

# Pre-Caucasoid and Caucasoid Genetic Features of the Indian Population, Revealed by mtDNA Polymorphisms

Giuseppe Passarino,<sup>1</sup> Ornella Semino,<sup>2</sup> Luigi F. Bernini,<sup>3</sup> and A. Silvana Santachiara-Benerecetti<sup>2</sup>

<sup>1</sup>Dipartimento di Biologia Cellulare, Università della Calabria, Rende (CS); <sup>2</sup>Dipartimento di Genetica e Microbiologia "A. Buzzati Traverso," Pavia; and <sup>3</sup>Department of Human Genetics, State University of Leiden, Leiden

## Summary

About 70 individuals from Punjab were examined for some mtDNA polymorphisms, namely, the RFLPs of the six classical enzymes (*HpaI*, *BamHI*, *HaeII*, *MspI*, *AvaII*, and *HincII*) and for the sites *AluI*<sub>7,025</sub>, *DdeI*<sub>10,394</sub>, and *AluI*<sub>10,397</sub>. The *AluI*<sub>7,025</sub> polymorphic site was also investigated in 96 Indians from Uttar Pradesh and Andhra Pradesh and in 163 Mediterranean Caucasoids. Moreover, 30 Indian *DdeI*<sub>10,394*AluI*<sub>10,397</sub> (+ +) mtDNAs were typed by the "high-resolution restriction analysis" with 14 endonucleases to estimate their divergence time. The results obtained are the following: (1) The RFLPs analysis has displayed some Caucasoid types as in Indians of Uttar Pradesh; (2) the *AluI*<sub>7,025</sub> (–) allele, which defines the most frequent Caucasoid-specific lineage (haplogroup H), ranges from 18% to 45% in the Mediterranean Caucasoids, whereas it has shown low frequencies in Punjab (6.0%) and in Uttar Pradesh (1.8%) and was not found in Andhra Pradesh; (3) the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype, which although previously was considered an East Asian marker (haplogroup M) and was found very frequently in India, is also frequent in Punjab (27%); this frequency is, however, much lower than in Uttar Pradesh (49%) and in Andhra Pradesh (74%), and a gradient decreasing from south to north is therefore observed; (4) the divergence time of the Indian *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) mtDNAs has been estimated to be 30,250–60,500 years, a value that is compatible with that of the homologous East Asian lineage. These results strongly support the hypothesis that the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype predated the Indo-European invasion and probably the split between proto-Indians and proto-Orientals. Its frequency cline well reflects the major influence of Indo-Europeans in the north and in the center of India.</sub>

## Introduction

It is known that 3,000–4,000 years ago a population of nomadic shepherds, the Indo-European speakers, coming from the steppes of Central Asia, settled into the Indus valley, in northwestern India. Thereafter, they invaded the Indian subcontinent, moving southward and eastward and, taking advantage of their organization, took power and imposed their culture on the local populations (Cavalli-Sforza et al. 1994; Tartaglia et al. 1995). Much less is known about the origin of the preexisting inhabitants of India. According to Cavalli-Sforza (1988) and Renfrew (1989a, 1989b), the most important pre-Indo-European migration led Dravidian speakers to India, who moved from the southeastern lobe of the Fertile Crescent (Zagros Mountains, in the present Iran) ~10,000 years ago. Nowadays, Dravidian languages are confined mostly (but not exclusively) to the south (fig. 1), which indicates that the Indo-European invasion affected mainly the north and the center of the Indian subcontinent.

