Am. J. Hum. Genet. 53:609-618, 1993

COII/tRNA^{Lys} Intergenic 9-bp Deletion and Other mtDNA Markers Clearly Reveal That the Tharus (Southern Nepal) Have Oriental Affinities

Giuseppe Passarino,* Ornella Semino,‡ Guido Modiano,§ and A. Silvana Santachiara-Benerecetti†,‡

*I. S. M. E. C., C. N. R., and †Dipartimento di Biologia Cellulare, Università della Calabria, Cosenza; ‡Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia; and §Dipartimento di Biologia, Università di "Tor Vergata," Rome

Summary

We searched for the East Asian mtDNA 9-bp deletion in the intergenic COII/tRNA^{Lys} region in a sample of 107 Tharus (50 from central Terai and 57 from eastern Terai), a population whose anthropological origin has yet to be completely clarified. The deletion, detected by electrophoresis of the PCR-amplified nt 7392–8628 mtDNA fragment after digestion with *Hae*III, was found in about 8% of both Tharu groups but was found in none of the 76 Hindus who were examined as a non-Oriental neighboring control population. A complete triplication of the 9-bp unit, the second case so far reported, was also observed in one eastern Tharu. All the mtDNAs with the deletion, and that with the triplication, were further characterized (by PCR amplification of the relevant mtDNA fragments and their digestion with the appropriate enzymes) to locate them in the Ballinger et al. phylogeny of East Asian mtDNA haplotypes. The deletion was found to be associated with four different haplotypes, two of which are reported for the first time. One of the deletions and especially the triplication could be best explained by the assumption of novel length-change events. Ballinger's classification of East Asian mtDNA haplotypes is mainly based on the phenotypes for the *Dde*I site at nt 10394 and the *Alu*I site at nt 10397. Analysis of the entire Tharu sample revealed that more than 70% of the Tharus have both sites, the association of which has been suggested as an ancient East Asian peculiarity. These results conclusively indicate that the Tharus have a predominantly maternal Oriental ancestry. Moreover, they show at least one and perhaps two further distinct length mutations, and this suggests that the examined region is a hot spot of rearrangements.

Introduction

The Tharu population lives mainly in Terai (southern Nepal), but some also live in Uttar Pradesh (India). Nepalese Tharus are concentrated in three groups in western, central, and eastern Terai (fig. 1). Until the 1960s, they were the only large permanent population to inhabit this previously heavily malaric area. Thereafter, malaria was almost completely eradicated. Since then, a massive immigration of other groups has taken place, so that the Tharus—about one-half million people—constitute only a minority in the present, highly heterogeneous population of Terai (Terrenato et al. 1988). On the basis of their physical features (Bista 1980) and of immunoglobulin markers (Chopra 1970), the Tharus are considered to be of East Asian extraction, although, when compared with other Oriental populations they are generally darker (Bista 1980). According to several popular stories, however, they are said to be largely derived from women who fled to Terai from Rajasthan —a Hindu area—at the time of the Islamic invasions (the 10th–12th centuries).

Studies on a number of mtDNA markers (Brega et al. 1986; Passarino et al. 1992) provided further support to the hypothesis of the Oriental origin of the Tharus. On the other hand, these studies, together with those on the distribution of malaria-related genes (Modiano et al. 1991; G. Modiano, unpublished data), showed a considerable degree of heterogeneity between western-central and eastern Tharus.

Received January 27, 1993; revision received April 22, 1993.

Address for correspondence and reprints: A. S. Santachiara-Benerecetti, Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Abbiategrasso 207, 27100 Pavia, Italy.

