Ethnic populations of India as seen from an evolutionary perspective

PARTHA P MAJUMDER

Anthropology and Human Genetics Unit, Indian Statistical Institute, 203 BT Road, Kolkata 700 035, India

(Fax, 91-33-5773049; Email, ppm@isical.ac.in)

It is now widely accepted that (i) modern humans, *Homo sapiens sapiens*, evolved in Africa, (ii) migrated out of Africa and replaced archaic humans in other parts of the world, and (iii) one of the first waves of out-of-Africa migration came into India. India, therefore, served as a major corridor for dispersal of modern humans. By studying variation at DNA level in contemporary human populations of India, we have provided evidence that mitochondrial DNA haplotypes based on RFLPs are strikingly similar across ethnic groups of India, consistent with the hypothesis that a small number of females entered India during the initial process of the peopling of India. We have also provided evidence that there may have been dispersal of humans from India to southeast Asia. In conjunction with haplotype data, nucleotide sequence data of a hypervariable segment (HVS-1) of the mitochondrial genome indicate that the ancestors of the present austro-asiatic tribal populations may have been the most ancient inhabitants of India. Based on Y-chromosomal RFLP and STRP data, we have also been able to trace footprints of human movements from west and central Asia into India.

1. Introduction

Data generated by the human genome sequencing project indicate that any two randomly drawn humans are genetically about 99·9% identical. Human geneticists who are intensively studying human genomic diversity engage themselves with a tiny fraction (about a 10th of one per cent) of the human genome, which some may consider as an insignificant endeavour. However, it is this small fraction that confers an element of uniqueness to every human. It is primarily this fraction on which various evolutionary forces, particularly natural selection, has acted on during the period of evolution of modern humans from its most recent common ancestor. Differences in this small fraction make some individuals susceptible to a disease, while conferring protection to others from the same disease. The study of human genomic variation among individuals, can help us understand the nature and intensity of actions of various forces that have modulated our evolutionary course. It can also provide valuable data for the understanding of various diseases that afflict us today. This paper looks at the contributions of molecular genetics to the understanding of modern human origins on the Indian sub-continent.

India occupies a centerstage in human evolution. It has served as a major corridor for the dispersal of modern humans that started from Africa about 100,000 years ago (Cann 2001). The date of entry of modern humans into India remains uncertain. However, modern human remains dating back to the late Pleistocene (55000–25000 years before present, ybp) have been found (Kennedy *et al* 1987) and by the middle paleolithic period (50,000–20,000 ybp), humans appear to have spread to many parts of India (Misra 1992, 2001). We (Majumder *et al* 1999) have recently provided molecular genetic evidence that a major population expansion of modern humans took place within India. Although the period of this demographic expansion remains uncertain, it has been speculated (Mountain *et al* 1995) that the event took place 60,000–85,000 ybp. Perhaps this expansion, followed by subsequent migration, resulted in the peopling of southeast Asia and later (50,000–60,000 ybp), of Australia (Crow 1998). About 60,000 ybp, there is believed to have been another independent expansion of modern humans in southern China (Ballinger *et al* 1992; Crow 1998), which may have resulted in human migration into India and also into southeast Asia.

Keywords. Demography; migration; mitochondrial DNA; polymorphism; Y-chromosome

India is a land of enormous genetic, cultural and linguistic diversity. With the exception of Africa, India harbours more genetic diversity than other comparable global regions (Majumder 1998). The enormous diversity in social and cultural beliefs and practices has been well documented and emphasized (Karve 1961; Beteille 1998). The population of India is culturally stratified, broadly into tribals and non-tribals. It is generally accepted that the tribal people, who constitute 8·08% of the total population (1991 Census of India), are the original inhabitants of India (Thapar 1966; Ray 1973). There are an estimated 461 tribal communities in India (Singh 1992), who speak about 750 dialects (Kosambi 1991) which can be classified into one of the following three language families: Austro-Asiatic (AA), Dravidian (DR) and Tibeto-Burman (TB). There is considerable debate about the evolutionary histories of the Indian tribals. The proto-Australoid tribals, who speak dialects belonging to the Austric linguistic group, are believed to be the basic element in the Indian population (Thapar 1966, p. 26). Other anthropologists, historians and linguists (Risley 1915; Rapson 1955; Pattanayak 1998) have supported the view that the Austro-Asiatic (a subfamily of the Austric language family) speaking tribals are the original inhabitants of India. Some scholars (Buxton 1925; Sarkar 1958) have, however, proposed that the Dravidians are the original inhabitants, the Austro-Asiatics being later immigrants. The Austro-Asiatic family is a fragmented language group. It is most widely spoken in Vietnam and Cambodia. Within India, only a small number of ethnic groups speak Austro-Asiatic languages. It is noteworthy that the Indian Austro-Asiatic speakers are exclusively tribal, which may be indicative of their being the oldest inhabitants of India (Pattanayak 1998; Gadgil *et al* 1998). Some believe that the Austro-Asiatic linguistic family evolved in southern China (Diamond 1997). If this is true, Indian Austro-Asiatic speakers must have entered India from southern China through the northeast. Many linguists (Renfrew 1987; Ruhlen 1991) contend that Elamo-Dravidian languages may have originated in the Elam province of southwestern Iran, and the dispersal of the Dravidian languages into India took place with migration of humans from this region who brought with them the technologies of agriculture and animal-domestication. The Tibeto-Burman speaking tribals, who primarily inhabit the northeast regions of India, are supposedly immigrants to India from Tibet and Myanmar (Guha 1935).

Contemporary non-tribal populations of India tend to belong to the overall Hindu religious fold and are hierarchically arranged in four main caste classes, viz. Brahmin (priestly class), Kshatriya (warrior class), Vysya (business class) and Sudra (menial labour class). In addition, there are several religious communities, who practice different religions, for instance Islam, Christianity, Sikhism, Buddhism, Jainism and Judaism. The non-tribals predominantly speak languages that belong to the Indo-Aryan or Dravidian families. These two linguistic groups have been the major contributors to the development of Indian culture and society (Meenakshi 1995), which have also been affected by multiple waves of migration that took place in historic and prehistoric times (Ratnagar 1995; Thapar 1995).

2. Testing anthropological and linguistic hypotheses using genomic data

As evident from the foregoing discussion, there are differences of opinion among anthropologists and linguists regarding the origins of Indian ethnic groups. During the past several years, we have attempted to see whether the differences might be resolved, at least in part, using genomic data. For the purpose of generating the data, we have obtained, with consent, blood samples of individuals drawn from a large number of population groups of India, of diverse geographical, linguistic and ethnic backgrounds. DNA was isolated from each of these blood samples and screened for various polymorphic markers using standard molecular genetic protocols (PCR amplification, restriction digestion, fragment visualization under UV transillumination, DNA sequencing). These data were then summarized either as allele or as haplotype frequencies, which were then statistically analysed to draw appropriate inferences. In some cases, raw DNA sequence data were statistically analysed.

The specific hypotheses that we have sought to test, and questions that we have asked, are as follows. (i) If indeed modern humans arrived early in India, then it is expected that there will be extensive sharing of ancient polymorphisms. Evidence in favour of this expectation will indicate a fundamental genomic similarity among humans in India. (ii) If India has served as a major corridor of migration of modern humans from Africa, where did modern humans move to using this corridor? (iii) Are the Austro-Asiatics the oldest inhabitants of India? (iv) Is there any evidence of human migration to India from central and west Asia?

In the following sections, we provide genomic evidence, primarily from our own work, pertaining to these issues.