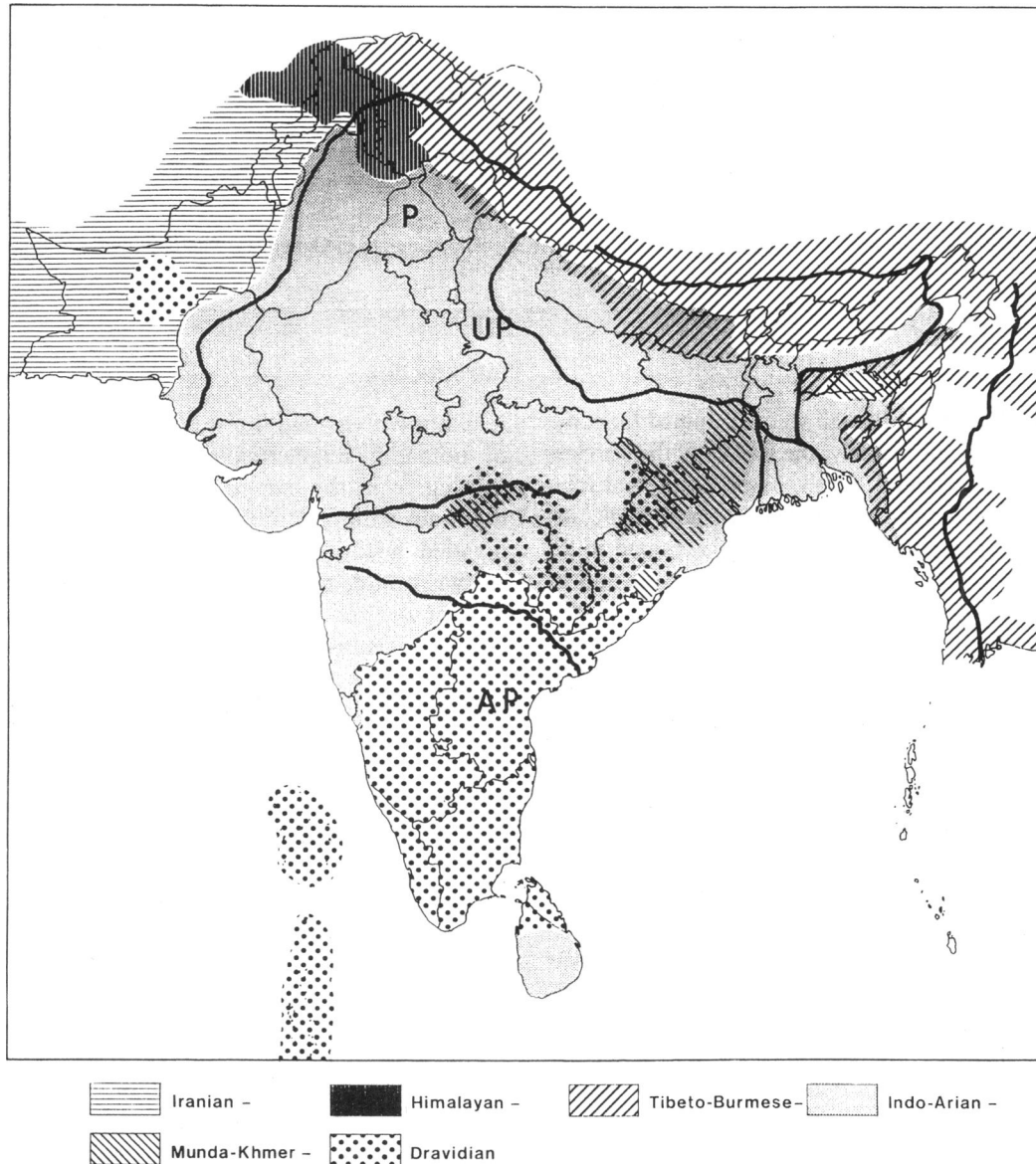
Genetic studies on the present-day Indian population highlighted Caucasoid features and showed that linguistic differences account for much of the genetic diversity. However, a considerable degree of admixture between Dravidians and Indo-Europeans has been observed, specifically in the north (Cavalli-Sforza et al. 1994).

Recently a survey of mtDNA RFLPs for the six classical enzymes (*HpaI*, *BamHI*, *HaeII*, *MspI*, *AvaII*, and *HincII*) in an Indian sample from New Delhi showed Caucasoid characteristics (types 6, 11, and 18) (Semino et al. 1991). By contrast, the analysis of the polymorphisms for the mtDNA sites *DdeI* at 10,394 bp and *AluI* at 10,397 bp unexpectedly displayed a very high frequency (49%; Passarino et al. 1996) of the simultaneous presence of these two restriction sites, the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype, which was considered a particular and very ancient characteristic of East Asians (Ballinger et al. 1992). It is indeed virtually absent in the other Caucasoids (Torroni et al. 1994a; Passarino et al. 1996), and it is present in peoples who moved from Asia very early (Australians, Papua New Guinea Highlanders, and Amerindians) (Ballinger et al. 1992; Chen et al. 1995). The extension of

Received April 3, 1996; accepted for publication July 1, 1996.