0002-9297/93/5303-0004$02.00

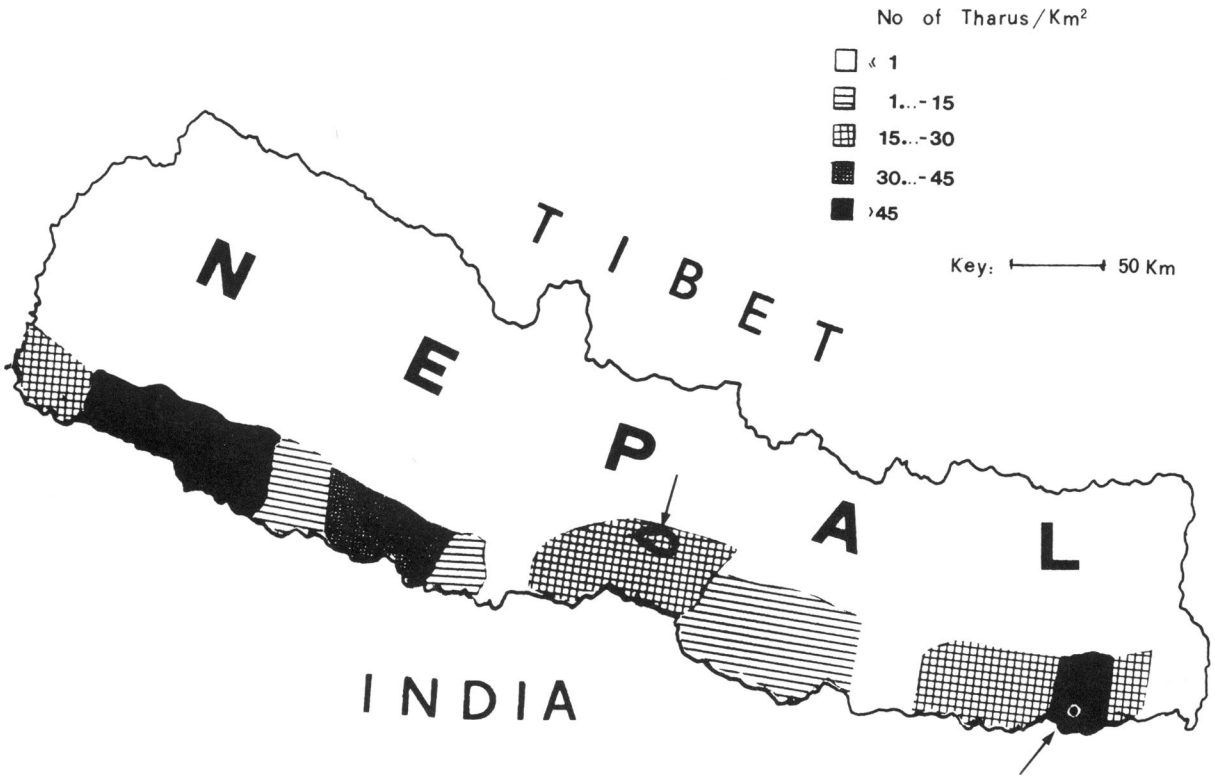

Figure 1 Map of Nepal showing the Terai areas inhabited by the Tharus. Arrows indicate the Chitwan (central Terai) and Morang (eastern Terai) districts where samples were collected.

An important marker to detect Oriental affinities is the 9-bp deletion of one of the two tandem repeats (CCCCCTCTA) between nt 8272 and nt 8289 in the COII/tRNA^{Lys} intergenic region of human mtDNA (Cann and Wilson 1983; Wrischnik et al. 1987). This marker was found in Polynesians (Hertzberg et al. 1989), Native Americans (Torroni et al. 1992), and Pacific coastal peoples (Ballinger et al. 1992), and it is

Table 1

Oligonucleotide Primers and Conditions Used for mtDNA Amplifications

Primers (5′→3′)[a]		Amplification Conditions		
Forward	Reverse	Denaturation	Annealing	Polymerization
534→553	1696→1677	93°C for 30 s	57°C for 60 s	72°C for 60 s
3007→3023	4508→4489	93°C for 30 s	55°C for 60 s	72°C for 60 s
3951→3970	5917→5898	93°C for 30 s	61°C for 60 s	72°C for 60 s
5720→5738	7608→7588	93°C for 30 s	57°C for 60 s	72°C for 60 s
7392→7410	8628→8608	93°C for 30 s	57°C for 60 s	72°C for 60 s
8282→8305	10107→10088	93°C for 30 s	57°C for 60 s	72°C for 60 s
9911→9932	11873→11851	93°C for 60 s	69°C for 60 s	72°C for 60 s
11711→11727	14208→14190	93°C for 60 s	49°C for 60 s	72°C for 120 s
13914→13930	15865→15845	96°C for 40 s	47°C for 60 s	72°C for 60 s
15553→15569	725→706	96°C for 40 s	45°C for 60 s	72°C for 60 s

[a] The primer coordinates are given according to the Anderson et al. (1981) sequence.