At this juncture, it may be useful to highlight some aspects of our data and limitations of our analyses. First, the traditional definition of an ethnic group as a group of individuals who share common cultural beliefs and practices and who are largely intra-marrying, leaves room for its unqualified acceptance in genomic research from an evolutionary perspective. Without getting into details, it must be emphasized that ethnic groups in India often form genetic subgroups, usually due to geographical

isolation and/or social regulations governing matings. Therefore, any general statement on an ethnic group must rule out the existence of such subgroups, or must be confined to the subgroups that have been studied. Identifying subgroups from a genetic standpoint is, by itself, a major scientific endeavour. In the present paper, we have used names of ethnic groups without investigating the existence of genetic substructuring within the groups. While this may be acceptable for providing insights in relation to some macro-level questions such as the ones entertained in the present study, certainly fine-tuning of our inferences may need to be done in the future. Second, because our studies and questions have evolved over a period of time, our choice of ethnic groups has sometimes been opportunistic, and sometimes more focussed in relation to our hypotheses. Therefore, not all hypotheses have been tested using DNA samples drawn from the same set of ethnic groups. We do not consider this to be a limitation, as not all hypotheses need to be or can be tested using a uniform set of ethnic groups. Third, although from each group we have sampled individuals who were unrelated at least to the first-cousin level, because we have sampled them from restricted geographical areas, it is possible that our sample may not represent the entire spectrum of genomic diversity of the group, especially if there are subgroups within the group. Fourth, our sample sizes from some of the ethnic groups are small. Although our inferences have been based on statistical tests that take variable sample sizes into account, we acknowledge that larger sample sizes would have added robustness to our inferences. However, it may be pointed out that it is well-known in population genetics that in comparison with inferences based on data of a large number of individuals, inferences based on data of a smaller number of individuals but a larger number of loci are equally robust. We have, therefore, studied a larger number of loci to compensate for our restricted sample sizes.

3. Fundamental genomic unity of India

Since the seminal study of Cann *et al* (1987), mitochondrial DNA (mtDNA) data have proven to be extremely useful in the study of human evolution, including prehistoric migrations and demographic events such as sudden population expansion or extreme bottlenecks (Sherry *et al* 1994). In other words, mtDNA enables to probe the distant past.

We have studied 644 mtDNA samples collected from 23 ethnic populations; 10 populations from the eastern states of West Bengal (5 populations), Orissa (4 populations) and Tripura (1 population), 1 population from the central state of Madhya Pradesh, 4 populations from the northern state of Uttar Pradesh, and 8 populations from the southern state of Tamil Nadu. These populations were chosen to include both tribal and caste populations at different levels of social hierarchy. One group of Muslims from Uttar Pradesh has also been included. The tribal populations belong to three different linguistic groups (Austro-Asiatic, Dravidian and Tibeto-Burman), and the caste populations are either Indo-Aryan speakers (northern Indian castes) or Dravidian speakers (southern Indian castes). The Muslims studied are all Indo-Aryan language speakers.

We have screened 8 mtDNA loci. The 9-bp COII/$tRNA^{Lys}$ intergenic length mutation revealed that all populations were monomorphic; no sampled individual showed the presence of the 9 bp deletion. The remaining seven RFLP loci were polymorphic in the pooled data set (see footnote of table 1 for details). Seven-locus haplotypes were constructed and their frequencies estimated in each population. A total of 19 different haplotypes were observed in the pooled data set. However, in none of the populations were all 19 haplotypes observed. The maximum number of haplotypes (13) was observed among Rajputs; the Kotas harboured only 2 haplotypes. Frequencies of haplotypes in each study population, as also in the pooled sample, are presented in table 1. The frequency distribution of haplotypes in the pooled data set is nearly unimodal; only one haplotype (0111011) accounted for about 50% of all mtDNA molecules. It can, therefore, be inferred that this is the most ancient haplotype in Indian populations. It is also seen from table 1 that in 20 of the 23 study populations, this modal haplotype is the most frequent. The three populations in which this haplotype is not the most frequent are all inhabitants of Uttar Pradesh in northern India – Brahmins, Rajputs and Muslims. Among Brahmins and Rajputs of Uttar Pradesh, the most frequent haplotype is 0100011. Among Muslims of Uttar Pradesh the most frequent haplotype is 0100111. However, all these three populations also harbour the 0111011 haplotype, which is modal in the remaining 20 populations, in fairly high frequencies (15%–25%).

Most of the mtDNA diversity observed in Indian populations are between individuals within populations; there is no significant structuring of haplotype diversity by socio-religious affiliation, geographical location of habitat or linguistic affiliation (Roychoudhury *et al* 2000).

The extensive sharing of one or two haplotypes across population groups within India, irrespective of their geographical location, linguistic affinity or social proximity reveals a fundamental unity of mtDNA lineages in India in spite of the extensive cultural and linguistic diversity. Since mtDNA is maternally inherited, we believe that there was a relatively small number of founding female lineages in India (Roychoudhury *et al* 2000). Ethnic differentiation took place subsequently through a series of

demographic expansions, geographic dispersal and social groupings. The lack of correspondence of clusters based on mtDNA haplotype frequencies with either geographical location of habitat, language or social proximity is consistent with such a model for the peopling of India. Further, because of the extensive haplotype sharing among ethnic groups, the extent of observed variation in haplotype frequencies attributable to differences between groups is small; most observed haplotype variation is between individuals within groups (Roychoudhury *et al* 2000).

4. Where to?

We have also compared the distributions of haplotypes found in the populations included in the present study with those of other populations of southeast Asia. For this purpose, we collated and compacted the haplotype data presented in Ballinger *et al* (1992). Since Ballinger *et al* (1992) did not study the RFLP site at nt 12308 studied by us, this locus had to be excluded for purposes of comparison. The results, presented in figure 1, show that there is a considerable sharing of haplotypes between Indian and southeast Asian populations. The distributions of haplotype frequencies are also similar. There is, however, one notable difference. The southeast Asian populations harbour a set of haplotypes, albeit with low or medium frequencies, on the 9 bp deletion background, which are completely absent in the present study populations. Most of these haplotypes are also on a DdeI(10394)-AluI(10397) –/– background. Ballinger *et al* (1992) have hypothesized that the 9 bp deletion arose in central China and radiated out from this region as migrants moved to populate parts of southeast Asia. If India was also populated by migrants radiating out from central China, one would have expected that a significant proportion of the migrants would carry the (–/– 9-bp-*del*) haplotype; hence this haplotype should be present in Indian populations in polymorphic frequencies. However, this haplotype has not been observed in any of the populations investigated in the present study, nor was it detected in an earlier study (Rickards 1995). On the other hand, a significant proportion of the southeast Asian populations possess the 9-bp 'non-deletion' allele on DdeI(10394)-AluI(10397) +/+ or +/– backgrounds. In fact, the two classes of haplotypes observed in southeast Asian populations (see figure 1), one of which is completely absent in Indian populations, leads us to believe that southeast Asian populations were derived from two sources; one from India and the other possibly from central or southern China. It may be noted that the 9 bp deletion is present in high frequencies among Tharus (Passarino *et al* 1993) and Japanese (Cann *et al* 1987; Horai and Matsunaga 1986) populations that are postulated to have arisen from human migrations originating from southern China. It is known (Beteille 1998; Diamond 1997) that there were two waves of human migration from mainland Asia through southeast Asia to New Guinea and Java. One of these was an early wave that occurred about 40,000 ybp. The other wave, the so-called Austronesian migration from south China, took place about 4000–3500 ybp. Although we cannot be certain, we postulate that the early wave of migration was from India and carried the +/+ haplotype into southeast Asia. The second wave of migration from south China may have carried the (–/– 9-bp-*del*) haplotype into this region. An early wave of migration from India, actually from Africa through India, to southeast Asia has also been proposed in a recent study (Chu *et al* 1998) using nuclear DNA microsatellite markers and subsequently supported by a study using Y-chromosomal DNA markers (Su *et al* 1999).