Address for correspondence and reprints: Dr. A. Silvana Santachiara-Benerecetti, Dipartimento di Genetica e Microbiologia "A. Buzzati Traverso," Via Abbiategrasso 207, 27100 Pavia, Italy. E-mail: santachiara@ipvgen.unipv.it

© 1996 by The American Society of Human Genetics. All rights reserved.  
0002-9297/96/5904-0025\$02.00



**Figure 1** Distribution of languages in the Indian subcontinent (modified from Istituto dell'Enciclopedia Italiana G. Treccani 1933, p. 46). P = Punjab; UP = Uttar Pradesh; AP = Andhra Pradesh; and black lines are rivers.

this analysis to a sample of Indians from southern India (Andhra Pradesh) confirmed the presence of the  $DdeI_{10,394}AluI_{10,397}$  (+ +) haplotype, which showed a frequency even higher (74%) than that observed in the New Delhi sample (Passarino et al. 1996). On the basis of this finding and of the observation that the  $DdeI_{10,394}AluI_{10,397}$  (+ +) haplotype was not carried by mtDNA molecules showing Caucasoid characteristics, we advanced the hypothesis that this marker was present in India before the Indo-European invasion (and, probably, also predated the split between proto-Indians and proto-Orientals [Passarino et al. 1996]). If so, a decreasing frequency cline of the  $DdeI_{10,394}AluI_{10,397}$  (+ +) haplotype from south to north, reflecting a different

degree of admixture between Indo-Europeans and preexisting Indian populations, could be expected. For this purpose, we analyzed a group from Punjab for the six classical enzymes RFLPs and the  $DdeI_{10,394}AluI_{10,397}$  haplotypes. That is a region situated in the northwest of India (fig. 1), whose inhabitants speak an Indo-European language (Punjabi) and show Caucasoid genetic features (Corbo et al. 1991; Tartaglia et al. 1995).

In addition, because the Caucasoid markers detectable in the RFLPs of the six classical enzymes have generally rather low frequencies (Johnson et al. 1983; Santachiara-Benerecetti et al. 1988; Semino et al. 1991; Brega et al. 1994), the Punjabi sample and the Indian groups

previously studied, together with control Caucasoid peoples from the Mediterranean area, were examined for a recently detected mtDNA marker. That marker is the absence of the *AluI* site at 7,025 bp, which defines the most frequent Caucasoid lineage (haplogroup H; 40% of U.S. Caucasoids) with an estimated divergence time of 23,000–45,000 years (Torroni et al. 1994a).

Further, a sample of *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) mtDNA molecules from individuals of different areas of India (Punjab, New Delhi, and Andhra Pradesh) was also typed by “high-resolution restriction analysis” with 14 endonucleases (Ballinger et al. 1992) in order to estimate their divergence time and compare it with that calculated for the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) lineage (haplogroup M) in East Asian populations (Chen et al. 1995).

## Material and Methods

### The Samples

The Punjabi sample consisted of 78 unrelated, apparently healthy individuals whose Punjabi ancestry was ascertained by interview. The sample collected in New Delhi ( $N = 56$ ) was mainly from Uttar Pradesh (Semino et al. 1991), where Indo-European languages (Hindi and Urdu) are spoken. The Andhra Pradesh sample ( $N = 40$ ) was collected in Kakinada. In this area a Dravidian language (Telugu) is spoken. Control Caucasoid groups consisted of 22 Italians, 50 Lebanese, 52 Jews, and 39 northern Africans.

### DNA Extraction

Five milliliters of blood from each individual were collected in EDTA, and DNA was extracted from buffy coats according to standard methods.

### Analysis of the Six Classical Enzymes' RFLPs

Restriction analysis, Southern blotting, and mapping were performed as described by Johnson et al. (1983). Restriction enzymes *HpaI*, *BamHI*, *HaeII*, *MspI*, *HincII* (Promega), and *AvaII* (Boehringer Mannheim) were used according to the suppliers' directions.

A pool of 11 partially overlapping fragments, obtained by PCR amplification and covering the entire mtDNA molecule, was used as a probe. (For primers and amplification conditions, see Passarino et al. [1993].) Labeling with <sup>32</sup>P-dCTP was performed as described by Feinberg and Vogelstein (1983). The new or rare morphs due to site gain were all further investigated through appropriate double digestions.

### *AluI*<sub>7,025</sub>, *DdeI*<sub>10,394</sub>, and *AluI*<sub>10,397</sub> Polymorphisms

*AluI*<sub>7,025</sub>, *DdeI*<sub>10,394</sub>, and *AluI*<sub>10,397</sub> sites were investigated by digestion of the relevant fragments after PCR

amplification with the appropriate enzymes. Electrophoretic analyses were carried out in 3.5% NuSieve GTG (FMC) Agarose gel in TBE.