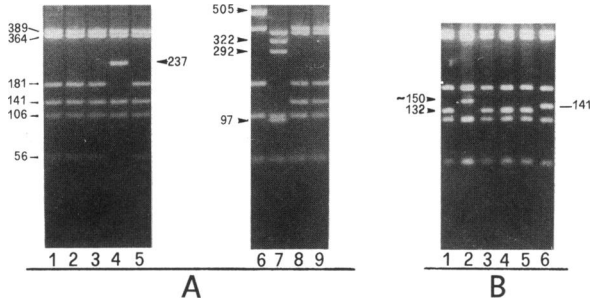

Figure 2 Electrophoresis of the amplified nt 7392–8628 fragment digested with *Hae*III. Fragment sizes are in bp. Arrows indicate variant fragments. *A*, Lanes 4, 6, and 7, Variant patterns due to single-base changes. Lane 4, Site loss at 8572. Lane 6, Site loss at 8250. Lane 7, Site loss at 8391 and site gain at 7792 (or 7789). Lanes 1–3, 5, 8, and 9, Normal patterns. *B*, Lanes 1–5, Variant patterns due to length variations. Lane 6, Normal pattern (141-bp band). Lanes 1, 3, and 4, Samples with the 9-bp deletion (132-bp band). Lane 5, Control sample with the deletion. Lane 2, Sample with the 9-bp triplication (150-bp band).

considered to be an indicator of Asian affinities in these populations. Furthermore, the occurrence of the 9-bp deletion in many distinct mtDNA haplotypes has been exploited to reconstruct the possible radiations of various groups within the Oriental major race (Horai and Matsunaga 1986; Cann et al. 1987; Stoneking et al. 1990; Ballinger et al. 1992; Harihara et al. 1992; Torroni et al. 1992). A detailed analysis of Southeast Asian mtDNA haplotypes interrelated in a maximum parsimony tree allowed 20 distinct groups to be identified, which, on the basis of the polymorphisms of *Dde*I at nt 10394 and *Alu*I at nt 10397, cluster into two major branches (Ballinger et al. 1992). This analysis also indicated that southern China is the center of Asian mtDNA radiation.

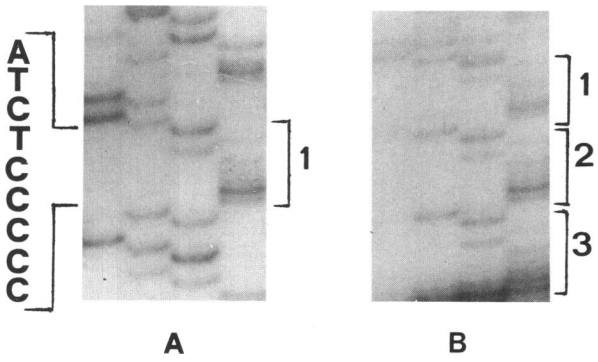

Figure 3 Sequencing analysis. *A*, Deletion. *B*, Triplication.

Table 2

Distribution of the Variants Found in the Tharus and Hindus, after *Hae*III Digestion of the nt 7392–8628 mtDNA Fragment

Variants	Status	No. of Tharus[a] (%)	No. of Hindus[b] (%)
Sites:			
8250	−	3 (2.8)	3 (3.9)
8572	−	2 (1.9)	1 (1.3)
7792	+		
8391	−		1 (1.3)
No. of 9-bp repeats:			
1		8 (7.5)	
3		1 (.9)	

[a] $N = 107$.
[b] $N = 76$.

We carried out a search for the COII/tRNA^{Lys} 9-bp deletion in the Tharus to verify their Oriental affinities, to further investigate heterogeneity between central and eastern Tharus, and to determine, through a more detailed analysis of the mtDNAs with the 9-bp deletion (if any), how these molecules are phylogenetically related. A sample of Hindus were also studied for comparative purposes.

Material and Methods

The sample consisted of 183 individuals, of which 107 were Tharus (57 from eastern Terai, Morang district, and 50 from central Terai, Chitwan district) (fig. 1), and 76 were Hindus (30 from Chitwan and 46 from New Delhi); the latter were taken as a non-Oriental neighboring control population. The absence of close relationships between these individuals was ascertained through interview data.