5. Are Austro-Asiatic speakers the oldest inhabitants of India?

A segment of the mtDNA, known as the hypervariable segment 1 (HVS1), is a fast-evolving stretch of about 400 nucleotides and has proven to be particularly useful in the study of short-term evolution. We have carried out DNA sequencing of HVS1 in 115 individuals belonging to various linguistically distinct tribal populations of India. The tribal groups were (i) Austro-Asiatic (AA) speakers: Santal (SA), Munda (MU), Lodha (LO); (ii) Dravidian (DR) speakers: Muria (MR), Kota (KT), Kurumba (KR), Irula (IR); and (iii) Tibeto-Burman (TB) speakers: Tipperah (TR). These tribal communities inhabit the eastern (SA, MU, LO), southern (KT, KR, IR), central (MR) and northeastern (TR) regions of India. Comparable data were obtained on a stretch of 338 nucleotides of the HVS1 segment. Deletions of one or two nucleotides were noted, at positions 16182 and 16183, in several individuals. These two nucleotides were not considered in the statistical analyses of the sequence data. Among the 115 individuals, there were a total of 104 mutations at 94 polymorphic sites. Twelve sequences were shared by at least two individuals. Sharing of sequences was primarily between individuals within the same population, although three of these sequences were shared among individuals belonging to different populations. No sequence was shared between populations belonging to different language groups. From these sequence data, we calculated nucleotide diversities and statistics (mean number of mismatches and raggedness) pertaining to mismatch distributions. The Austro-Asiatics exhibited the maximum genomic diversity (table 2), both nucleotide diversity and mean number of mismatches. We estimated the parameters of a population

Table 1. Absolute and percentage frequencies of 7-locus mtDNA haplotypes in 23 ethnic populations of India.

Haplotype Type*	AGH	AMB	BAG	BRA (UP)	BRA (WB)	CHA	GAU	ILA	IYN	IYR	KOT	KUR	LOD	MAH	MUN	MUR	MUS	PLN	RAJ	SAN	TAN	TRI	VAN	Total
0111111							1											2						3
							7·7											6·7						0·47
0110111							2												1·0					3
							15·4												2·0					0·47
0100111	2		2	5	1	6	1	1		4		2	6	2		1	8	1	6		1	2	3	54
	8·3		6·5	18·5	4·6	24·0	7·7	3·3		13·3		6·7	18·7	6·1		3·3	28·6	3·3	11·7		6·25	4·44	10·0	8·39
1100110														1·0										1·0
														3·0										0·16
0100110	3	2	2	2		2	1	6	3	3				4		1	1	1	2	2	1	1		37
	12·5	6·7	6·5	7·4		8·0	7·7	20·0	10·0	10·0				12·1		3·3	3·6	3·3	3·9	10·0	6·25	2·22		5·75
0111011	17	17	15	4	12	14	4	14	15	14	29	23	15	14	2	23	7	17	10	12	12	14	12	316
	70·8	56·7	48·4	14·8	54·6	56·0	30·8	46·7	50·0	46·8	96·7	76·7	46·9	42·5	28·6	76·7	25·0	56·7	19·6	60·0	75·0	31·1	40·0	49·07
0011011				1		1									1				3			2		8
				3·7		4·0									14·2				5·9			4·44		1·24
0110011			1		3									1					3					8
			3·2		13·6									3·0					5·9					1·24
1100011												1·0												1·0
												3·3												0·16
0100011			6	11	3	1	1	7	10	4	1	4		8	2	4	3	6	13	3	1	11	6	105
			19·3	40·8	13·6	4·0	7·7	23·4	33·4	13·3	3·3	13·3		24·2	28·6	13·4	10·6	20·0	25·4	15·0	6·25	24·4	20·0	16·3
0111001								1·0		1·0									2·0			1	4·0	5·0
								3·3		3·3									3·9			2·22	13·3	0·78
0100001				1															1			1·0		3
				3·7															2·0			2·22		0·47
0111010	1	10	4		1		1	1					11	1	2	1	1	3	1	2	1	4		49
	4·2	33·3	12·9		4·6		7·7	3·3					34·4	3·0	28·6	3·3	3·6	10·0	2·0	10·0	6·25	8·89		7·61
0011010						1											2					2		5
						4·0											7·1					4·44		0·78
0110010									1								1		1					3
									3·3								3·6		2·0					0·47
1100010				1						1				2			4		3			4		15
				3·7						3·3				6·1			14·3		5·9			8·89		2·33
0100010	1	1	1	1	2		2		1	3							1		5	1		1	5	25
	4·2	3·3	3·2	3·7	9·0		15·3		3·3	10·0							3·6		9·8	5·00		2·22	16·7	3·88
0000010				1																				1
				3·7																				0·16
0000000																						2		2
																						4·44		0·31

*Order of loci: HaeIII (663), AluI (5176), DdeI (10394), AluI (10397), HinfI (12308), HincII (13259), HaeIII (16517).
1, Present of restriction site; 0, absence of restriction site.
Population codes and descriptions: AGH, Agharia (caste, east); AMB, Ambalakarer (caste, south); BAG, Bagdi (caste, east); BRA (UP), Brahmin-Uttar Pradesh (caste, north); BRA (WB), Brahmin–West Bengal (caste, east); CHA, Chamar (caste, north); GAU, Gaud (caste, east); ILA, Irula (tribe, south); IYN, Iyengar (caste, south); IYR, Iyer (caste, south); KOT, Kota (tribe, south); KUR, Kurumba (tribe, south); LOD, Lodha (tribe, east); MAH, Mahishya (caste, east); MUN, Munda (tribe, east); MUR, Muria (tribe, Central); MUS, Muslim (religious group, north); PLN, Pallan (caste, south); RAJ, Rajput (caste, north); SAN, Santal (tribe, east); TAN, Tanti (caste, east); TRI, Tipperah (tribe, northeast); VAN, Vanniyar (caste, south).