### High-Resolution Restriction Analysis

The high-resolution analysis was carried out on the 11 above-mentioned amplified fragments covering the whole mtDNA molecule. Digestions were performed with *AluI*, *DdeI*, *HaeIII*, *HhaI*, *HinfI*, *HpaI*, *MspI*, *MboI*, *RsaI*, *TaqI*, *BamHI*, *HaeII*, *HincII* (Promega), and *AvaII* (Boehringer Mannheim) enzymes, used according to manufacturers' instructions. Restriction digests were electrophoresed in 2%–3.5% NuSieve GTG Agarose gel, in TBE.

### Statistics

Evolutionary relationships among haplotypes defined by high-resolution restriction analysis in the molecules sharing the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) marker, were inferred by PAUP 3.0 analysis (Swofford 1992). All dendrograms were rooted using the Senegalese haplotype Af71 (Chen et al. 1995). The Tree Bisection and Reconnection (TBR) algorithm was used to generate maximum parsimony trees. We stopped our search after 1,074 replications. Four trees were found to be the shortest, each of which was 51 steps long. The strict consensus tree of these four trees was then generated, and it turned out to be 52 steps long. The intragroup sequence divergence was estimated according to the method of Ballinger et al. (1992).

## Results

### RFLPs of the Six Classical Enzymes

In table 1, the Punjabi mtDNA types and their frequencies are reported, together with those of Indians from New Delhi (Semino et al. 1991). With the exception of a new *AvaII* morph (*AvaII* 37), all of the morphs had already been described. In the new *AvaII* electrophoretic pattern, the 9.8-kb band is split into two fragments of ~7.5 and 2.5 kb. Double digestion with *PstI* allowed the *AvaII* site to be located at 5,260 bp (G→A or T at 5,262 bp). This site had already been observed in a Hindu from New Delhi (Semino et al. 1991) where, however, it was associated with another *AvaII* site gain (at 15,882 or 15,887 bp), and therefore a differing morph (*AvaII* 29) was produced.

Unlike previously surveyed Indians (Semino et al. 1991), Punjabis do not show the *BamHI* 0, and *MspI* 4 morphs. As a consequence, types 18, 28, and 84 are absent. The Caucasoid types are here represented by types 6 and 11.

The very rare type 66 deserves some comments. It was previously observed only in three subjects, one

**Table 1****Frequencies (in percent) of the mtDNA Types in the Punjabi Population**

TYPES <sup>a</sup>	MORPH <sup>b,c</sup>	INDIANS	
		Punjab (N = 65)	New Delhi (N = 79)
1-2	(2.1.1.1.1.2)	72.3 ± 5.5	64.6 ± 5.4
1-7	(2.1.1.1.1.7)	1.5	2.5 ± 1.8
1-10	(2.1.1.1.1.10)	1.5	1.3
6-2	(2.1.2.1.1.2)	4.6 ± 2.6	1.3
8-1	(1.1.1.1.1.1)	1.5	1.3
9-1	(1.1.2.1.1.1)	1.5	—
11-2	(2.2.3.1.5°.2)	6.2 ± 3.0	2.5 ± 1.8
18-2	(2.3.1.4.9*.2)	—	2.5 ± 1.8
18-9	(2.3.1.4.9*.9)	—	1.3
21-2	(2.1.1.1.2.2)	4.6 ± 2.6	—
21-9	(2.1.1.1.2.9)	3.1 ± 2.1	1.3
23-2	(2.2.1.1.3°.2)	—	1.3
28-2	(2.1.1.4.1.2)	—	7.6 ± 3.0
47/55-2	(2.1.1.1.3.2)	—	1.3
66-2	(2.3.1.1.9*.2)	1.5	—
72-2	(2.1.1.1.12.2)	—	1.3
82-2	(2.1.1.1.16.1.2)	—	1.3
83-2	(2.1.1.1.29.2)	—	1.3
84-2	(2.0.1.1.30.2)	—	5.1 ± 2.4
85-2	(2.1.1.7.18.2)	—	1.3
208-2	(2.1.1.1.37.2)	1.5	—
F		.53	.42

NOTE.—Data from a Hindu sample collected in Delhi (Semino et al. 1991) are also given, for comparison. Standard errors, when appropriate, are given in italics.

