-Khmer – and physically belong to two racial types, the Mongoloids and Australoids. Therefore, before inferring on the origin, migration and/or antiquity of these people, it is necessary to first ascertain if all the Austro-Asiatic groups in India genetically constitute a single entity or are they derived from different sources. It is in this context that we have taken up a major endeavor to study the extent of genomic diversity within this broad linguistic category, covering almost all the major groups. As a first step towards this endeavor we gathered all the available geographic, cultural and biological data and analysed critically to examine the nature and extent of heterogeneity among the Austro-Asiatic populations the results of which are presented in the subsequent sections.

## 2. Geographic affinity

Figure 2 presents the nature of distribution of Austro-Asiatic speakers in India. The Mon-Khmer group of Austro-Asiatic people includes the Khasi tribes, which are confined to the Khasi and Jantia hills of Meghalaya, and Nicobarese and Shompen of the Nicobar inlands. The Austro-Asiatic tribal communities affiliated to the Mundari branch are mainly concentrated in and around Chota-Nagpur plateau. While the Southern Mundari groups such as Juang, Gata, Bondo, Bodo Gadaba, Paranga and Saora inhabit the Koraput and adjoining districts of Orissa, Kherwarian groups comprising of Asur, Birhor, Ho, Korwa, Santhal, Turi and Munda are widely distributed from Jharkhand (Ranchi, Gumla, Lohardaga and Singhbhum districts) to Orissa (Mayurbhanj, Keonjhar and Sundergarh districts) to Madhya Pradesh (Raigarh and Jashpur districts) and to West Bengal (Birbhum, Nadia and Bakura districts). A section of Korku is the only tribe that inhabits the north-eastern border areas of Maharashtra. From the distributional pattern one can infer that these communities bear a strong affinity in that they occupy contiguous patches of ecologically similar habitats characterized by hilly and heavily forested areas.

## 3. Cultural homogeneity

All these communities are exclusively tribes, which are further divided into endogamous sub-tribes. Each of these tribes and sub-tribes constitutes within them a number of exogamous clans that are generally totemic i.e. named after plants, animals, birds and insects. These are basically hunting and food-gathering societies, supplemented by shifting cultivation. These tribes are also unique in practicing primitive technologies (Gadgil *et al* 1998).

Their political organization is found to be very traditional, constituting a very simple structure. One of the cultural traits exclusively found in Mundari groups is the use of vermilion by the women of these tribes to indicate their marital status (S Barua, personal communication). From the foregoing account, it appears probable that these Austro-Asiatic tribal communities, particularly the Mundari speakers, are a homogenous cultural entity.

## 4. Ethnic affiliation

With exceptions to the Mon-Khmer groups, which are classified as Mongloid, ethnically these tribal communities belong to Proto-Australoid type that is believed to be the basic element in the Indian population (Thapar 1966). Many scholars have classified these tribal communities under different headings. Lapicque (1905) used the term pre-Dravidian while Chandra (1916) favoured the term Nishada. Hutton (1933) used the term proto-Australoid and even put the Veddas of Sri Lanka within this ambit. While Hooton (1930) replaced the term proto-Australoid as pseudo-Australoid, Guha (1937) used the term proto-Australoid to designate these tribal groups. Taking into account the different classifications proposed, it is clear that all these authors agree that these tribal communities belong to one ethnic group although there is no agreement on the names used for the classification. A comprehensive picture will probably emerge only when we look at biological similarities shared by Austro-Asiatic speaking tribal communities.

## 5. Biological affinities

Recently published data pertaining to molecular genetic markers are very patchy and quite inadequate to unequivocally conclude that the Austro-Asiatic tribal groups belong to a single and genetically homogeneous ethnic group. Although data on anthropometric variables such as stature, cephalic index and nasal index etc. and on certain genetic marker were available in literature on a number of Austro-Asiatic tribes, no comprehensive analysis has yet been attempted. Therefore, in order to empirically ascertain the nature and extent of genetic heterogeneity among the Austro-Asiatic groups, we tried to collate published data on these groups and have attempted a comprehensive analysis of the anthropometric and traditional genetic marker data. We have also included few transitional groups, which were once considered to be Austro-Asiatic and at present are identified with the neighbouring non Austro-Asiatic groups. For the sake of comparison we have further included few non Austro-Asiatic populations, which have geographic proximity to the Austro-Asiatic tribes.

A total of thirty populations were considered for the present study, which includes twenty-three tribal groups and seven caste populations. Out of these, fourteen communities (all tribal groups) are affiliated to Austro-Asiatic linguistic family, ten to Dravidian and one tribal group speaks Tibeto-Burman language. The remaining five groups are affiliated to Indo-European linguistic family. The

name, geographic distribution and linguistic affiliation of the populations are given in table 1.

## 6. Anthropometric variation

Four anthropometric variables viz. stature, cephalic index, nasal index and facial index that are common to 22 of the

**Table 1.** Geographical distribution and the linguistic affiliations of the thirty investigated populations.

Sl. No.	Name of population	Distribution	Linguistic family	Genetic markers	Anthropometric measures
1	Agharias (T)	Uttar Pradesh; Madhya Pradesh, Mandla, Bilaspur, Rewa, Maikal hills; Bihar	Austro-Asiatic	–	+ <sup>1</sup>
2	Bado Gadaba (T)	Andhra Pradesh, Visakhapatnam District; Orissa, Koraput	Austro-Asiatic	+ <sup>2</sup>	+ <sup>3</sup>
3	Bareng Paroja (T)	Orissa, Koraput	Dravidian	+ <sup>2</sup>	–
4	Bhumij (T)	Bihar, Ranchi	Austro-Asiatic	+ <sup>4</sup>	–
5	Brahmin (UC)	Orissa (Os.)	Indo-European	–	+ <sup>5</sup>
6	Brahmin (UC)	Andhra Pradesh (A.P.)	Dravidian	–	+ <sup>6</sup>
7	Brahmins (UC)	West Bengal (W.B.)	Indo-European	+ <sup>7,8,9</sup>	–
8	Chenchu (T)	Andhra Pradesh, Kurnool, Nallamalla Hills; Karnataka	Dravidian	+ <sup>10,11</sup>	–
9	Dhurwa (T)	Madhya Pradesh, Bastar southeast Jagdalpur; Orissa, Koraput	Dravidian	+ <sup>12</sup>	–
10	Dudh Kharia (T)	Bihar, Ranchi; Madhya Pradesh, Raigarh, Jashpur; Orissa, Sundargarh	Austro-Asiatic	+ <sup>13</sup>	–
11	Ho (T)	Bihar Singhbhum; Orissa, Mayurbhanj and Koenjhar; West Bengal	Austro-Asiatic	+ <sup>14</sup>	+ <sup>15</sup>
12	Juang (T)	Orissa, Southern Keonjhar, Northern Angul, and Eastern Dhenkanal	Austro-Asiatic	–	+ <sup>16</sup>
13	Kamma (MC)	Andhra Pradesh	Dravidian	+ <sup>17,18</sup>	–
14	Khasi (T)	Assam; Meghalaya, Khasi-Jaintia hills	Austro-Asiatic	+ <sup>19,20</sup>	+ <sup>21</sup>
15	Khond (T)	Orissa, Phulbani, Koraput, Ganjam	Dravidian	+ <sup>22,23</sup>	+ <sup>24</sup>
16	Konda Paroja (T)	Orissa, Koraput	Dravidian	+ <sup>25</sup>	–
17	Korku (T)	Maharashtra, Amravati; Madhya Pradesh	Austro-Asiatic	–	+ <sup>26</sup>
18	Kyastha (UC)	West Bengal	Indo-European	–	+ <sup>5</sup>
19	Lanjia Saora (T)	Orissa, Ganjam Koraput and Phulbani; Andhra Pradesh, Srikakulam	Austro-Asiatic	+ <sup>22,23</sup>	+ <sup>15</sup>
20	Lodha (T)	Orissa, Mayurbhanj and Baleswar; West Bengal, Midnapore	Austro-Asiatic	–	+ <sup>27</sup>
21	Madiga (LC)	Andhra Pradesh	Dravidian	–	+ <sup>28</sup>
22	Mal Paharia (T)	Bihar, Ramgarh Hills, Dumka, Pakaur, Godda, Deoghar	Indo-European	–	+ <sup>29</sup>
23	Munda (T)	Bihar, Ranchi; Orissa Mayurbhanj	Austro-Asiatic	+ <sup>30</sup>	+ <sup>31</sup>
24	Nicobarese (T)	Nicobar Island	Austro-Asiatic	+ <sup>32,33</sup>	+ <sup>34</sup>
25	Ollaro Gadaba (T)	Orissa, Koraput	Dravidian	+ <sup>25</sup>	+ <sup>2</sup>
26	Oraon (T)	Bihar; West Bengal; Orissa	Dravidian	+ <sup>35,36</sup>	+ <sup>37</sup>
27	Pareng Gadaba (T)	Orissa, Koraput	Austro-Asiatic	+ <sup>25</sup>	+ <sup>2</sup>
28	Rajbanshi (LC)	West Bengal	Indo-European	+ <sup>38,39</sup>	+ <sup>40</sup>
29	Riang (T)	Tripura	Sino-Tibetan	+ <sup>41</sup>	+ <sup>42</sup>
30	Santhals (T)	Bihar, Santhal Pargana; Orissa; West Bengal	Austro-Asiatic	+ <sup>43,44</sup>	+ <sup>45</sup>