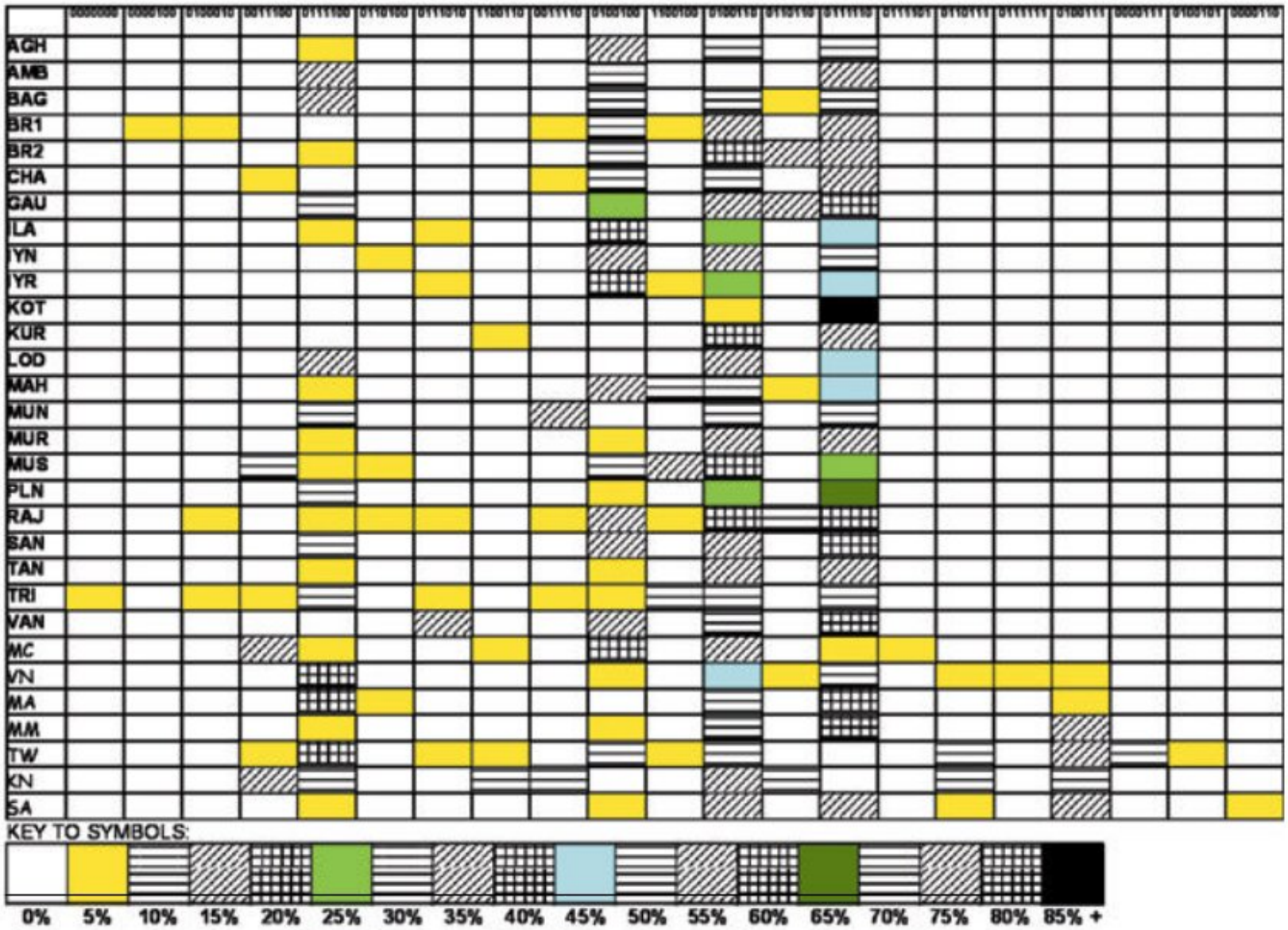

Figure 1. mtDNA haplotype diversities in 23 ethnic populations of India.

expansion model (Harpending *et al* 1993) and examined the fit of the observed and expected mismatch distributions for the three linguistic groups of tribals. The observed and expected distributions are presented in figure 2(a–c). From the unimodal nature of the observed mismatch distributions, their smoothness [as revealed by the very small values of the raggedness statistic (Harpending *et al* 1993); table 2] and the reasonably good fits with expected distributions, it is clear that there were significant expansions of these linguistic groups of tribals. To detect traces of population expansions, we also used a second approach. We computed Fu's (1997) F_S statistic, which is particularly sensitive to population growth. Significantly large negative values indicate population expansion (Fu 1997), which is what is observed (table 2) in our data set for each of the three language groups. We estimated the expansion times, which are also presented in table 2, using the methodology proposed by Slatkin and

Table 2. Descriptive statistics and estimated expansion times of Indian tribals belonging to various language groups and haplogroups.

Linguistic group	No. of sequences	No. of polymorphic sites	No. of mutations	Nucleotide diversity (π) ± 2 SD	Mean No. of mismatches (k)	Raggedness (r)	F_S (P-value)	Expansion time in years before present (95% confidence interval)
Austro-Asiatic	34	54	59	0·023 ± 0·002	7·747	0·0157	– 19·176 (0·000)	56098 (51220–60975)
Dravidian	61	59	66	0·016 ± 0·002	5·432	0·0058	– 25·417 (0·000)	39024 (34146–43902)
Tibeto-Burman	20	42	42	0·021 ± 0·001	7·170	0·0281	– 12·203 (0·000)	51220 (48780–53659)

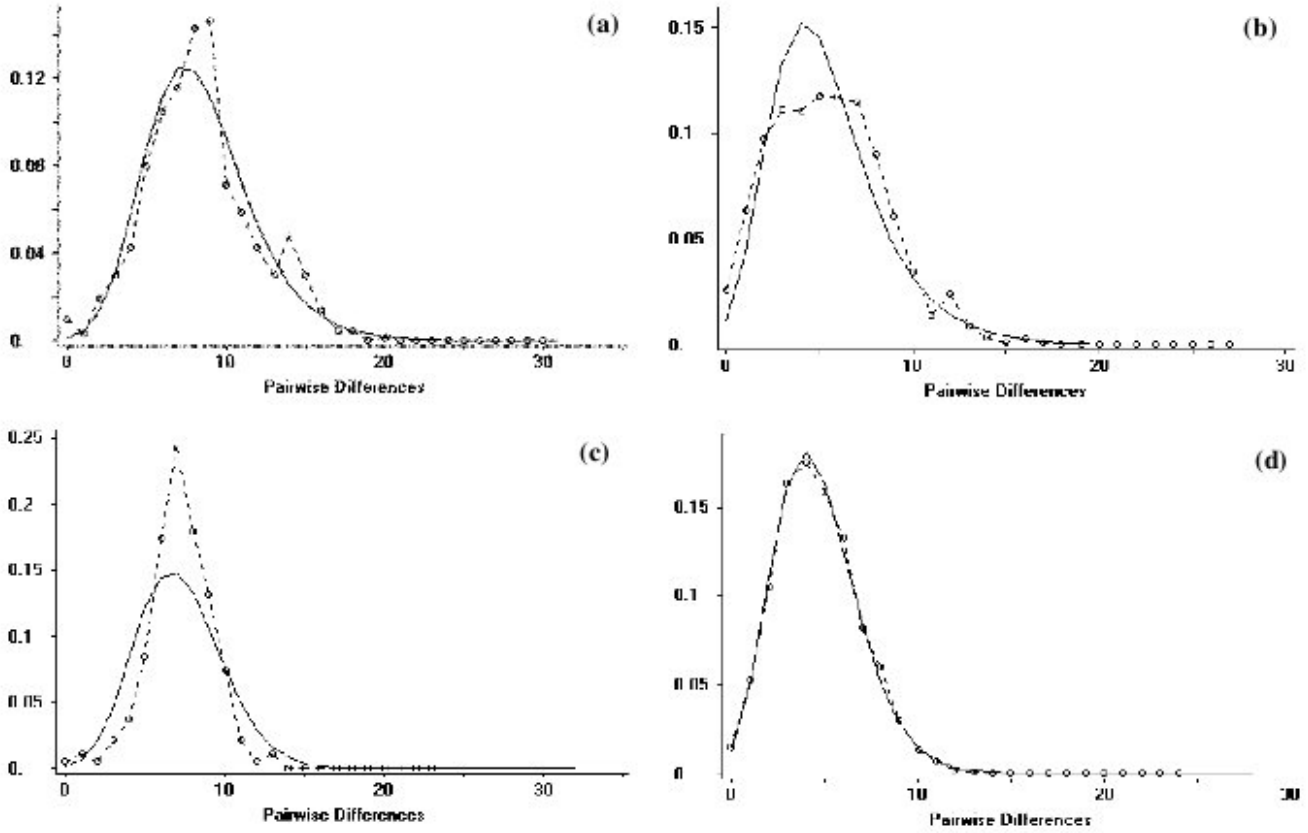

Figure 2. Observed (--o--o--) and expected (——) mismatch distributions based on mtDNA HVS-1 sequences for (**a**) Austro-asiatic, (**b**) Dravidian, (**c**) Tibeto-Burman speaking tribals of India, and (**d**) tribals belonging to mtDNA haplogroup M.