<sup>a</sup> Since *HincII* data were not always included in the previous studies, the classification for this enzyme is indicated as a subtyping.

<sup>b</sup> The enzyme morphs are indicated in the order: *HpaI*, *BamHI*, *HaeII*, *MspI*, *AvaII*, and *HincII*.

<sup>c</sup> An \* or a ° indicates the loss of the *d* or *f* site respectively, due to a known specific nucleotide substitution (Santachiara-Benerecetti et al. 1988).

from Senegal (Scozzari et al. 1988), the others in the same isolated highland area of Sardinia (Sartoris et al. 1988). It harbors the mutation G→A at 13,368 bp, which simultaneously creates the Caucasoid *BamHI* 3 and *AvaII* 9 morphs (Santachiara-Benerecetti et al. 1988). Among Caucasoids, these two morphs are usually in molecules that are *MspI* 4 (Santachiara-Benerecetti et al. 1988), so it is likely that type 66 is caused by repeated mutations. This is also supported by the observation that the mtDNA with type 66 we found in Punjabis displays the non-Caucasoid *DdeI*<sub>10,394</sub>-*AluI*<sub>10,397</sub> (+ +) marker. Compared to the New Delhi sample, Punjabis appear less heterogeneous (*F* = .53 vs. .42).

#### *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> Haplotypes

Table 2 shows the distribution of the *DdeI*<sub>10,394</sub>-*AluI*<sub>10,397</sub> haplotypes among Punjabis. Data from East

Asian, Caucasoid, and previously studied Indian populations are shown for comparison. As for previous Indian data, important differences appear not only for the frequency of the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype (26.9% vs. 48.7% in New Delhi and 73.7% in Andhra Pradesh) but also for the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub>(+ -) association (12.8% vs. ~2% in the others), which has a greater incidence in Caucasoids (21.4%) (fig. 2).

#### *AluI*<sub>7,025</sub> Polymorphism

Table 3 shows the distribution of the *AluI*<sub>7,025</sub> polymorphism in Indians and in other Caucasoid populations. The very low frequency of the *AluI*<sub>7,025</sub> (-) allele in India is remarkable, compared with its incidence in the other groups.

#### High-Resolution Restriction Analysis

High-resolution restriction analysis (Ballinger et al. 1992) was carried out on 30 samples (9 from Punjab, 11 from New Delhi, and 10 from Andhra Pradesh), all being *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +). This analysis produced 23 different haplotypes, which are shown in table 4. All of the haplotypes were observed in one individual, with the exception of haplotypes 1 (five subjects), and 9, 17, and 19 (two subjects each). The strict consensus parsimony tree generated with these haplotypes is shown in figure 3.

The intragroup sequence divergence was 0.121%, corresponding to a radiation time of 41,724–55,000 or 30,250–60,500 years, when the evolution rates of 2.2%–2.9% (Torroni et al. 1994c) or 2%–4% (Cann et al. 1987; Torroni et al. 1993) per million years, respectively, were used.

#### Discussion

Classical genetic analyses clustered Indians with west Asians, but one could not deduce the origin of the first inhabitants of India (Cavalli Sforza et al. 1994; Mountain et al. 1995). Recently, studies on mtDNA shed new light on the peopling of India. The mtDNA D-loop sequences showed lineages probably predating the divergence of Eurasian populations (Mountain et al. 1995). In addition, an ancient “East” Asian-specific marker, the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype, which in East Asian lineages was estimated to be 55,000–73,000 years old (Chen et al. 1995), was found at a very high frequency among Indians (Passarino et al. 1996) who still displayed Caucasoid features in the mtDNA RFLPs (Semino et al. 1991).

The analysis of a sample from Punjab (northwest India) for the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> haplotypes and for the six classical enzymes RFLPs showed both Caucasoid features at RFLPs and a high frequency of the *DdeI*<sub>10,394</sub>-