PCR Analysis

All samples were analyzed by PCR amplification and *Hae*III digestion of the fragment between nt 7392 and nt 8628 of the mtDNA sequence (Anderson et al. 1981). mtDNA of the nine samples showing length mutations was amplified by PCR in 11 overlapping fragments and digested with the enzymes (tables 3–5) necessary for a detailed haplotype characterization according to Ballinger et al. (1992). In addition, all the Tharu samples were examined for the sites *Dde*I at nt 10394 and *Alu*I at nt 10397 by digesting the amplified nt 9911–11853 mtDNA fragment. Primers and amplification

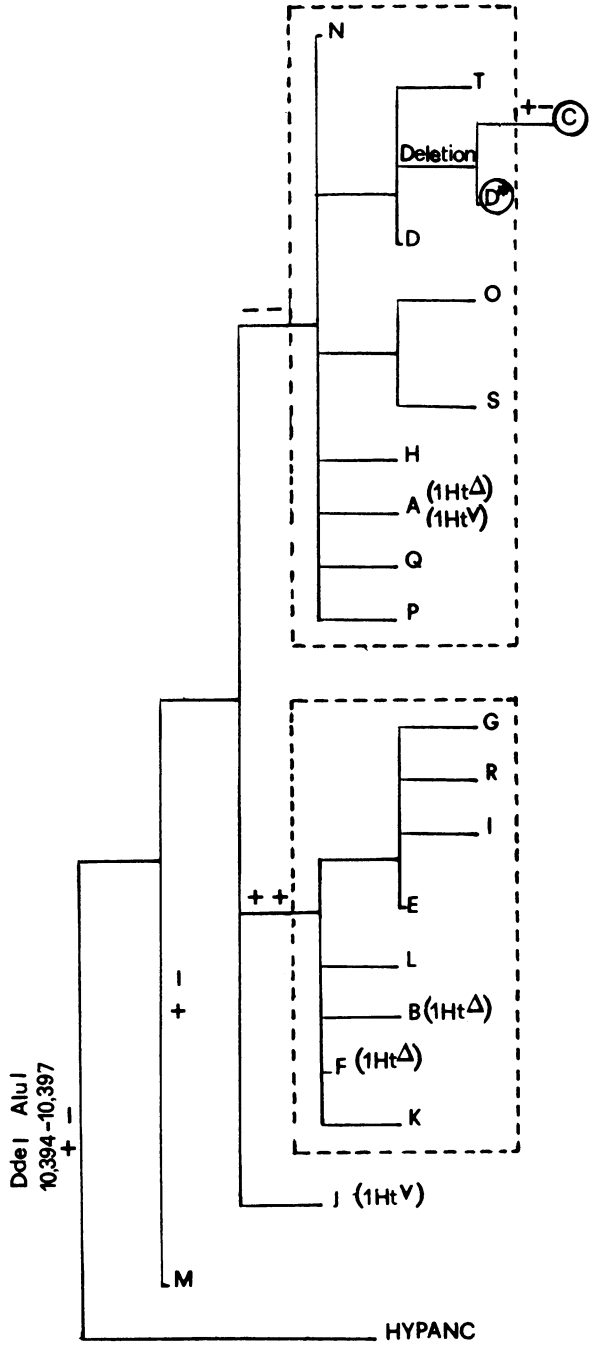

Figure 4 Simplified illustration of Ballinger's phylogeny of Southeast Asian mtDNA haplotype groups. The main classification based on the $DdeI_{10394}$, $AluI_{10397}$ phenotypes is indicated, whereas the various haplotypes that belong to each group with their additional mutations are not reported. Circles indicate those groups (C and D*) that comprise only haplotypes with the 9-bp deletion. Branch length is not meant to be proportional to numbers of mutations. Each of the groups A, B, and F include, inter alia, one haplotype with the deletion

conditions for each fragment are given in table 1. Reactions were performed in a final volume of 100 µl containing 10 mM Tris-HCl (pH 8.3, at 25°C), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, 0.2 mM of each dNTP mix, 0.3 µM of each appropriate oligonucleotide primer, 2.5 U of *Taq* DNA polymerase (Promega), and about 250 ng of genomic DNA. The reaction mixtures were overlaid with 1–2 drops of mineral oil (Sigma). At the end of each amplification (35 rounds), the incubation was prolonged for 10 min at 72°C. Amplifications were verified by electrophoresis of 5 µl of each product on 0.85% agarose. Ten microliters of each amplified fragment was digested for 3 h with 4 U of the appropriate restriction enzymes. DNA digests were electrophoresed in 2.5% NuSieve GTG + 0.9% SeaKem ME agarose gel (FMC Bio-Products) containing 1 µg ethidium bromide/ml. Restriction patterns were determined with UV fluorescence.