Hudson (1991) assuming a mutation rate of 20·5% per site per million years (which is appropriate for the HVS1 region; Bonatto and Salzano 1997). The 95% confidence interval of an estimated expansion time was taken to be twice the standard deviation of the sampling variance of nucleotide diversity. The estimated expansion time of the Austro-Asiatics (≈56,000 ybp) is much older than that of the Dravidians (≈39,000 ybp). This difference is significant as indicated by the disjoint 95% confidence intervals of the estimates. Our tentative estimate, in view of the limited sample size, of the expansion time of the Tibeto-Burmans is ≈ 52,000 ybp.

Although anthropologists, archaeologists and historians accept that the tribal populations are the original inhabitants of India, most studies on Indian populations using DNA markers have not included the tribals. It has been argued (Risley 1915; Thapar 1966; Pattanayak 1998) that the Austro-Asiatic speaking tribals are the original inhabitants of India. Some other scholars have, however, argued that tribal groups speaking Dravidian and Austro-Asiatic languages have evolved from an older original substrate of proto-Australoids (Keith 1936), while the Tibeto-Burman speaking tribals are later immigrants from Tibet and Myanmar (Guha 1935). Parpola (1975) has contended that the different language families in India may represent different lineages, which is consistent with Cavalli-Sforza *et al*'s (1994) finding that within India, linguistic differences account for much of the genetic diversity.

Based on mtDNA HVS1 sequence data, we find that the Austro-Asiatic tribals show a higher diversity than Dravidian tribals. This is consistent with Renfrew's (1992) observation that the present distribution of the Austric language group is due to the initial dispersal process out of Africa, whereas later agricultural dispersal can account for the Elamo-Dravidian or Sino-Tibetan (to which family Tibeto-Burman languages belong) distributions. Our observation is also consistent with the view of many scholars (Risley 1915; Rapson 1955; Thapar 1966; Pattanayak 1998) that the Austro-Asiatics in India may have been the original inhabitants. Indeed, if the Austro-Asiatic speaking tribals are the most ancient group of humans in India, they are expected to show the highest genetic diversity. Of course, it is possible that this group has descended from a group of founders after the Dravidian or the Tibeto-Burman speakers, but the founding group had a larger effective population size. Our data indicate that Austro-Asiatic speakers underwent a population expansion about 17000 years prior to the Elamo-Dravidian speakers, and about 5000 years prior to the Tibeto-Burman speakers. The confidence intervals of the expansion times of Austro-Asiatic and Tibeto-Burman speakers are non-overlapping with that of the Dravidian speakers, while those of Austro-Asiatic and Tibeto-Burman speakers are overlapping, indicating that the antiquity of expansion of the Austro-Asiatics is significantly greater than that of the Dravidians, but not of the Tibeto-Burmans. These data do not provide any evidence that the expansions took place within India. However, among hunter-gatherers, in particular, population expansions lead to enormous pressures on natural resources, which result in population movements. Therefore it is probable that the Austro-Asiatic speakers who expanded earlier also migrated earlier. Our data (Roychoudhury *et al* 2001) support the theory that different language groups in India represent distinct founding groups and that the Austro-Asiatic speakers are likely to have been the most ancient inhabitants of India.

6. Footprints from central and west Asia

From the 8 tribal populations described in the previous section, DNA samples were obtained from a total of 224 individuals. (Because of resource limitations, HVS1 sequencing reported in the previous section was performed only on a subset of 115 individuals.) Each DNA sample was screened for 10 mtDNA restriction site polymorphisms (RSPs) and 1 insertion/deletion polymorphism (IDP). The RSPs screened were HaeIII nt 663, HpaI nt 3592, AluI nt 5176, AluI nt 7025, DdeI nt 10394, AluI nt 10397, HinfI nt 12308, HincII nt 13259, AluI nt 13262, HaeIII nt 16517; and the IDP screened was the COII/$tRNA^{Lys}$ intergenic 9 bp deletion. These sites were chosen such that individuals could be classified into haplogroups that are most relevant for Indian populations (Roychoudhury *et al* 2001). With respect to the 11 biallelic mtDNA polymorphisms, 13 distinct haplotypes were observed among the 224 individuals screened from the 8 populations. None of the individuals possessed the COII/$tRNA^{Lys}$ intergenic 9 bp deletion. Two of the 10 restriction sites (HpaI nt 3592 and AluI nt 7025) were also monomorphic. It was found that only 4 of the 13 distinct haplotypes were present in many populations; the remaining 9 haplotypes occurred sporadically. The Tibeto-Burman speaking Tipperahs harboured the maximum number (12 out of 13) of haplotypes. In the pooled sample, only one haplotype was present in about 60% of the individuals. (These features have been noted in a more diverse set of populations, as reported in an earlier section.) Based on the RSP data, we were able to classify individuals into 8 haplogroups: Asian and Amerindian haplogroups A, B, C and D; east Asian haplogroup M; Caucasian haplogroups U and H; and, African haplogroup L. The frequencies of the various haplogroups are presented in table 3. Haplogroups B, H and L were not observed. Thirty seven (16·5%) individuals could not be classified into any of the 8 haplogroups based on the RSP

sites examined by us. Haplogroup M was found to be the most frequent – 71·4% of the individuals in the pooled sample belonged to this haplogroup. The frequency (51·11%) of this haplogroup was found to be significantly lower among Tibeto-Burman tribals compared to the Austro-Asiatic (76·27%) and the Dravidian (76·66%). Of the remaining haplogroups observed in the study populations, haplogroup U was also found to occur in most populations. This haplogroup is known to occur in high frequencies among Caucasian populations, including those of central and west Asia. The frequency of this haplogroup in our pooled tribal sample was about 10%. Considerable differences in the frequencies of this haplogroup were observed among Austro-Asiatic (13·56%), Dravidian (9·17%) and Tibeto-Burman (6·7%) tribals; these differences were, however, not statistically significant at the 5% level. Kivisild *et al* (1999) found that there are several subclusters of haplogroup U, of which they had found 6 to be present in their sample of Indians. We have found only 2 of these subclusters to be present among the tribals in India. These are subclusters U2i and subcluster U1, with frequencies 77·3% and 9·1%, respectively. Interestingly, we have found that all the 6 Irulas who belonged to haplogroup U also possessed transitions at nucleotide positions 16051, 16189, 16234 and 16247. This association was not found in any other tribal belonging to haplogroup U. It is remarkable that our estimate (77·3%) of the proportion of tribals belonging to Indian-specific subcluster U2i of haplogroup U coincided with that (77·9%) estimated earlier by Kivisild *et al* (1999) based on samples primarily from caste populations. Because the antiquities of the tribal populations are far greater than the time of entry (3000–4000 ybp) of Indo-Aryan speakers in India, our data support Kivisild *et al*'s (1999) conclusion that haplogroup U was introduced in India by an ancestral population that preceded the arrival of Indo-Aryan speakers into India. However, while Kivisild *et al* (1999) found several western-Eurasian mtDNA lineages belonging to haplogroup H and subcluster U1, K, U4, U5 with frequencies between 1%–5% in their samples from India, we found only the subcluster U1 in our tribal samples with a frequency of 9%. The subcluster U7, found at a frequency of about 13% in Kivisild *et al*'s (1999) samples but not found in our tribal samples, may also be western-Eurasian. Since the samples included in Kivisild *et al*'s (1999) study were obtained primarily from Indo-Aryan speaking caste populations, it is possible that these western-Eurasian specific haplogroups and subclusters, except U1, which are not found among the tribals in India may have been introduced in India with the entry of Aryan speakers from west and central Asia. This is contrary to Kivisild *et al*'s (1999) suggestion that all of the western-Eurasian subclusters of haplogroup U were present in India before the entry of the Aryan speakers. Thus, mtDNA provides signatures of population movements into India from central and west Asia.