<sup>1</sup>Mahalanobis *et al* 1948–49; <sup>2</sup>Das *et al* 1963; <sup>3</sup>Das *et al* 1968; <sup>4</sup>Gupta 1962; <sup>5</sup>Ghua 1935; <sup>6</sup>Guha 1933; <sup>7</sup>Bhattacharjee 1956; <sup>8</sup>Chokraborty *et al* 1975; <sup>9</sup>Chaudhari *et al* 1969; <sup>10</sup>Simmons 1953; <sup>11</sup>Ramesh *et al* 1980; <sup>12</sup>Negi and Ahmad 1963; <sup>13</sup>Bhattacharjee and Kumar 1969; <sup>14</sup>Kumar and Mukherjee 1975; <sup>15</sup>Majumdar and Sen 1950; <sup>16</sup>Ray 1958; <sup>17</sup>Deep Kumar *et al* 1985; <sup>18</sup>Naidu and Veerraju 1982; <sup>19</sup>Miki *et al* 1960; <sup>20</sup>Bhattacharjee 1975; <sup>21</sup>Raychaudhuri 1935; <sup>22</sup>Hargrave 1963; <sup>23</sup>Papiha *et al* 1988; <sup>24</sup>Ray 1949; <sup>25</sup>Das *et al* 1966; <sup>26</sup>Basu 1970; <sup>27</sup>Bhowmick 1956; <sup>28</sup>Gupta and Basu 1960; <sup>29</sup>Sarkar 1935–36; <sup>30</sup>Kumar and Bhattacharjee 1976; <sup>31</sup>Basu 1932–33; <sup>32</sup>Gupte and Dash Sarma 1973; <sup>33</sup>Polunin and Sneath 1953; <sup>34</sup>Ganguly 1976; <sup>35</sup>Buechi 1954; <sup>36</sup>Kirk *et al* 1962; <sup>37</sup>Basu 1933–34; <sup>38</sup>Das and Bhattacharjee 1963; <sup>39</sup>Das *et al* 1967; <sup>40</sup>Ray 1946; <sup>41</sup>Kumar and Shastry 1961; <sup>42</sup>Mitra 1956; <sup>43</sup>Bhattacharjee 1969; <sup>44</sup>Kumar N and Mukherjee D P (unpublished); <sup>45</sup>Sarkar 1954.

T, tribe; UC, upper caste; MC, middle caste; LC, lower caste; '+', present; '–', absent.

30 populations are analysed for the present study. Out of these, twelve belong to Austro-Asiatic family whereas five, four and one group(s), respectively, affiliated to Dravidian, Indo-European and Sino-Tibetan linguistic families. For the sake of brevity, the basic data are not presented here and can be supplied on request.

Given the data in the form of only means and standard deviations, we have computed the three measures of Penrose's distance (Penrose 1954) – mean square, shape and

size components – to visualize the nature of population relationships. For the sake of comparison we have also computed Sanghvi's T-square (Sanghvi 1953). However, the mantel test for matrix correspondence (table 2) suggests that the correlation between  $mC^2_H$  and its shape distance matrix is almost one and even between  $mC^2_H$  and  $T^2$  or between  $T^2$  and shape distance it is very high. Furthermore, it is shape or morphology of the object, which enables it to be classified. Therefore, we have presented the neighbour joining (NJ) tree (figure 3) based on the Penrose's shape distance only.

Broadly speaking, this tree depicts two major clusters, one formed mostly by the Mundari groups and the second by the others. Although the Santhal joins the main cluster of Mundari groups, it is a single point cluster. The clustering of Oraon and the Mal Paharia, the two non-Mundari tribes, with the Mundari groups can perhaps be explained by their overlapping geographic distribution and a possible role of admixture. Basu (1934), Dutta and Gupta (1967) and Bhattacharjee and Kumar (1969) suggest that the current affiliation of the Oraons is with the Munda speaking tribal groups and the results of anthropometric analysis corroborate with this. In fact, Kumar and Mukherjee (1975) goes to the extent of saying that the Dravidian language of the Oraon is a later adoption and the Oraon originally belonged to some Munda speaking group. On the other hand, based on a study of myth of

**Table 2.** Values of mantel correlation between pairs of distance matrices.

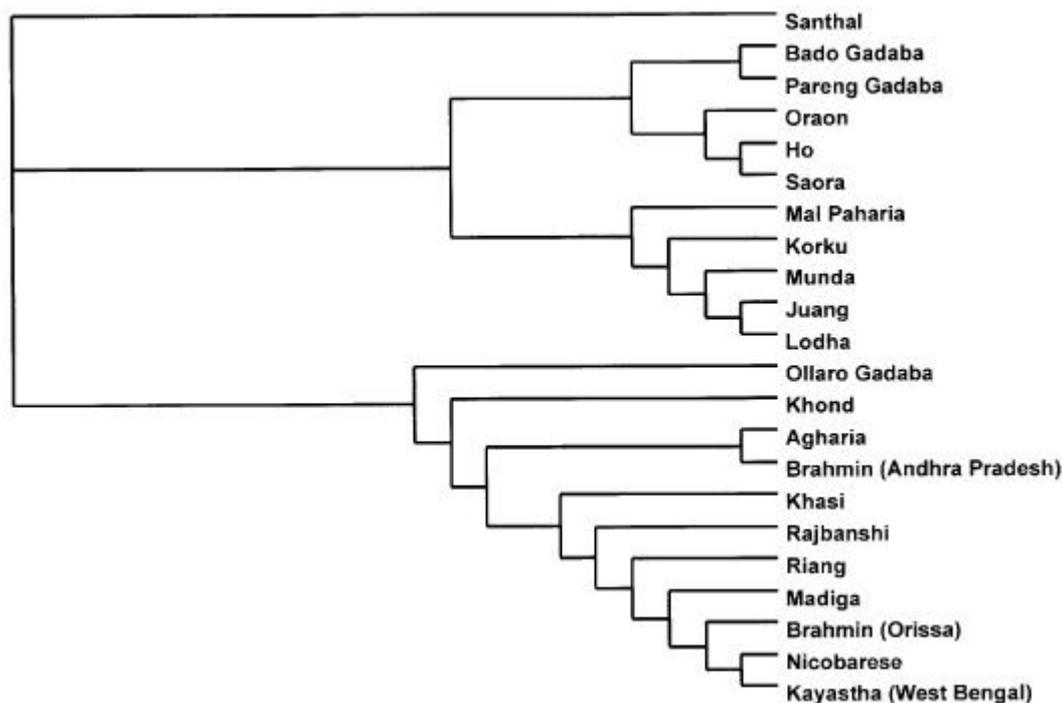
Distance matrix	<i>r</i>	<i>P</i>
$T^2$ and $mC^2_H$	0.8544	0.0010
$T^2$ and Penrose's shape	0.8442	0.0010
$mC^2_H$ and Penrose's shape	0.9894	0.0010
Penrose's shape and <i>DA</i>	0.6267	0.0910
$mC^2_H$ and <i>DA</i>	0.6275	0.0140
$T^2$ and <i>DA</i>	0.5419	0.0890
Penrose's shape and <i>Ds</i>	0.3833	0.1330
$mC^2_H$ and <i>Ds</i>	0.3722	0.1550
$T^2$ and <i>Ds</i>	0.3729	0.1080

$mC^2_H$  is Penrose's mean square distance.