Sequencing Analysis

Sequencing analysis was carried out with the *f*-mol Sequencing Kit (Promega) by using the fragment nt 7392–8628 (amplified as described above) as template. The primer, CTACGGTCAATGCTCTG, which corresponds to nt 8161–8177 of the mtDNA sequence (Anderson et al. 1981), was radiolabeled by incubating a mix containing 10 pmol of the primer, T4 Kinase (5 U) with the supplied buffer and [γ^{32}P]ATP (10 pmol) for 30 min at 37°C in an Eppendorf tube. The reaction was stopped by heating the mix for 2 min at 90°C. Three picomoles of the 5′-labeled primer was added to a tube that contained the template DNA (about 100 ng), and sequencing buffer was added to a final volume of 16 µl. This mix was subdivided into four tubes, each containing all four deoxynucleotides (16 µM each) and one dideoxinucleotide (about 200 µM). To the tubes were then added 1.25 U of *Taq* polymerase sequencing grade and a few drops of mineral oil. The tubes were transferred into a Perkin Elmer Cetus cycler for the amplification. The program consisted of 30 rounds, with the following three steps each: 95°C for 30 s (denaturation), 55°C for 30 s (annealing), and 70°C for 60 s (polymerization). The reaction was stopped by the ad-

(Ht$^{\Delta}$). While for the Ht$^{\Delta}$ A the deletion may have a common origin with other Hts$^{\Delta}$, two new deletion events were postulated for the Hts$^{\Delta}$ B and F because they are separated by numerous mutations (not reported in this scheme). Ht^{V} indicates a haplotype with a 4-bp insertion in the same region as that in which the two tandem repeats occurred. HYPANC = Hypothetical ancestor.

Table 3

The Haplotypes of the Five $DdeI_{10394}$ $AluI_{10397}$ (+−) Samples with the Deletion, Belonging to the C Group of Ballinger's Classification

Haplotype[a]	N	*Mbo*I 3090	*Dde*I/ ***Alu*I** 3534/37	*Hae*III 6957	***Dde*I** 10394	***Alu*I** 10397	*Dde*I 10746	*Dde*I 11146	*Xho*I 14157	*Taq*I 14168	*Hin*fI/ *Mbo*I 15234/35	*Alu*I 16254	*Taq*I 16512
55	1	−	−/−	−	+	−	−	−	−	−	−/+	−	−
60	2	−	−/−	+	+	−	−	+	−	−	−/+	−	−
$1^{Tharu\ b}$	2	−	−/−	+	+	−	+	+	−	−	−/+	−	−

NOTE.—In **boldface** are the features which, when added to the deletion, characterize the C group. Slashes indicate two phenotypes produced by a single event.

[a] To classify these haplotypes according to Ballinger et al. (1992), all the sites for which the authors found variation within the C group were examined.

[b] This new type can be derived from type 60 by a single *Dde*I site gain at nt 10746.

dition of 3 μl of a stop solution that contained 10 mM NaOH, 95% formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol. The samples were denaturated at 70°C for 2 min and were loaded in a 0.4-mm thick, 6% polyacrylamide sequencing gel. After electrophoresis the gel was exposed for autoradiography without intensifying screens for 12–24 h.

Nomenclature

Since, on several occasions, different mtDNA morphs and types were given the same (numerical) symbols, we adopted a compound symbol (i.e., 1^{Tharu}), as for hemoglobin variants. The phenotypes for the *Dde*I site at nt 10394 and the *Alu*I site at nt 10397 are indicated by the restriction enzyme symbol, with the site as a subscript and with, in parentheses, a plus (+) for site presence and a minus (−) for site absence.

Results

Figure 2 shows the electrophoretic patterns for some amplified nt 7392–8628 fragments, after *Hae*III digestion. We observed, in addition to the most common pattern, five variant phenotypes: three were due to single-base substitutions (fig. 2*A*), and two were caused by length variations (fig. 2*B*).

In figure 2*A*, lane 4, fragments of 181 and 56 bp are missing because of the site loss at nt 8572, which produces a new 237-bp band. The pattern in lane 6 is determined by a site loss at nt 8250, which causes the fusion of the two 141- and 364-bp contiguous fragments into a new 505-bp fragment. Both of these variants have been described elsewhere (Cann et al. 1987; Stoneking et al. 1990; Ballinger et al. 1992).