Since many contend that immigrants from central and west Asia were predominantly males, we also sought to find similar signatures on Y-chromosomal DNA (Mukherjee *et al* 2001). Prehistoric, historic and linguistic evidences have suggested that Middle Eastern/west Asian and central Asian gene pools have contributed to the Indian gene pool. The northern exit route of humans from Africa to India was through the Middle East and west Asia. Subsequently, with the development of agriculture in the fertile crescent region that extends from Israel through northern Syria to western Iran, there was possibly migration of humans from this region into India. More recently, pastoral nomads originating in the central Asian

Table 3. Haplogroup frequencies among 8 tribal populations of India.

Population name	Geographical region of sampling	Haplogroup frequency* (%)					
		A	C†	D†	M	U	Other
Lodha	East				26 (81·3)	6 (18·7)	
Munda	East				5 (71·4)		2 (28·6)
Santal	East				14 (70·0)	2 (10·0)	4 (20·0)
Irula	South		1 (3·3)		16 (53·3)	7 (23·3)	7 (23·3)
Kota	South				29 (96·7)		1 (3·3)
Kurumba	South	1 (3·3)			23 (76·7)	2 (6·7)	4 (13·3)
Muria	Central				24 (80·0)	2 (6·7)	4 (13·3)
Tipperah	Northeast	4 (8·9)	1 (2·2)	4 (8·9)	23 (51·1)	3 (6·7)	15 (33·3)
Pooled		5 (2·2)	2 (0·9)	4 (1·8)	160 (71·4)	22 (9·8)	37 (16·5)

*Haplogroups B, H and L were not observed in the study samples.
†Haplogroups C and D are subsets of haplogroup M; therefore, individuals belonging to haplogroups C and D are also counted as belonging to haplogroup M.

steppes may also have contributed to the gene pool of India. The entry of humans from these regions into India was through the northwest corridor of India (Thapar 1975). We have, therefore, chosen to investigate gene pools of contemporary population groups inhabiting northern India, since traces of ancient admixture are likely to be more easily detected in northern India than elsewhere. We have studied four groups inhabiting the northern Indian state of Uttar Pradesh. The ethnic groups were: Brahmin (BRA), Chamar (CHA), Muslim (MUS) and Rajput (RAJ). The Brahmins, Rajputs and Chamars all belong to the Hindu caste fold and occupy upper, middle and lower ranks, respectively, in the caste hierarchy. The Muslim is an Islamic religious group. Most individuals belonging to this group are religious converts from various other populations that inhabited this geographical location. The study is based on Y-chromosomal polymorphisms; our inferences, therefore, reflect male population movements and admixture. We have collated data, mostly published and some unpublished, from several Middle Eastern population groups (Hammer 2000; Nebel 2000) and have performed comparative statistical analyses to draw inferences.

DNA samples were typed in respect of 12 binary polymorphic markers – YAP, 92r7, SRY 4064, sY81, SRY+465, TAT, M9, M13, M17, M20, SRY10831 and p12f2. Based on the UEP markers, we have classified Y chromosomes of each population into haplogroups (HGs) as defined by Rosser *et al* (2000). The haplogroup frequencies are presented in table 4. Because the set of markers screened in this study were not exactly the same as those screened in the published studies (Hammer *et al* 2000; Nebel *et al* 2000), some of the haplogroups could not be resolved and had to be pooled. It is seen that there is substantial overlap in the types of haplogroups observed in the north Indian and in the Middle Eastern regions.

The distributions of Y haplogroups, defined (Rosser *et al* 2000) on the basis of 12 biallelic UEPs reveal many interesting patterns. The haplogroup diversities in the populations of northern India and the Middle East are quite high, which is indicative of large long-term effective population sizes or high rates of gene flow from disparate populations or both. Among the north Indian populations, the differences in the frequency distributions of haplogroups are not statistically significant, but these differences among the Middle Eastern populations are significant. Although several haplogroups are common to the north Indian and Middle Eastern populations, the haplogroup frequency distributions in these two regions are substantially different. In northern India, HG-3 is the most frequent (35%–58%), while HG-9 is the most frequent

Table 4. Haplogroup frequencies (%) in 4 populations of north India and 8 populations of Middle East and west Asia.

	Haplogroup[a]							
Population	1	2	3	9	21	26	28	Other
North India								
Brahmin ($n = 17$)	11·8	23·5	35·3	23·5	0·0	0·0	5·9	0·0
Chamar ($n = 18$)	11·1	44·4	44·4	0	0·0	0·0	0·0	0·0
Muslim ($n = 19$)	15·8	0·0	57·9	10·5	0·0	15·8	0·0	0·0
Rajput ($n = 35$)	11·4	25·7	37·1	17·1	0·0	2·9	5·7	0·0
Total	13·2	23·1	41·7	14·3	0·0	4·4	3·3	0·0
Middle East and west Asia								
Israeli and Palestinian Arab ($n = 143$)	8·4	6·3	1·4	55·2	20·3	7·0	0	1·4[b]
Kurdish Jew ($n = 50$)	16·0[c]	4·0	4·0	44·0	8·0	24·0[d]	0·0	0·0
Yemenite Jew ($n = 30$)	27·0[c]	0·0	3·0	43·0	17·0	7·0[d]	0·0	3·0
Palestinian ($n = 73$)	9·0[c]	5·0	0·0	51·0	19·0	10·0[d]	0·0	5·0
Syrian ($n = 91$)	10·0[c]	3·0	9·0	57·0	10·0	11·0[d]	0·0	0·0
Lebanese ($n = 24$)	0·0	13·0	4·0	46·0	29·0	4·0[d]	0·0	4·0
Druze ($n = 21$)	5·0	0·0	0·0	38·0	19·0	38·0[d]	0·0	0·0
Saudi Arabian ($n = 21$)	5·0	5·0	19·0	33·0	5·0	29·0[d]	0·0	5·0
Total	10·2	4·8	4·0	50·5	16·1	12·4	0·0	2·0

[a]Haplogroup definitions are as given in Rosser *et al* (2000).
[b]All belong to Haplogroup 7.
[c]Includes Haplogroups 1 and 22.
[d]Includes Haplogroups 26 and 16.

(33%–57%) in Middle Eastern populations. Globally, the peak of HG-9 frequency is in the Caucasus-Anatolia region (Rosser *et al* 2000). This haplogroup is thought to have arisen about 5,500–17,400 ybp (Hammer *et al* 2000; Quintana-Murci *et al* 2001) in this region (south-western Iran). Our estimate (table 5) of the age of this haplogroup from data on the Middle Eastern populations is in fair agreement with this previous estimate. As noted in a previous study (Quintana-Murci *et al* 2001), this haplogroup may have been brought into India by Indo-European speakers from the Middle East. The frequency of this haplogroup is highest (23·5%) among the upper-ranked caste Brahmins and is lower (17·1%) among the middle-ranked caste Rajput. It is known that after the entry of the Aryan-speakers into India, the Brahmins were the torchbearers and promoters of Aryan rituals (Karve 1961). Therefore, it is likely that this group had the highest genetic contact with the Aryan-speaking peoples. This observation is consistent with the high frequency of HG-9 observed among them. This haplogroup may have percolated into the middle-ranked Rajputs, either through admixture with Brahmins or directly with the Aryan-speaking immigrants. It is noteworthy that HG-9 is absent among the low-ranked caste group, Chamar. A large section of the Muslims of Uttar Pradesh are known to be religious converts from both upper and middle-ranked caste groups. Our observation that HG-9 occurs in a lower frequency (10·5%) among the Muslim compared to the Brahmin and the Rajput is consistent with the known social history of this group.