**Table 2**

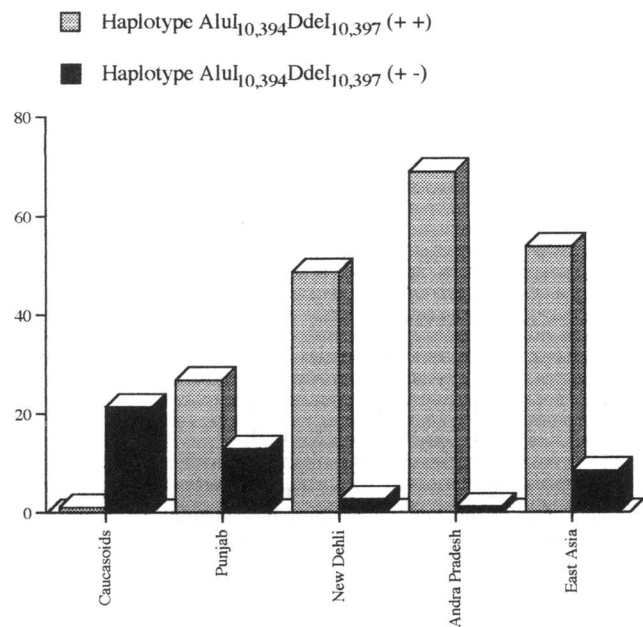
**Frequencies of the *DdeI*<sub>10,394</sub> *AluI*<sub>10,397</sub> Haplotypes in Punjab**

	No.	HAPLOTYPES			REFERENCES
		++ (%)	+ - (%)	-- (%)	
Caucasoids	383	1.0 ± .5	21.4 ± 2.0	77.6 ± 2.1	Torroni et al. (1994a) and Passarino et al. (1996)
<b>Punjabi</b>	<b>78</b>	<b>26.9 ± 5.0</b>	<b>12.8 ± 3.8</b>	<b>60.2 ± 3.5</b>	<b>Present study</b>
Indians from Delhi	76	48.7 ± 5.7	2.6 ± 1.8	48.7 ± 5.7	Passarino et al. (1996)
Indians Andhra Pradesh	57	73.7 ± 5.8	1.7	24.6 ± 5.7	Passarino et al. (1996)
Tribals Andhra Pradesh	30	60.0 ± 8.9	0	40.0	Passarino et al. (1996)
East Asians	153	44.4 ± 4.0	7.2 ± 2.1	48.4 ± 4.0	Ballinger et al. (1992)
Siberians	412	56.3 ± 2.4	10.0 ± 1.5	33.7 ± 2.3	Torroni et al. (1993)
Tibetans	54	61.1 ± 6.6	0	38.9	Torroni et al. (1994b)
Sub-Saharan Africans	197	1.5 ± .8	88.8 ± 2.2	9.6 ± 2.1	Chen et al. (1995) and Passarino et al. (1996)

NOTE.—Standard errors, when appropriate, are in italics. Relevant available data are also given, for comparison.

*AluI*<sub>10,397</sub> (+ +) haplotype. That frequency, however, was lower than in the previously studied Indian groups. A frequency cline decreasing from south to north (74% in Andhra Pradesh vs. 49% in New Delhi and 27% in Punjab) is observed. This cline can be explained by considering that

1. The *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype predates the arrival of Indo-Europeans.



**Figure 2** Distribution of the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) and (+ -) haplotypes in some Indian groups, in Caucasoids and in East Asians. Data from table 3; the two groups of Andhra Pradesh are pooled; “East Asia” comprises East Asians, Siberians, and Tibetans.

2. Indo-Europeans, who virtually were lacking this haplotype, entered India and spread from the northwest.
3. Indo-European migration affected the population structure mainly in the north and in the center of the Indian subcontinent, diluting the preexisting marker.

Further support is given by the independent complementary information of the frequency of the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ -) haplotype, which accounts for 21% of Caucasoids and decreases in the opposite direction (table 2 and fig. 2).

The divergence (0.121%) of *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) mtDNA molecules from individuals of different areas of India implies a radiation time of 41,724–55,000 or 30,250–60,500 years, when evolution rates of 2.2%–2.9% or 2%–4% per million years are used. These val-

**Table 3**

**Distribution of the *AluI*<sub>7,025(-)</sub> Allele Among Indian and Some Caucasoid Groups**

	No.	<i>AluI</i> <sub>7025(-)</sub> (%)
Indians		
Punjab	67	6.0 ± 2.9
New Delhi	56	1.8
Andhra Pradesh	40	0
Lebanese	50	18.0 ± 5.4
Jews	52	38.5 ± 6.7
North Africans	39	28.2 ± 7.2
Northern Italians	22	45.5 ± 10.6
U.S. Caucasoids <sup>a</sup>	175	40.0 ± 3.7