$T^2$  is Sanghvi's T-square distance.

*DA* is modified Cavalli-Sforza chord distance of Nei *et al* 1983.

*Ds* is Nei's standard genetic distance.

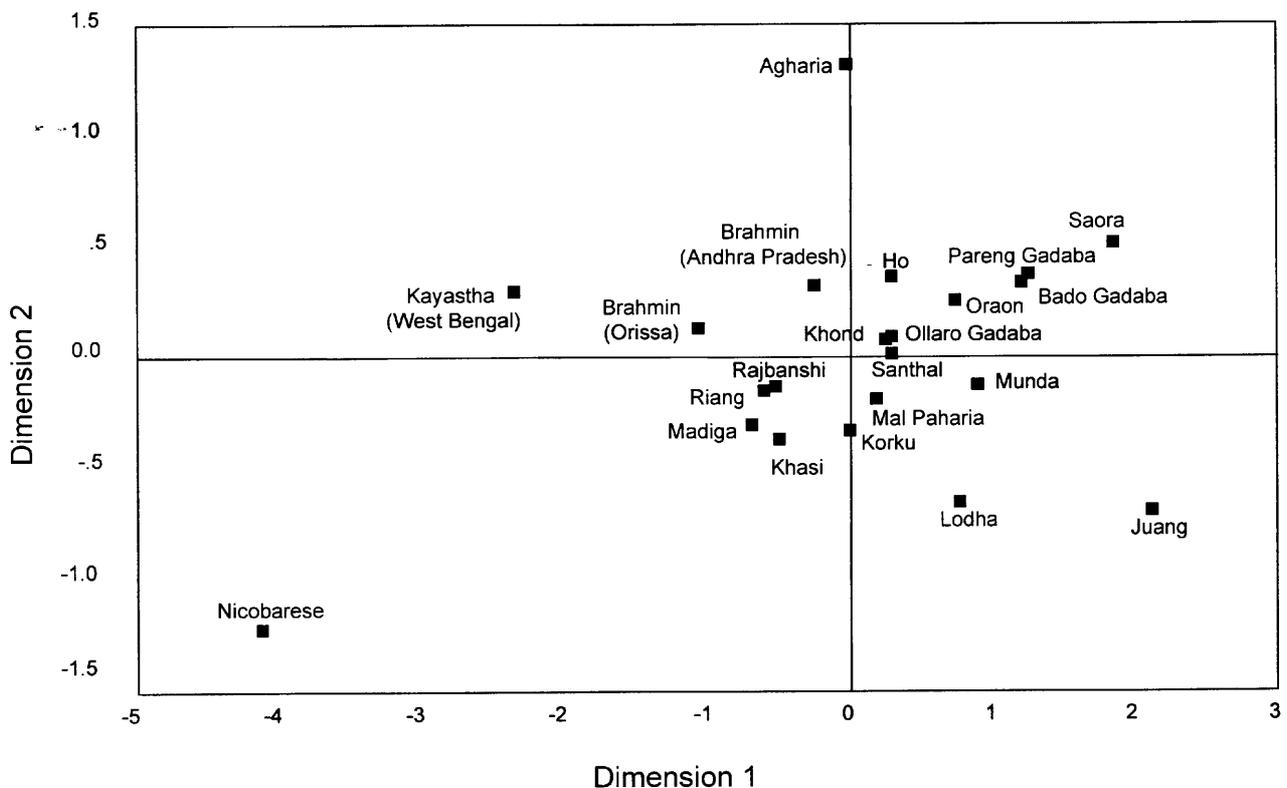


**Figure 3.** Neighbour joining tree constructed on the Penrose's shape distance matrix based on the four anthropometric variables among the 22 populations.

the Oraon Roy (1915) suggests that it is an immigrant population from South India. It is intriguing to note that within the second cluster constituting non-Mundari groups certain Mon-Khmer groups such as Khasi and Nicobarese, along with the other Mongloid populations like Riang from Tripura and other eastern Indian populations are included as a major subcluster, reflecting geographical contiguity and ethnic similarity. This pattern is also reflected in the study done by Dixon (1922) who on the basis of physical measurements of the Khasi and the surrounding Mongoloid populations concludes that the Khasi are racially related to the majority of the Tibeto-Burman tribes. The Khasi and the Nicobarese being ethnically Mongloid and linguistically belong to subfamily Mon-Khmer of the larger Austro-Asiatic family cluster neither with the Mundari groups nor closely with each other. It seems probable that the Khasi and Nicobarese genetically drifted apart, as geographical, and social and cultural factors favoured prolonged isolation between these groups (Kumar and Mukherjee 1975). Four other groups, Agharia from Uttar Pradesh and Ollaro Gadaba and Khond from Orissa, along with Brahmins of Andhra Pradesh form as outer elements to this subcluster and are a part of the 2nd main cluster. The clustering of Agharia with the Brahmins of

Andhra Pradesh is intriguing since they are linguistically and geographically different. Overall, it appears that the biological affinities among the Austro-Asiatic populations are structured mainly along the broad linguistic divisions and to a degree on geographic affiliations.

The limitations inherent in the dendrogram resulting from the imposition of bifurcation topology can be partially overcome by using multidimensional scaling for depicting relationships among populations based on frequency or distance data (Cavalli-Sforza *et al* 1994). In figure 4 we present the two-dimensional plot of population configurations based on the multidimensional scaling of Penrose's shape distances which accounts for 91% of the total variation. The following salient observations emerge: Broadly speaking, most of the Mundari groups form a compact constellation, separated from particularly the non Austro-Asiatic groups on the first dimension, having placed on the right half of the multivariate space. Whereas Austro-Asiatic Agharia is separated from the rest on the second dimension, the Nicobarese from Nicobar islands form as an extreme outlier, separated from the rest of the groups on both the dimensions. Lodha and Juang also seem to have relatively more diverged from the rest of Austro-Asiatic groups on the second dimension.



**Figure 4.** Plot of 22 populations on the first two dimensions derived from the multidimensional scaling of the anthropometric distances (Penrose's shape distance).

Overall, both the NJ tree and the multidimensional scaling depict the same pattern of population configurations on the basis of linguistic criteria. Therefore, the results reflect congruence between biological and linguistic similarity.

### 7. Genetic heterogeneity

Gene frequency data for five genetic markers namely A<sub>1</sub>A<sub>2</sub>BO, Rh and MN blood groups, ABH secretion in saliva and sickle cell trait, commonly available for a total of twenty-one populations, were gathered from the literature. Out of these, eighteen are tribal populations and three caste populations. Ten of these tribal populations are affiliated to Austro-Asiatic linguistic family, eight are Dravidians, one Tibeto-Burman tribe and the remaining two groups belong to Indo-European linguistic family. We have already discussed the criteria for selecting populations other than Austro-Asiatic groups. Since the data pertaining to sickle cell trait was not available for all the communities and is also absent in most of the Austro-Asiatic groups, the gene diversity and genetic distance analyses were done on the basis of remaining four loci only. For the sake of brevity, we are not presenting allele frequencies for these loci and the data can be supplied on request.

To examine the degree of gene diversity within and between the populations, Nei's (1987) method was used and  $H_T$ ,  $H_S$  and  $G_{ST}$  were computed.  $G_{ST}$  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 total intrapopulation heterozygosity. We have also computed average heterozygosity and their standard errors for all the populations (table 3). The range and the average for different categories of populations are given in table 4. The relatively higher values of average heterozygosity, hence  $H_S$ , is observed for both the tribal as well as pooled Dravidian groups compared to the Mundari or to the Austro-Asiatic populations as a whole. The  $G_{ST}$  values, locus-wise as well as averages for different categories of populations are furnished in table 5. The  $G_{ST}$  value is highest for ABH locus (0.0614) and least for the A<sub>1</sub>A<sub>2</sub>BO locus (0.0313). While the average  $G_{ST}$  value is 0.0485 when all the twenty-one populations were considered, it is lowest for the Mundari groups (0.0226) and highest for the Dravidian groups, both for tribes (0.0533) and for pooled category (0.0548), reflecting partly the heterogeneity of the populations included in this category.