The pattern shown in lane 7 is due to the combination of two different changes: a site loss at nt 8391, which produces a new 322-bp band (fusion of the 141- and 181-bp fragments), and a site gain, which splits the 389-bp fragment into two new ones of about 290 and 100 bp. Of the two latent sites, one at nt 7789 and the other at nt 7792, which are compatible with the observed pattern, the latter is the more likely because its variation does not change the amino acid sequence. Though both the mutations yielding this pattern have also been described, the pattern is new because they were observed separately, in two distinct mtDNAs (Cann et al. 1987; Ballinger et al. 1992).

Figure 2*B* shows variant patterns for length polymorphisms. The 141-bp band (lane 6), which harbors the normally sized COII/$tRNA^{Lys}$ region, is replaced in lanes 1, 3, and 4 by a faster one (132 bp), the position of which corresponds in all cases to that (lane 5) of a control sample with the COII/$tRNA^{Lys}$ Oriental deletion. In lane 2, the 141-bp segment is replaced by a slower band of 150 bp.

Sequencing of the fragment of interest in two samples with the faster band (fig. 3*A*) confirmed in both cases the classical 9-bp repeat deletion (Cann and Wilson 1983; Wrischnik et al. 1987). Sequencing of the sample with the slower band showed the insertion of a third copy of the 9-bp repeat (fig. 3*B*). The frequency of all these variants is reported in table 2.

To better characterize the eight mtDNA molecules with the deletion, we analyzed them for the most informative RFLP sites of the Ballinger et al. (1992) classifica-