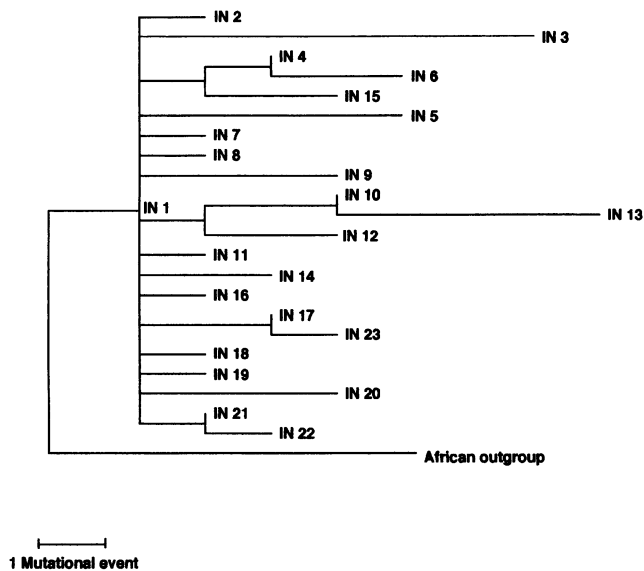
NOTE.—Standard errors, when appropriate, are in italics.  
<sup>a</sup> From Torroni et al. (1994a).

**Table 4**

**mtDNA Haplotypes Defined through High-Resolution Analysis of 30 Indian Samples Sharing the *Ddel*<sub>10,394</sub>*Alu*<sub>10,397</sub> (++) Haplotype**

	1	2	3	4	5	6	7	8	9	1	1	1	1	1	1	1	1	1	2	2	2	2	
										0	1	2	3	4	5	6	7	8	9	0	1	2	3
245 l	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
626 e	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3388c/3391a	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3534c/3537a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
5176 a	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6000 c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
6618 e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
6856 l	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6904 j	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7607 e	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
7859 j	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
8074 a	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8148 e	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8249b/8250e	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8309 c	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8391 e	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
8678 a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8882 c	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
9052n/9053f	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
9134 g	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
9156 e	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9470 c	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9524 a	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10364 e	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10394c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10397a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10431 j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10631 c	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11163 j	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11362 a	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
12185 e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
12212 a	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12507 b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
12560 a	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1
13332 e	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13366m/67b/68j	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14440 k	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15236 a	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15754 c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15925 i	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
15954 j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16085 c	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
16248 l	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
16303 k	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
16517 e	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1

NOTE.—“1” indicates the presence of a site, while “0” indicates the absence of one. The restriction enzymes used in the analysis are designated by the following single-letter code: a = *Alu*I; b = *Ava*II; c = *Ddel*; e = *Hae*III; f = *Hha*I; g = *Hinf*I; h = *Hpa*I; i = *Msp*I; j = *Mbo*I; k = *Rsa*I; l = *Taq*I; m = *Bam*HI; n = *Hae*II; and o = *Hinc*II. Sites separated by a slash indicate either simultaneous site gain or site loss for two different enzymes or a site gain for one enzyme and a site loss for another (in this case, the classification “0” or “1” refers to the first site) due to a single nucleotide substitution.



**Figure 3** Possible phylogeny of the Indian mtDNA molecules sharing the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype. This is the strict consensus tree generated from the four shortest trees, which were found by PAUP analysis after 1,074 replications by TBR algorithm (Swofford et al. 1992). Its length is 52 steps; the consistency index and retention index are 0.942 and 0.769, respectively. The “African outgroup” is the haplotype Af71 (Chen et al. 1995). The horizontal branch lengths are proportional to the number of mutational events that separate the haplotypes. “IN” stands for Indian.

ues are compatible with the divergence (0.161%) estimated in the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) East Asian lineage, which implies a divergence time of 55,517–73,180 or 40,250–80,500 (Chen et al. 1995). Thus, the hypothesis is strongly supported that the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype is an ancient marker of Indians and probably predates the split between proto-Indians and proto-East Asians (Passarino et al. 1996).

Moreover, if one considers the distribution of the *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) haplotype, the *AluI*<sub>7,025</sub> (–) allele, and the RFLPs mtDNA markers specific to Caucasoids (types 6, 11, and 18), some speculations are possible. Let us consider the ancient divergence time of both *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +) and *AluI*<sub>7,025</sub> (–) lineages and that “in the Paleolithic, when population densities were low, drift ... led to a mosaic genetic geography” (Cavalli Sforza et al. 1993, p. 639). It is therefore likely that, at the time of the major expansion leading to the peopling of Eurasia (Cavalli-Sforza et al. 1994; Mountain et al. 1995), Europe was colonized by group(s) predominantly *AluI*<sub>7,025</sub> (–), while India and East Asia were colonized by population(s) predominately *DdeI*<sub>10,394</sub>*AluI*<sub>10,397</sub> (+ +). On the other hand, the Caucasoid types 6, 11, and 18 would have been spread by subsequent migrations starting from the Middle East.