Based on the allele frequency data Nei's standard genetic distances ( $D_S$ ) were computed for different pairs of populations. Although we have drawn both the NJ and the UPGMA trees because of the relatively greater congruence of the latter to the anthropometric pattern of

clustering, we have presented only the UPGMA tree (figure 5). In this tree the earliest divergence is shown by the Oraon, which is considered to be Dravidian linguistically but otherwise has greater affinity with the Mundari speaking tribal groups. Broadly speaking, this dendrogram consists of three major clusters, the first one constituted by nine out of the twenty-one populations, most of which (with exception to Brahmins of West Bengal) can be considered as Mundari or closely affined to Mundari groups. Within this broad cluster there are three sub-clusters: the first formed by the Santhal, Kharia and Munda; the second by Ho, Bareng Paroja, Bhumij and Rajbanshi; and the third by Brahmins of West Bengal in association with the Dhurwa. It may be noted that Kumar and Mukherjee (1975), on the basis of a small genetic distance of Bareng Paroja with most of the Munda speaking tribes conclude that Bareng Paroja is an offshoot of

**Table 3.** Average heterozygosity and its standard error based on four loci for twenty-one populations.

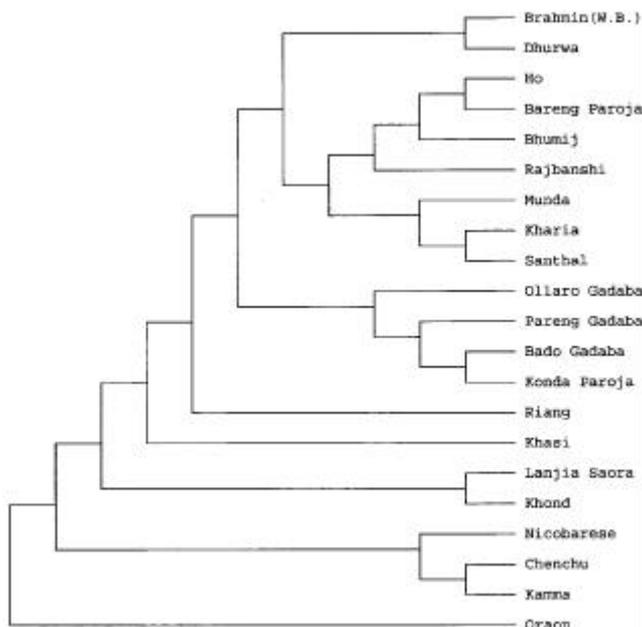
Name of population	Average heterozygosity	standard error
Bhumij	0.44415	0.06528
Bado Gadaba	0.48982	0.05619
Ollaro Gadaba	0.45265	0.06538
Pareng Gadaba	0.45998	0.07060
Ho	0.48606	0.05345
Kharia	0.50040	0.05540
Khasi	0.45176	0.03925
Munda	0.47259	0.06654
Nicobarese	0.26798	0.07650
Santhal	0.47485	0.06214
Saora	0.42201	0.07104
Bareng Paroja	0.48086	0.05339
Konda Paroja	0.46157	0.06436
Brahmins (WB)	0.50700	0.04109
Chenchu	0.52907	0.06715
Dhurwa	0.51928	0.04749
Kamma	0.54359	0.06547
Khonds	0.45735	0.06486
Oraon	0.52541	0.03384
Rajbanshi	0.44166	0.07056
Riang	0.40815	0.09775

**Table 4.** The range and the average values of heterozygosity for different groups of populations.

Groups of populations	Range	Average
All groups	0.26798–0.54359	0.46376
Austro-Asiatic	0.26798–0.50040	0.44424
Munda	0.42201–0.50040	0.46607
Munda+transitional	0.42201–0.52541	0.46822
Dravidian (pooled)	0.45265–0.54359	0.49317
Dravidian (tribes)	0.45265–0.52907	0.48634

**Table 5.** Locus-wise and average gene diversity ( $G_{ST}$ ) values among different linguistic categories.

Locus	All groups	Austro-Asiatic	Munda	Munda + transitional	Dravidian (tribes)	Dravidian (pooled)
ABH	0.0614	0.0592	0.0535	0.0433	0.0936	0.0830
A <sub>1</sub> A <sub>2</sub> BO	0.0313	0.0511	0.0047	0.0055	0.0055	0.0125
Rh	0.0505	0.0223	0.0165	0.0231	0.0399	0.0564
MN	0.0557	0.0234	0.0179	0.0635	0.0905	0.0795
Average	0.0485	0.0417	0.0226	0.0315	0.0533	0.0548

**Figure 5.** UPGMA tree based on Nei's standard genetic distances between the 21 populations.

some Munda speaking group, which is perhaps concurrent with the present study. On the other hand, the Rajbanshi of West Bengal has been considered by Grierson (1906) to be formerly Austro-Asiatic group, and Bhattacharjee and Kumar (1969) observed complete homogeneity of Rajbanshi with the Munda speaking groups in a study of four genetic characters. Therefore, the clustering of Rajbanshi with the Mundari groups is expected. Ollaro Gadaba, Pareng Gadaba, Bado Gadaba and Konda Paroja form the second cluster. The clustering of Ollaro Gadaba and Konda Paroja with the other two Gadaba tribes, although linguistically different, is not surprising because they are geographically proximate to each other, inhabiting the Koraput district of Orissa. Furthermore, the Ollaro Gadaba is considered to be a subtribe of the larger Gadaba tribe. This cluster joins the first cluster to form a major cluster. Thus the pattern of sub-clustering within

the major cluster, consisting of 13 of the 21 studied populations subtly reflects linguistic, ethno-historic and geographic affiliations. The remaining eight populations, form a loose congregation of independent elements of Oraon, Rieng and Khasi and the two subclusters formed by Lanjia Saora and Khond on the one hand and Nicobarese, Chenchu and Kamma on the other. Even in this certain degree of geographic and linguistic affiliations can be perceived. For example, the cluster of Khasi and Rieng is consistent with the study of Bhattacharjee (1975) who infer that probably Khasi are genetically affiliated to the adjoining Tibeto-Burman Mongloid populations, despite the linguistic barrier. Similarly, Chenchu and Kamma from Andhra Pradesh and Lanjia Saora and Khond from the neighbouring Orissa tend to form subclusters suggesting implicit geographic and/or linguistic similarity. Lanjia Saora is a subgroup of Saora of Andhra Pradesh and clustering of this Austro-Asiatic group with the three Dravidian groups may also reflect possible admixture.

The results of multidimensional scaling of the genetic distances of the twenty-one populations are shown in figure 6. The plot of populations on the two-dimensions, which explains 88.3% of the total variation depicts Oraon, Chenchu, Kamma and Nicobarese on the one hand and Lanjia Saora and Khond on the other hand as clear outliers. Except in case of Oraon and Nicobarese, most of the differentiation is on the first axis. It is apparent that most of the Mundari groups lie close to the centroid, while transitional groups like Gadaba and Mon-Khmer groups, Khasi and Nicobarese, form as peripheral elements to the core cluster of Mundari groups, subtly depicting relative linguistic and geographical similarities.