Haplogroup-3, which is the most frequent haplogroup in India, is known to be widely found in Asia, except Eastern Asia, and is virtually absent in Africa and the Americas (Karafet *et al* 1999). HG-3 is found in high frequencies in central Asia (Russia and Altai region) and east Europe (Poland and Hungary). It appears that this haplogroup arose in central Asia about 7,500 ybp (Karafet *et al* 1999; Zerjal *et al* 1999; Rosser *et al* 2000), and the distribution of this haplogroup reflects a recent and major population expansion within Eurasia. HG-3 shows a decreasing frequency cline from central Asia westward into Europe (Rosser *et al* 2000) and from Iran towards India (Quintana-Murci *et al* 2001). Our data, however, are somewhat at variance with previous reports (Rosser *et al* 2000; Quintana-Murci *et al* 2001). We have found that the frequency of HG-3 in Uttar Pradesh is quite high (35%–58%). Although some published data from India are available (Zerjal *et al* 1999; Rosser *et al* 2000), the earlier reported frequencies of HG-3 from other parts of India are much lower. Our present data also indicate that the decreasing clinal pattern of HG-3 from Iran to India may not be as smooth as previously reported (Rosser *et al* 2000). However, our estimate of the age of this haplogroup (5,200 ybp) from the data on north Indian populations does not contradict the higher estimated age (Karafet *et al* 1999) of 7,500 ybp from data on central Asian populations.

Table 5. Estimated ages of common Y-chromosomal haplogroups found among northern Indians and among Israeli and Palestinian Arabs.

Population	Haplogroup	Estimated age (95% CI)
Northern Indians	1	7,200 (4,200–13,160)
	2	5,800 (3,400–10,700)
	3	5,200 (3,000–9,500)
	9	5,200 (3,000–9,500)
Israeli and Palestinian Arabs	1	6,000 (3,500–11,000)
	2	10,000 (5,800–18,400)
	9	8,300 (4,800–15,300)
	26	7,300 (4,250–13,400)

In addition to the data on HG-9 and HG-3 that provide clear indications of population movements from Iran and central Asia into India, some other haplogroups (HG-1 and HG-2) that are common in Europe (Rosser *et al* 2000) are also found in both north Indian and Middle Eastern populations, indicating genomic closeness of populations of these two regions to the Caucasoids. It is known that a major influx of people into India was of the central Asian tribal Huns, who at the end of the fifth century broke through into northern India (Thapar 1975; Kochhar 2000) after several aborted attempts. The Hun dominion extended from Persia right across to Khotan, the main capital being Bamiyan in Afghanistan. Together with the Huns came a number of other central Asian tribes and peoples, some of whom remained in northern India and others moved further to the south and the west. Some of these tribal people became the ancestors of the Rajput families (Thapar 1975). It is interesting that HG-21, which is thought to have originated in Africa about 31,000 ybp (Hammer *et al* 1998), is present in the Middle East in moderately high frequencies, but is completely absent in northern India (table 4). This haplogroup is on the YAP(+), presence of the Y *Alu* polymorphic element, background. We have previously noted (Bhattacharyya *et al* 1999) that the YAP element is absent in India. HG-21 is common (Rosser *et al* 2000) in many northern-African populations, and is largely confined to the southern region of Europe (Greek and Cyprus). Rosser *et al* (2000) contend that this haplogroup may reflect a barrier to gene flow between Africa and Europe.

We have also obtained another evidence, albeit somewhat indirect, of a possible admixture of northern Indian populations with central/west Asians. This evidence was obtained from a recent study on genomic polymorphisms with disease (Majumder and Dey 2001). Certain members

of the chemokine family of receptors serve as critical portals for the entry of HIV-1 into target cells. A mutant allele (Δ*ccr5*) of the β-chemokine receptor gene *CCR5* carrying a 32 base-pair deletion prevents cell invasion by the primary transmitting strain of HIV-1. The frequency of this mutant allele is known to be high among Caucasian populations, including populations of central and west Asia (Martinson *et al* 1997). We have screened about 1500 individuals, drawn from 40 diverse ethnic populations of India. We have found that the Δ*ccr5* allele is completely absent or only sporadically present in most populations. However, among the Muslims of Uttar Pradesh, the frequency of this allele was 5·36%, which may be due to admixture with immigrants from central and west Asia.

Acknowledgements

The studies described in this article were carried out with the active collaboration of scientists of various Indian institutions. Funding for these studies were obtained primarily from the Department of Biotechnology, New Delhi. I am grateful to my various collaborators, notably Susanta Roychoudhury, Nitai P Bhattacharyya, Bidyut Roy, A Ramesh, M V Usha Rani and Samir K Sil. I am also grateful to Sangita Roy, Namita Mukherjee, Badal Dey, Monami Roy, Sanat Banerjee, Madan Chakraborty, H Vishwanathan, N Prabhakaran and Analabha Basu, for help in data collection, DNA analyses, statistical analyses and other aspects of these studies. Ariella Oppenheim, Almut Nebel and Marina Faerman actively collaborated with me on an Indo-Israeli project on genomic continuity from west and central Asia to India.

References

Ballinger S W, Schurr T G, Torroni A, Gan Y Y, Hodge J A, Hassan K, Chen K-H and Wallace DC 1992 Southeast Asian mitochondrial DNA analysis reveals continuity of ancient mongoloid migrations; *Genetics* **130** 139–152

Beteille A 1998 The Indian heritage – a sociological perspective; in *The Indian human heritage* (eds) D Balasubramanian and N A Rao (Hyderabad: Universityies Press) pp 27–94

Bhattacharyya N, Basu P, Das M, Pramanik S, Banerjee R, Roy B, Roychoudhury S and Majumder P P 1999 Negligible gene-flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms; *Genome Res.* **9** 711–719

Bonatto S L and Salzano F M 1997 A single and early origin for the peopling of the Americas supported by mitochondrial DNA sequence data; *Proc. Natl. Acad. Sci. USA* **94** 1866–1871

Buxton L H D 1925 *The peoples of Asia* (London)

Cann R L 2001 Genetic clues to dispersal of human populations: Retracing the past from the present; *Science* **291** 1742–1748

Cann R L, Stoneking M and Wilson A C 1987 Mitochondrial DNA and human evolution; *Nature* (*London*) **325** 31–36

Cavalli-Sforza L L, Menozzi P and Piazza A 1994 *History and geography of human genes* (Princeton: Princeton University Press)

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 Chaina; *Proc. Natl. Acad. Sci. USA* **95** 11763–11768

Crow T J 1998 Was the speciation event on the Y chromosome?; *Abstracts of Contributions to the Dual Congress 1998* (Johannesburg: University of Witwatersrand Medical School) p. 109

Diamond J 1997 *Guns, germs and steel: The fates of human societies* (London: Jonathan Cape)

Fu Y-X 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection; *Genetics* **147** 915–925

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: Universities Press) pp 100–129

Guha B S 1935 The racial affinities of the people of India; in *Census of India, 1931, Part III – Ethnographical* (Simla: Government of India Press)

Hammer M F, Karafet T, Rasanayagam A, Wood E T, Altheide T K, Jenkins T, Griffiths R C, Templeton A R and Zegura S L 1998 Out of Africa and back again: nested cladistic analysis of human Y chromosome variation; *Mol. Biol. Evol.* **15** 427–441