## Acknowledgments

This work has been partially supported by funds Ministero dell’Universita e della Ricerca Scientifica e Tecnologia (40% and 60%) and the Consiglio Nazionale delle Ricerche Progetto Finalizzato “Ingegneria Genetica” granted to A.S.S.-B.

## References

- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 130:139–152
- Brega A, Mura G, Cacciò S, Semino O, Brdicka R, Santachiara Benerecetti AS (1994) mtDNA polymorphism in a sample of Czechoslovaks and in two groups of Italians. *Gene Geogr* 8:45–54
- Cann R, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36
- Cavalli-Sforza LL (1988) The Basque population and ancient migrations in Europe. *Munibe Suppl* 6:129–137
- Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. *Science* 259:639–646
- (1994) *The history and geography of human genes*. Princeton University Press, Princeton
- Chen Y-S, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57:133–149
- Corbo RM, Tartaglia M, Scacchi R, Rickards O, Dash S, Sharma SK, Ramdeo IN, et al (1991) A survey of six genetic markers on the populations of Punjab and Rajasthan (India). *Gene Geogr* 5:113–121
- Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragment to high specific activity. *Anal Biochem* 132:6–13
- Istituto dell’Enciclopedia Italiana G. Treccani (1933) *Enciclopedia Italiana di lettere scienze ed arti*. Istituto dell’Enciclopedia Italiana G. Treccani, Rome
- Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL (1983) Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 19:255–271
- Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Otolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 56:979–992
- Passarino G, Semino O, Modiano G, Bernini LF, Santachiara-Benerecetti AS (1996) mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids: it likely predates the diversification between Indians and Orientals. *Ann Hum Biol* 23:121–126
- Passarino G, Semino O, Modiano G, Santachiara-Benerecetti AS (1993) COII/tRNA<sup>Lys</sup> 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
- Renfrew C (1989a) Before Babel: speculations on the origins of linguistic diversity. *Camb Archaeol J* 1:3–23

- (1989*b*) The origins of IndoEuropean languages. *Sci Am* 261:106–112
- Sartoris S, Varetto O, Migone N, Cappello N, Piazza A, Ferrara GB, Ceppellini R (1988) Mitochondrial DNA polymorphism in four Sardinian villages. *Ann Hum Genet* 52:327–340
- Santachiara-Benerecetti AS, Scozzari R, Semino O, Torroni A, Brega A, Wallace DC (1988) Mitochondrial DNA polymorphisms in Italy. II. Molecular analysis of new and rare morphs from Sardinia and Rome. *Ann Hum Genet* 52:39–56
- Scozzari R, Torroni A, Semino O, Sirugo G, Brega A, Santachiara-Benerecetti AS (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphism. *Am J Hum Genet* 43:534–544
- Semino O, Torroni A, Scozzari R, Brega A, Santachiara-Benerecetti AS (1991) Mitochondrial DNA polymorphisms among Hindus: a comparison with the Tharus of Nepal. *Ann Hum Genet* 55:123–136
- Swofford D (1992) Phylogenetic analysis using parsimony (PAUP). Version 3.0s. Illinois Natural History, Champaign
- Tartaglia M, Scacchi R, Corbo RM, Pompei F, Rickards O, Ciminelli BM, Sangatramani T, et al (1995) Genetic Heterogeneity among Hindus and their relationships with the other “Caucasoid” populations: new data on Punjab-Haryana and Rajasthan Indian States. *Am J Phys Anthropol* 98:257–273
- Torroni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC (1994*a*) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-Loop region. *Am J Hum Genet* 55:760–776
- Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994*b*) Mitochondrial DNA analysis in Tibet. Implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 93:189–199
- Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994*c*) A mitochondrial DNA “clock” for the Amerinds and its implication for timing their entry into North America. *Proc Natl Acad Sci USA* 91:1158–1162
- Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993) Mitochondrial DNA variation of aboriginal Siberians reveals distinct affinities with Native Americans. *Am J Hum Genet* 53:591–608