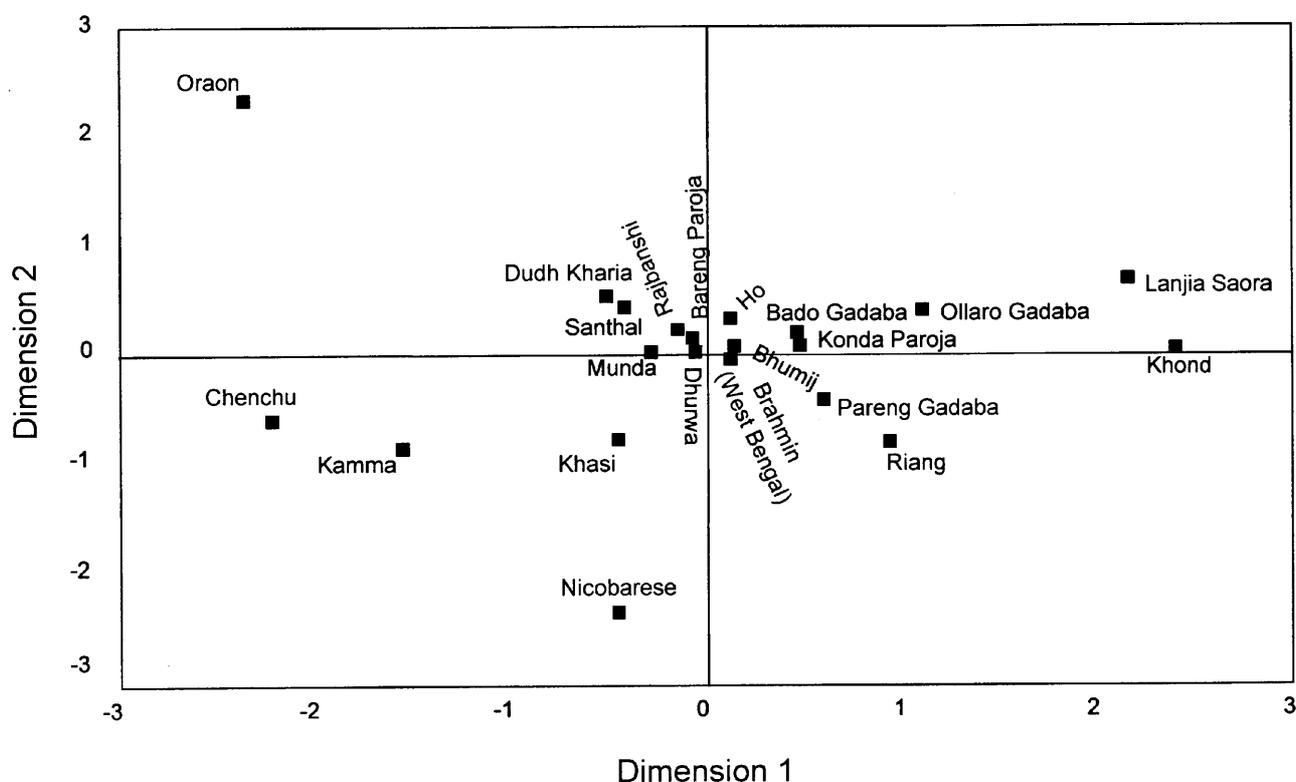
The degree of congruence in the pattern of population relationships based on anthropometry and the genetic markers is examined by computing mantel correlation between pairs of distance matrices (table 2). It is evident from the table that the Mantel correlation is relatively high between Nei's  $DA$  distance matrix and all the anthropometric distance matrices, although the correlation reaches statistical significance only between Nei's  $DA$  and Penrose's  $mC_H^2$ . This may suggest fair degree of simila-

rity in the pattern of population relationships based on the two sets of variables. However, mantel correlation of  $D_s$  distance matrix with each of the anthropometric distance matrices is relatively low despite showing greater congruence with the anthropometric tree based on 21 populations. It may be noted that the distance matrices used for Mantel correlation were computed for only 13 of the 30 populations for which data on both anthropometry and genetic markers were commonly available.

### 8. 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 Austro-Asiatic groups, we have used Harpending and Ward (1982) model, which is somewhat analogous to the Wright's island model. According to this model, given the uniform systematic pressure (gene flow) from outside, the average heterozygosity of 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. Regression plot of average heterozygosity vs. distance from the centroid for the Austro-Asiatic groups is shown in figure 7, along with the theoretical regression line. In the regression plot, the Nicobarese of Nicobar Island is an extreme outlier below the theoretical regression line and also far moved from the centroid. This may suggest their long isolation and effect of drift. This is concurrent with the fact that the Nicobarese are both culturally and geographically isolated even from their neighboring groups. On the other hand, Dudh Kharia and Bado Gadaba are outliers above the regression line, which may suggest that they have experienced more external gene flow than any of the other tribes. While Bado Gadaba inhabits the Koraput district in Southern Orissa, Dudh Kharia is the most advanced sub-section of the Kharia having much greater interactions with outsiders. Therefore, it is probable that they have received considerable gene flow from the surrounding populations. The remaining populations are placed nearer to theoretical regression line as well as relatively closer to the centroid. Since the Khasi and Nicobarese do not belong to the same linguistic sub-



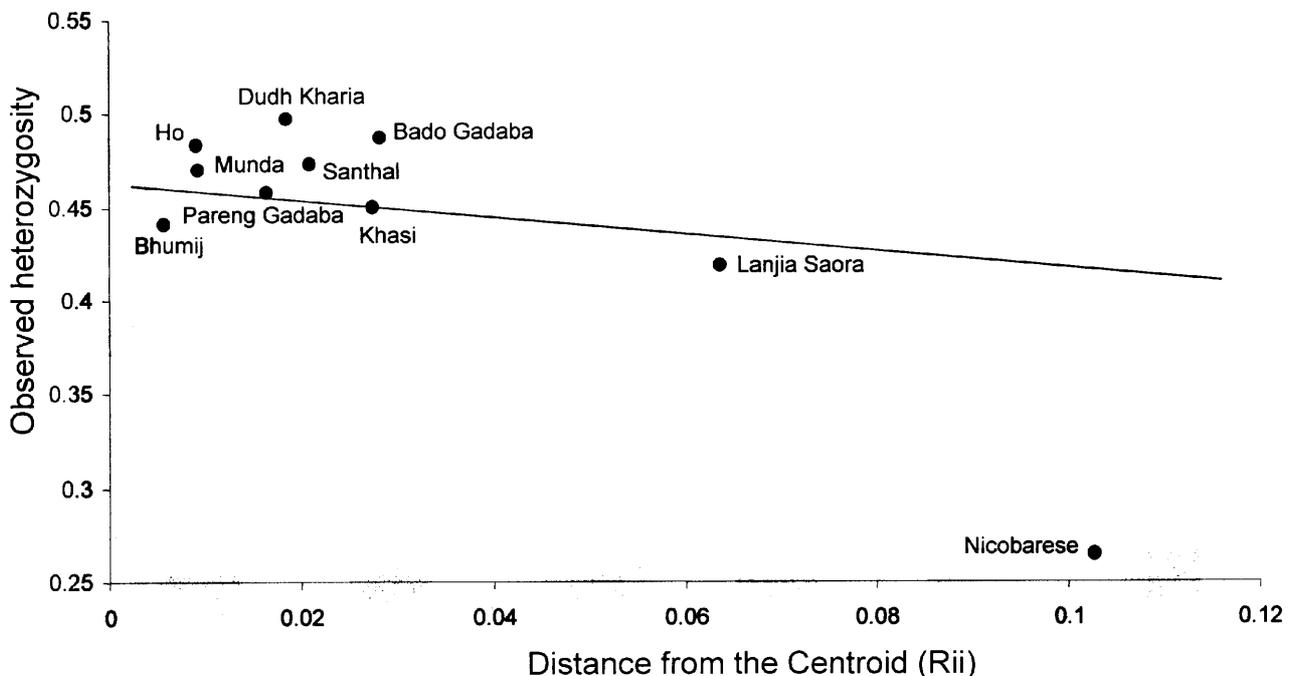
**Figure 6.** Plot of 21 populations on the first two dimensions derived from the multidimensional scaling of the Nei's standard genetic distances.

family as that of the Mundaris and are also geographically somewhat separated from them, the Harpending and Ward model was employed exclusively for the eight Munda tribes (figure 8). In this regression plot, while Dudh Kharia and Bado Gadaba are placed as extreme outliers above the theoretical regression line, Bhumij and Lanjia Saora are placed below the line as outliers. As discussed earlier the ethnography of Lanjia Saora suggest that it has migrated to the forests and hills of Koraput and Ganjam districts of Orissa from the plains of Mandasa in the Srikakulam district of Andhra Pradesh after they were driven away by their kinsmen. Therefore, their position in the above regression plot may reflect bottleneck effect of certain degree. Another regression plot (figure 9) considering transitional groups (Ollaro Gadaba, Bareng Paroja, Rajbanshi and Oraon), which are once considered to be Mundari and at present share geographic proximity with them, presents the Rajbanshi and Oraon as outliers below and above the regression line, respectively, suggesting greater isolation and drift in case of the former and excessive gene flow into the latter. It is probable that the Oraon have not only adapted the Dravidian language but indeed have also experienced genetic admixture.