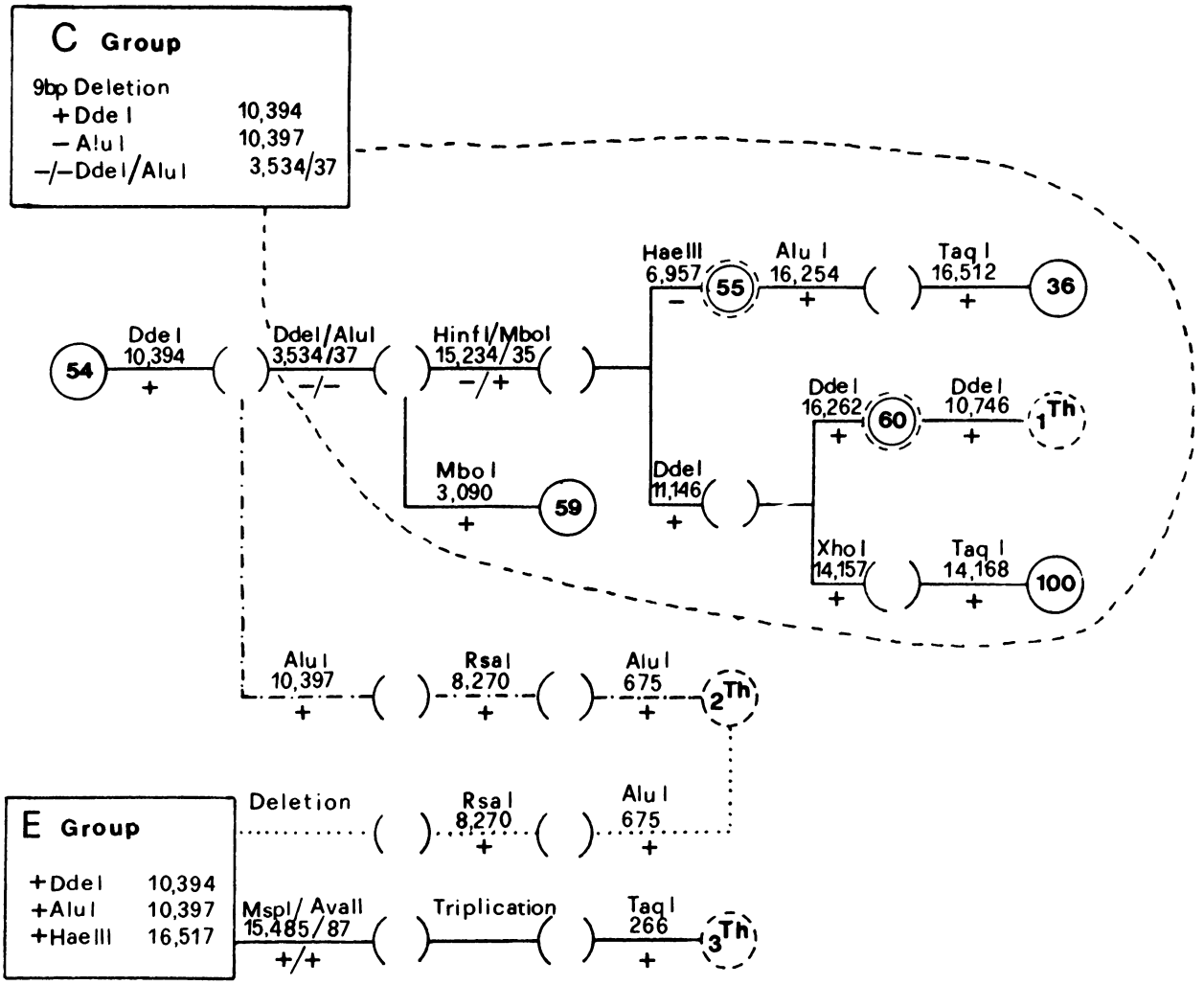

Figure 5 Relationship of the Tharu length-variation haplotypes with Ballinger's mtDNA phylogeny, in which the mutational steps are given in detail. Only the relevant part of this phylogeny is illustrated. Squares enclose C and E group characteristics. Dashed lines enclose haplotypes of the C group. ( ) = undiscovered intermediate haplotypes; ○ = Ballinger haplotypes; ◎ = Ballinger's haplotypes found in the present study; and ◌ = haplotypes found only in the present study. The order of the intermediates required for types 2^{Th} and 3^{Th} is arbitrary, with the exception of the first intermediate in the 3^{Th} pathway. The *Msp*I/*Ava*II mutation at nt 15487 was found in about 20% of the sample, while triplication and the *Taq*I site at nt 266 were present in only one subject. Haplotypes 60 (2 subjects) and 1^{Th} (2 subjects) were found in central Tharus; haplotypes 55 (1 subject) and 2^{Th} (3 subjects) were found in eastern Tharus.

tion (fig. 4). On the basis of the two main sites (*Dde*I at nt 10394 and *Alu*I at nt 10397) of this classification, five samples (four in central Tharus and one in eastern Tharus) were $DdeI_{10394}$, $AluI_{10397}$ (+ −). Thus they showed the basic features of the C group. The remaining three samples (all from eastern Tharus) were $DdeI_{10394}$, $AluI_{10397}$ (+ +) and were of an unclear derivation.

We obtained a more detailed haplotype classification within the C group of the five $DdeI_{10394}$, $AluI_{10397}$ (+ −) samples by analysis with the appropriate enzymes (table 3). One of these samples was classifiable as haplotype 55, two were classifiable as haplotype 60, and the remaining two showed a new type, which we named haplotype 1^{Tharu} and which could be considered as a single-step derivative from haplotype 60 (fig. 5 and table 3).

We characterized the three $DdeI_{10394}$, $AluI_{10397}$ (+ +) samples by analysis with 15 restriction enzymes (table 4). All samples exhibited a new haplotype which we

Table 4

The Haplotype (2^{Tharu}) of the Three $DdeI_{10,394}$, $AluI_{10,397}$ (++) Samples with the Deletion

Haplotype	No. of Repeats	*DelI/AluI*[a] 3534/37	*AluI* 675	*RsaI* 8270	*DdeI* 10394	*AluI* 10397	*HaeIII* 16517
2^{Tharu}	1	+/+	+	+	+	+	+
C group ...	1	−/−	−	−	+	−	+
E group ...	2	+/+	−	−	+	+	+

NOTE.—For comparison, the relevant features of the C and E groups of Ballinger's classification are also reported. The haplotype 2^{Tharu} and features characterizing each group are in **boldface**. The samples were analyzed with the following 15 enzymes: *AluI*, *AvaII*, *BamHI*, *DdeI*, *HaeII*, *HaeIII*, *HhaI*, *HincII*, *HinfI*, *HpaI*, *HpaII*, *PstI*, *RsaI*, *TaqI*, and *XbaI*. Only the variant sites have been reported.

[a] Slashes indicate two phenotypes produced by a single event.

called haplotype 2^{Tharu}. This haplotype differed greatly from the only two haplotypes of Ballinger's phylogeny, which carried both the $DdeI_{10394}$, $AluI_{10397}$ sites, together with the deletion (fig. 4, Ht^{Δ} in B and F). These haplotypes, in turn, differed to the extent that each was interpreted as the result of a repeated deletion event (Ballinger et al. 1992).

As shown in table 4 and figure 5, the haplotype 2^{Tharu} can be derived either from a precursor of the C group or from the E group. In the second case, a new deletion event is required.