Hammer M F, Redd A J, Wood E T, Bonner M R, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling M A, Jenkins T, Ostrer H and Bonne-Tamir B 2000 Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes; *Proc. Natl. Acad. Sci. USA* **97** 6769–6774

Harpending H C, Sherry S T, Rogers A R and Stoneking M 1993 The genetic structure of ancient human populations; *Curr. Anthrop.* **34** 483–496

Horai S and Matsunaga E 1986 Mitochondrial DNA polymorphism in Japanese: II. Analysis with restriction enzymes of four or five base pair recognition; *Hum. Genet.* **72** 105–117

Karafet T M, Zegura S L, Posukh O, Osipova L, Bergan A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Weibe V, Griffiths R C, Templeton A R and Hammer M F 1999 Asian source(s) of New World Y-chromosome founder haplotypes; *Am. J. Hum. Genet.* **64** 817–831

Karve I 1961 *Hindu society: An interepretation* (Poona: Deshmukh Prakashan)

Keith A 1936 Review of B S Guha's Racial Affinities of the Peoples of India; *Man* **29** 37

Kennedy K A R, Deraniyagala S U, Roertgen W J, Chiment J and Sisotell T 1987 Upper Pleistocene fossil hominids from Sri Lanka; *Am. J. Phys. Anthrop.* **72** 441–461

Kivisild T, Bamshad M J, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins W S, Dixon M E, Papiha S S, Mastana S S, Mir M R, Ferak V and Villems R 1999 Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages; *Curr. Biol.* **9** 1331–1334

Kochhar R 2000 *The vedic people* (Hyderabad: Orient Longman Ltd.)

Kosambi D D 1991 *The culture and civilisation of ancient India in historical outline* (New Delhi: Vikas Publishing House)

Majumder P P 1998 People of India: Biological diversity and affinities; *Evol. Anthrop.* **6** 100–110

Majumder P P and Dey B 2001 Absence of the HIV-1 protective Δ*ccr5* allele in most ethnic populations of India; *Eur. J. Hum. Genet.* (in press)

Majumder P P, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Guha Thakurta P 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

Martinson J, Chapman N H, Rees D C, Liu Y-T and Clegg J B 1997 Global distribution of the CCR5 gene 32-basepair deletion; *Nature Genet.* **16** 100–103

Meenakshi K 1995 Linguistics and the study of early Indian history; in *Recent perspectives of early Indian history* (ed.) R Thapar (Bombay: Popular Prakashan) pp 53–79

Misra V N 1992 Stone age in India: an ecological perspective; *Man Env.* **14** 17–64

Misra V N 2001 Prehistoric human coonization of India; *J. Biosci. (Suppl.)* **26** 491–531

Mountain J L, Hebert J M, Bhattacharyya S, Underhill P A, Ottolenghi C, Gadgil M and Cavalli-Sforza L L 1995 Demographic history of India and mtDNA sequence diversity; *Am. J. Hum. Genet.* **56** 979–992

Mukherjee N, Nebel A, Oppenheim A and Majumder P P 2001 High resolution mapping of Y-chromosomal polymorphisms reveals signatures of population movements from central and west Asia into India; *Hum. Genet.* (submitted)

Nebel A, Filon D, Weiss D A, Weale M, Faerman M, Oppenheim A and Thomas M G 2000 High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews; *Hum. Genet.* **107** 630–641

Parpola A 1975 On the protohistory of the Indian languages in the light of archaeological, linguistic and religious evidence: an attempt at integration; in *South Asian Archaeology* (ed.) J E van Lohuizen-De Leeuw (New York: Brill Academic Pub.) pp 73–84

Passarino G, Semino O, Modiano G, Santachiara-Benerecetti A S and Wallace D C 1993 COII/tRNA^{lys} intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities; *Am. J. Hum. Genet.* **53** 609–618

Pattanayak D P 1998 The language heritage of India; in *The Indian human heritage* (eds) D Balasubramanian and N A Rao (Hyderabad: Universities Press) pp 95–99

Quintana-Murci L, Krausz C, Zerjal T, Sayar S H, Hammer M F, Mehdi S Q, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling M A, Tyler-Smith C and McElreavey K 2001 Y-chromosome lineages trace diffusion of people and languages in southwestern Asia; *Am. J. Hum. Genet.* **68** 537–542

Rapson E J 1955 Peoples and languages; in *Cambridge history of India*, Vol. 1: *Ancient India* (ed.) E J Rapson (Delhi: S Chand and Co.)

Ratnagar S 1995 Archaeological perspectives on early Indian societies; in *Recent perspectives of early Indian history* (ed.) R Thapar (Bombay: Popular Prakashan) pp 1–52

Ray N 1973 *Nationalism in India* (Aligarh: Aligarh Muslim University)

Renfrew C 1987 *Archaeology and language: The puzzle of Indo-European origins* (London: Jonathan Cape)

Renfrew C 1992 Archaeology, genetics and linguistic diversity; *Man* **27** 445–487

Risley H H 1915 *The people of India* (Calcutta: Thacker Spink)

Rosser Z H, Zerjal T, Hurles M E, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley D G, Brede G, Cooper G, Corte-Real H B, de Knijff P, Decorte R, Dubrova Y E, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill E W, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko S A, Krumina A, Kucinskas V, Lavinha J, Livshits L A, Malaspina P, Maria S, McElreavey K, Meitinger T A, Mikelsaar A V, Mitchell R J, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis P C, Pereira L, Peterlin B, Pielberg G, Prata M J, Previdere C, Roewer L, Rootsi S, Rubinsztein D C, Saillard J, Santos F R, Stefanescu G, Sykes B C, Tolun A, Villems R, Tyler-Smith C and Jobling M A 2000 Y-chromosomal diversity in Europe is clinal and is influenced primarily by geography, rather than by language; *Am. J. Hum. Genet.* **67** 1526–1543

Rickards O 1995 Analysis of the region V mitochondrial marker in two Black communities of Ecuador and in their parental populations; *Hum. Evol.* **10** 5–16

Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakarn 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** 1182–1192

Roychoudhury S, Roy S, Basu S, Banerjee R, Vishwanathan H, Usha Rani M V, Sil S K, Mitra M and Majumder P P 2001 Genomic structures and population histories of linguistically distinct tribal groups of India; *Hum. Genet.* (in press)

Ruhlen M 1991 *A guide to the world's languages* (Stanford: Stanford University Press)

Sarkar S S 1958 Race and race movements in India; in *The Cultural Heritage of India* (ed.) S K Chatterjee (Calcutta: The Ramakrishna Mission Institute of Culture) vol. 1, pp 17–32

Sherry S T, Rogers A R, Harpending H, Soodyall H, Jenkins T and Stoneking M 1994 Mismatch distributions of mtDNA reveal recent human population expansions; *Hum. Biol.* **66** 761–775

Singh K S 1992 *People of India: An introduction* (Calcutta: Anthropological Survey of India)

Slatkin M and Hudson R R 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations; *Genetics* **129** 555–562

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, 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 1995 The first millennium B.C. in northern India; in *Recent perspectives of early Indian history* (ed.) R Thapar (Bombay: Popular Prakashan) pp 80–141

Thapar R 1966 *A history of India* Vol. 1 (Middlesex: Penguin)

Zerjal T, Pandya A, Santos F R, Adhikari R, Tarazona E, Kayser M, Evgrafov O, Singh L, Thangarj K, Destro-Bisol G, Thomas M G, Qamar R, Mehdi S Q, Rosser Z H, Hurles M E, Jobling M A and Tyler-Smith C 1999 The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe; in *Genomic diversity: Applications in human population genetics* (eds) S S Papiha, R Deka and R Chakraborty (New York: Plenum Press) pp 91–102