### 9. Conclusion

From the available prehistoric, ethno-historic, linguistic and biological evidences, it appears that the Indian subcontinent was probably first peopled approximately 60,000

years BP. There has also been some speculation that populations affiliated to the Austro-Asiatic linguistic family were probably the earliest to settle in the subcontinent. One of the hypotheses tracing the migratory routes indicate that Austro-Asiatic languages speaking tribes probably came from Africa, entered India through western corridor (Nei and Ota 1991; Chu *et al* 1998; Su *et al* 1999; Majumder 2001). There is an alternative hypothesis in vogue, which speculates that these groups have migrated from China into India through north-eastern corridor (Ballinger *et al* 1992; Gadgil *et al* 1998; Diamond 1997). Neither of these hypotheses however could so far muster adequate and unequivocal empirical support so as to validate any of the competing hypotheses with reasonable confidence. The absence of 9 bp deletion in most Indian populations, which was claimed as evidence for the western Indian route of migration does not seem to be valid any more as many populations recently studied show 9 bp deletion in varying frequency, ranging even up to 50% (Watkins *et al* 1999; Clark *et al* 2000; Reddy B M, Naidu V M, Madhavi V K, Thangaraj K and Singh L, unpublished results). As has been pointed out earlier, there have also been problems on several other counts. First of all, the so called Austro-Asiatic groups in India represent more than 30 endogamous tribal populations but only a group or two were included in the past studies, leading to the all important conclusions on the peopling of the subcontinent, disregarding the implicit genetic/linguistic heterogeneity among them. Further-



**Figure 7.** Regression plot of the average heterozygosity vs. distance from the centroid of the Austro-Asiatic populations.

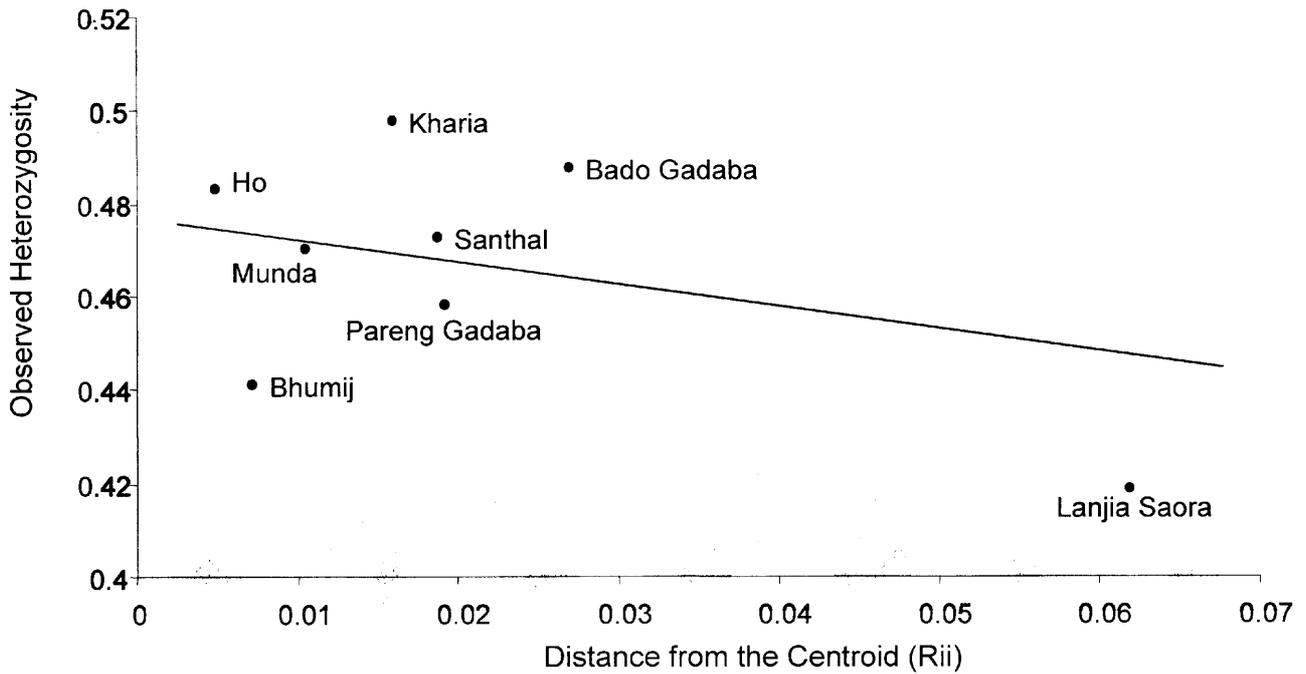


Figure 8. Regression plot of the average heterozygosity vs. distance from the centroid of the Mundari populations.

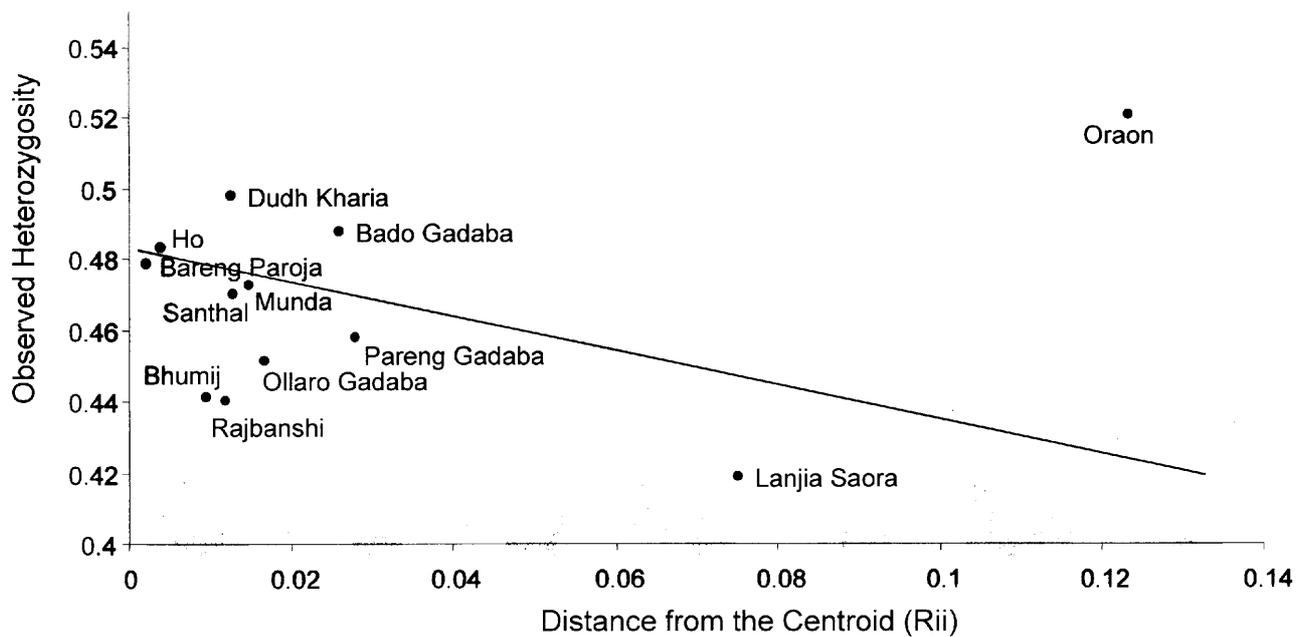


Figure 9. Regression plot of the average heterozygosity vs. distance from the centroid of the Mundari populations, along with certain transitional groups.

more, those groups were represented only by a handful of individuals constituting very small samples. In addition to this, the representation of the other linguistic and ethnic elements forming the comparative perspective was very inadequate, given that they constitute predominant sec-

tion of the Indian population. It is therefore imperative that a comprehensive study including representative samples from the Austro-Asiatic tribes should be undertaken to ascertain the magnitude of heterogeneity among them *vis-à-vis* the other ethnic elements in the Indian popula-

tion before any formidable conclusions can be drawn on the peopling of the subcontinent.

The analyses of few genetic markers and the anthropometric variables attempted here although suggest relative homogeneity particularly among the Mundari component of the Austro-Asiatic tribes, the number of markers/variables available for this study makes this observation rather tentative. In any case, the Mon-Khmer constituents of the Austro-Asiatic tribes behave somewhat differently than the others in this category, hence lie somewhat as outliers in the multivariate space. Ethnically, Mundari groups are considered to be Proto-Australoid whereas Mon-Khmer groups are classified as Mongloid. However, that the transitional groups, particularly the Ollaro Gadaba and Oraon, lie as outliers above the line in the regression plot may hint at the possibility of assimilation of certain Austro-Asiatic groups into later migrants speaking different other languages. This may indirectly support the hypothesis that the Austro-Asiatic people might have entered through the western Indian route, although this observation again needs validation with more appropriate data. Relatively greater heterozygosity of the Austro-Asiatic groups, when compared to others that was considered as one of the evidences for inferring greater antiquity of these people in India (Majumder *et al* 1999; Majumder 2001), is not reflected in the traditional markers and in fact, on the contrary, it is the Dravidian groups that show greater heterozygosity. Therefore, we cannot be certain that this inference is robust and can withstand scrutiny if the proper representation of Austro-Asiatic groups as a whole is made and compared with other linguistic and/or ethnic groups. In collaboration with Centre for Cellular and Molecular Biology, Hyderabad, we are currently generating DNA data on a number of Austro-Asiatic and other ethnic/linguistic groups of India and should be able to address this problem comprehensively in the near future.

#### Acknowledgements

We are grateful to Prof. K C Malhotra for a number of illuminating discussions in the process of developing this manuscript. It is indeed just to acknowledge that it was he who had floated the idea of working on this problem jointly with one of us (BMR) during 1992, although for a number of reasons it could not be initiated. We thank the two anonymous reviewers for their comments, which helped in improved presentation of the manuscript.

#### References

- Ballinger S W, Schurr T G, Torroni A, Gan Y Y, Hodge J A, Hassan K, Chen K H and Wallace D C 1992 Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations; *Genetics* **130** 139–152
- Basu A 1970 Anthropometry of the Korkus of Meghat forest; *Man India* **50** 149–156
- Basu P C 1932–33 Racial affinities of Mundas; *Trans. Bose Res. Inst.* **8** 211–247
- Basu P C 1933–34 Racial affinities of Mundas; *Trans. Bose Res. Inst.* **9** 132–176
- Basu P C 1934 The racial affinities of the Oraon; *Trans. Bose Res. Inst.* **9** 132–176
- Bhattacharjee P N 1956 A genetic survey in the Rahri Brahmans and the Muslim of West Bengal: A<sub>1</sub>A<sub>2</sub>BO, MN, Rh, blood groups, ABH secretion, sickle cell, P.T.C. taste, middle phalangeal hair and colour blindness; *Bull. Anthropol. Surv. India* **5** 18–28
- Bhattacharjee P N 1969 A genetical study of the Santals of Santal Parganas; *Anthropologist* (Spl. Vol.) 93–103
- Bhattacharjee P N 1975 Sero-genetical study of the Khasi and their genetic relationship; *East. Anthropol.* **28** 171–177
- Bhattacharjee P N and Kumar N 1969 A blood group genetic survey in the Dudh Kharia of the Ranchi district (Bihar India); *Hum. Herd.* **19** 385–397
- Bhowmick P K 1956 Physical affinity of the Lodhas of Midnapore; *Man India* **36** 110–131
- Buechi E C 1954 ABH secretion in India; *Anthropologist* **1** 3–9
- Cavalli-Sforza L L, Menozzi P and Piazza A 1994 *The history and geography of human genes* (Princeton, NJ: University Press)
- Chakroborty R, Das S K and Roy M 1975 Blood group genetics of some caste groups of southern 24 Parganas West Bengal; *Hum. Hered.* **25** 218–225
- Chandra R 1916 *The Indo Aryan races: A study of the origin of Indo-Aryan people and institutions* (New Delhi: Cosmo Publications)
- Chaudhuri S, Mukherjee B, Ghosh J and Roychoudhury A K 1969 Study of blood groups ABH secretors and haemoglobin variants in the three upper castes of West Bengal India; *Am. J. Phys. Anthropol.* **30** 129–132
- Chu J Y, Huang W, Kuang S Q, Wang J M, Xu J J, Chu Z T, Yang Z Q, Lin K Q, Li P, Wu M, Geng Z C, Tan C C, Du R F and Jin L 1998 Genetic relationship of populations in China; *Proc. Natl. Acad. Sci. USA* **95** 11763–11768
- Clark V J, Sivendren S, Saha N, Bentley G R, Aunger R, Sirajuddin S M and Stoneking M 2000 The 9-bp deletion between the mitochondrial lysine tRNA and COII genes in tribal populations of India; *Hum. Biol.* **72** 273–285
- Das S R and Bhattacharjee P N 1963 Blood groups (A<sub>1</sub>A<sub>2</sub>BO), ABH secretion, sickle cell, PTC taste and colour blindness in the rajbanshi of Midnapur district West Bengal; *Bull. Anthropol. Surv. India* **12** 1–6
- Das S R, Bhattacharjee P N, Sastry D B and Mukherjee D P 1963 Blood groups (ABO, MN, Rh), ABH secretion sickle cell trait and colour blindness in the Bado Gadaba and the Pareng Paroja of Koraput district in Orissa; *Bull. Anthropol. Surv. India* **11** 145–151
- Das S R, Mukherjee D P and Bhattacharjee P N 1967 Survey of blood groups and PTC taste among the rajbanshi caste of West Bengal (ABO, MNS, Rh, Duffy, Diego); *Acta Genet. Stat. Med.* **17** 433–445
- Das S R, Mukherjee D P and Sastry D B 1968 A somatological survey of five tribes in the Koraput district Orissa; *Bull. Anthropol. Surv. India* **17** 400–422
- Das S R, Sastry D B and Mukherjee D P 1966 Blood groups (A<sub>1</sub>A<sub>2</sub>BO, MN, Rh) and ABH secretion in the Pareng Gadaba

- the Ollaro Gadaba and the Konda Paroja of Koraput district in Orissa; *Acta Genet. Stat. Med.* **16** 169–183
- Deep Kumar V S, Rao P P, Ramesh A and Reddy A P 1985 A genetic study of the Kamma and Kapu caste of Andhra Pradesh; *J. Indian Anthropol. Soc.* **20** 59–65
- Deraniyagala S U 1985 The prehistory of Sri Lanka An outline; in *James Thevathsan Ruthnam Festschrift* (eds) A R B Amerasinghe and S J Sumansekara Banda (Sri Lanka UNESCO National Commission) pp 14–21
- Deraniyagala S U 1986 Prehistoric Sri Lanka The late Quaternary; in *The Pleistocene Perspective* (London: World Archaeology Congress) pp 1–39
- Diamond J 1997 Location, location, location: The first farmers; *Science* **278** 1243–1244
- Dixon R B 1992 The Khasi and the racial history of Assam; *Man India* **2** 2–13
- Dutta P C and Gupta P 1967 Finger print patterns of the Austra tribe of Chotanagpur India; *Acta Genet. Basel.* **17** 248–259
- Gadgil M, Joshi N, Manoharan S, Patil S and Prasad U V S 1998 Peopling of India; in *The human heritage* (eds) D Balasubramanian and N A Rao (Hyderabad: Hyderabad University Press) pp 100–129
- Ganguli P 1976 *Physical anthropology of the Nicobarese* Memoir No 34 (Calcutta: Anthropological Survey of India)
- Gordon D H 1958 *The pre-historic background of Indian culture* (Mumbai: Bhulabai Memorial Institute)
- Grierison G A 1906 *Linguistic surveys of India. IV. Munda and Dravidian languages* (Calcutta: Office of Superintendent Gov. printing)
- Guha B S 1933 Anthropometric notes on Telugu Brahmins and Kumits; in *Census of India 1931*
- Guha B S 1935 The racial affinities of the people of India; in *Census of India 1931*
- Guha B S 1937 An outline of Racial Ethnology of India; in *An outline of field sciences in India* (ed.) S L Hora (Calcutta) pp 125–139
- Gupta P and Basu A 1960 Physical characters of the Madiga of Guntur – A Pariah caste of Andhra; *Bull. Anthropol. Surv. India* **9** 53–60
- Gupta S P 1962 Analysis of some aspects of human biology of Bhumij, Bathudi and Savar of Dhalbhum; *Bull. Bihar Tribal Res. Inst.* **4** 69–96
- Gupta S P 1979 *Archaeology of Soviet Central Asia and the Indian borderlands* (Delhi: B R Publishing)
- Gupta S P and Dash Sharma P 1973 A genetic survey of the Car-Nicobarese of Nicobar Inlands; *Man India* **53** 315–320
- Hargrave D C 1963 *The distribution of human blood groups and other polymorphisms* (London: Oxford University Press)
- Harpending H C and Ward R 1982 Chemical systematics and human evolution; in *Biochemical aspects of evolutionary biology* (ed.) M Nitecki (Chicago: Chicago University Press) pp 213–256
- Hooton E A 1930 *The Indians of Pucos Pueblo* (New Haven)
- Hutton J H 1933 *Census of India* Vol. I, Report 1931, Delhi
- Kennedy K A R 1984 Biological adaptations and affinities of Mesolithic South Asians; in *The people of south Asia. The biological anthropology of India Pakistan and Nepal* (ed.) J R Lukas (New York, London: Plenum Press) pp 29–57
- Kennedy K A R 2000 *God-Apes and fossil men. Paleoanthropology in South Asia* (Ann Arbor: University of Michigan Press)
- Kirk R L, Lai L Y C, Vos G S and Vidyarthi L P 1962 A genetical study of the Oraons of the Chotanagpur plateau (Bihar India); *Am. J. Phys. Anthropol.* **20** 375–386
- Kumar N and Bhattacharjee P N 1976 Blood groups of the Munda of Ranchi district in Bihar (India); *Indian J. Hum. Genet.* **2** 201–214
- Kumar N and Mukherjee D P 1975 Genetic distances among the Ho tribe and other groups of Central India; *Am. J. Phys. Anthropol.* **42** 489–494
- Kumar N and Sastry D B 1961 A genetic survey among the Riang, a Mongloid tribe of Tripura; *Z. Morphol. Anthropol.* **51** 346–355
- Lal B B 1956 Paleolithics from the Beas and Banganga valleys Punjab; *Ancient India* **12** 58–92
- Lapicque L 1905 Note Semmaire sur une mission ethnologique dans le Sud de l'Inde; *Bull. Mus. Nat. Hist.* **2**
- Mahalanobis P C, Majumdar D N and Rao C R 1948–49 Anthropometric survey of the United Provinces 1941 A statistical study; *Sankhya* **9** 89–324
- Majumdar D N and Sen A R 1950 Anthropometric status of castes and tribes; in *Race realities in cultural Gujrat* (Gujrat: Research Society Part II)
- Majumdar P P, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta P G and Sil S K 1999 Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications; *Eur. J. Hum. Genet.* **7** 435–446
- Majumdar P P 2001 Ethnic populations of India as seen from an evolutionary perspective; *J. Biosci.* **26** 533–545
- Miki T, Tanaka T, Hayashida Y and Furuhashi T 1960 On the distribution of secretor character of the Lepchas and the Khasis; *Proc. Jpn. Acad.* **36** 229–233
- Mitra A K 1956 The Riang of Tripura; *Bull. Anthropol. Surv. India* **5** 21–120
- Mohapatra G C 1975 Acheulian Element in Soan Culture Area; *J. Archaeol. Soc. Nippon* **60** 4–18
- Mohapatra G C 1985 Cultural ecology of early man in India; in *Recent advances in Indian archaeology* (eds) S B Deo and K Paddaya (Poona: Deccan College) pp 23–73
- Naidu J M and Veerajay P 1982 Blood groups among Brahmin and Kamma caste populations of coastal Andhra Pradesh; *Am. J. Phys. Anthropol.* **58** 197–200
- Negi R S and Ahmad S H 1963 A study of A<sub>1</sub>A<sub>2</sub>BO, M-N, Rh blood groups, A-B-H secretion in saliva and colour blindness in the Doria and Dhurwa of Bastar; *Bull. Anthropol. Surv. India* **12** 21–29
- Nei M 1987 *Molecular evolutionary genetics* (New York: Columbia University Press)
- Nei M and Ota T 1991 Evolutionary relationships of human population at molecular level; in *Evolution of life fossils molecules and culture* (eds) S Osawa and T Honjo (Heidelberg: Springer-verlag) pp 415–428
- Nei M, Tajima F and Tetemo Y 1983 Accuracies of estimated phylogenetic trees from molecular data; *J. Mol. Ecol.* **19** 153–170
- Papiha S S, Roberts D F and Mishra S C 1988 Serogenetic studies among an urban and two tribal populations of Orissa India; *Ann. Hum. Biol.* **15** 143–152
- Passarino G, Semino O, Bernini L F and Santachiara-Benerecetti A S 1996 Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms; *Am. J. Hum. Genet.* **59** 927–934
- Penrose L S 1954 Distance, size and shape; *Ann. Eugen.* **18** 337–343

- Qunitana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K and Santachiara-Benerecetti A S 1999 Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa; *Nat. Genet.* **23** 437–441
- Ramesh A, Blake N M, Vijaykumar M and Murthy J S 1980 Genetic studies on the Chenchu tribe of Andhra Pradesh India; *Hum. Hered.* **30** 291–298
- Ray A K 1958 A note on Juang anthropometry; *Man India* **38** 56–59
- Ray G S 1946 A study of physical characteristics of the Rajbansis; *J. R. Asiat. Soc. Bengal Lett.* **12** 31–42
- Ray G S 1949 A study of physical characters of the Khonds; *East. Anthropol.* **2** 138–144
- Raychoudhury T C 1935 The Khasis; *Anthropol. Papers J. Dept. Lett. Univ. Calcutta* **26** 249–272
- Roy S C 1915 *The Oraons of Chota Nagpur: Their history economic life and social organisation with an introduction by C A Haddon* (Ranchi)
- Roychoudhuri S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran N, Usha Rani M V, Vishwanathan H, Mitra M, Sil S K and Majumder P P 2000 Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA; *Curr. Sci.* **79** 1182–1191
- Sarkar S S 1935–36 Racial affinities of the inhabitants of Rajmahal Hills; *Trans. Bose Res. Inst.* **11** 123–153
- Sarkar S S 1954 *The Aboriginal Races of India* (Calcutta: Bookland)
- Sergent B 1997 *Genese de l'Inde* (Paris: Payot)
- Simmons R J, Graydon J J, Semple N M and D'Sena G W L 1953 A genetical survey in Chenchu south India blood test and secretion; *Med. J. Aust.* **1** 497–503
- Singhvi A K, Deraniyagala S U and Sengupta D 1986 Thermoluminescence dating of Quaternary red sand beds in Sri Lanka; *Earth Planet. Sci. Lett.* **80** 139–144
- Singhvi L D 1953 Comparison of genetical and morphological methods for a study of biological differences; *Am. J. Phys. Anthropol.* **11** 385–404
- Polunin I and Sneath P H A 1953 Studies of blood group in South-east Asia; *J. R. Anthropol. Inst.* **83** 215–251
- Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z and Jin L 1999 Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age; *Am. J. Hum. Genet.* **65** 1718–1724
- Thapar R 1966 *A history of India* Vol. 1 (Middlesex: Penguin)
- Watkins W S, Bamshad M, Dixon M E, Rao B B, Naidu J M, Reddy P G, Prasad B V R, Das P K, Reddy P C, Gai P B, Bhanu A, Kusuma Y S, Lum J K, Fisher P and Jorde L B 1999 Multiple origins of the mtDNA 9-bp deletion in populations of South India; *Am. J. Phys. Anthropol.* **109** 147–158

*MS received 29 October 2002; accepted 15 April 2003*

Corresponding editor: VIDYANAND NANJUNDIAH