The haplotype with the triplication, which we called haplotype 3^{Tharu}, very likely derives from the E group through two intermediate types (table 5 and fig. 5).

Recently, it was suggested that, while the *DdeI* polymorphism at nt 10394 has been found in every major ethnic group (Cann et al. 1987; Brown et al. 1992), the combination of the two sites $DdeI_{10394}$, $AluI_{10397}$ (+ +) represents a very ancient Asian marker that is particularly frequent in Southeast Asian populations (Ballinger et al. 1992; Wallace and Torroni 1992). We therefore extended the analysis of these sites to the whole Tharu sample. Digestion of the mtDNA (nt 9911–11873)-amplified fragment with the above-mentioned enzymes split the Tharus into two ($DdeI_{10394}$, $AluI_{10397}$ [+ +] and [− −]) major phenotype clusters, as occurs with the other Asian populations. As shown in table 6, the $DdeI_{10394}$, $AluI_{10397}$ (+ +) phenotype, which represents about 50% of the previously studied East Asian populations (Ballinger et al. 1992), accounts for more than 70% of both the central and eastern Tharus. Very few $DdeI_{10394}$, $AluI_{10397}$ (+ −) mtDNAs were found in any of these groups (table 6), and most of them (all in our sample) carry the deletion.

Discussion

We searched for the deletion of one of the two 9-bp repeats, between nt 8272 and 8289 in the COII/$tRNA^{Lys}$ region of human mtDNA (Cann and Wilson 1983), in the Tharus of Terai. We used this valuable Oriental marker (Wrischnik et al. 1987) to support the commonly accepted, but still debated, hypothesis of this population's Oriental extraction. The marker's oc-

Table 5

The Haplotype (3^{Tharu}) of the Sample with Triplication

Type	No. of Repeats	*TaqI* 266	***DdeI*** **10394**	***AluI*** **10397**	*MspI/AvaII*[a] 15485/487	***HaeIII*** **16517**
3^{Tharu}	3	+	+	+	+/+	+

NOTE.—The features characterizing the E group (Ballinger et al. 1992) are in **boldface**. The sample was analyzed with the following 15 enzymes: *AluI*, *AvaII*, *BamHI*, *DdeI*, *HaeII*, *HaeIII*, *HhaI*, *HincII*, *HinfI*, *HpaI*, *HpaII*, *PstI*, *RsaI*, *TaqI*, and *XbaI*. Only the variant sites have been reported.

[a] Slashes indicate two phenotypes produced by a single event.

Table 6

Distribution of *DdeI*$_{10394}$, *AluI*$_{10397}$ Phenotypes in the Tharus

	No. (%) in			
Phenotype	Central Tharus[a]	Eastern Tharus[b]	Whole Sample[c]	East Asians[d]
++	37 (74.0)	41[e] (73.2)	78 (73.6)	69 (45.1)
+−[f]	4 (8.0)	1 (1.8)	5 (4.7)	7 (4.6)
+−				4 (2.6)
−−	9 (18.0)	14 (25.0)	23 (21.7)	73 (47.7)

[a] $N = 50$.
[b] $N = 56$.
[c] $N = 106$.
[d] From Ballinger et al. (1992). $N = 153$.
[e] Includes three samples with the deletion and one with the triplication.
[f] All of these mtDNAs carry the deletion.

currence at an appreciable frequency (7.5%) in both groups of the Tharus examined and its absence from the sample of 76 Hindus (table 2) support the hypothesis.

We also observed a case of triplication of the same 9-bp unit. As far as we know, this is the second instance of a complete triplication of this repeat, the first having been found by Shields et al. (1992) in a single Chukchi from Siberia.

The relationships between the Tharu deletion or triplication mtDNA haplotypes and those of the phylogenetic tree of Ballinger et al. (1992) are shown in figure 5, where only the relevant part of that tree is represented. Four different deletion haplotypes were observed in the Tharus, three of which fall into the C group with a new single-step derivative, haplotype 1^{Tharu}; the fourth, haplotype 2^{Tharu}, is another new haplotype which can be related, through three steps, either to the C or the E group. The derivation from the C group requires a new *Alu*I site gain at nt 10397, whereas that from the E group requires the occurrence of a new deletion event. The second hypothesis appears more likely because it requires a less rare event: five length changes in the relevant region and only one acquisition of the *AluI*$_{10397}$ site are postulated in Ballinger's parsimonious tree.

The mtDNA with triplication derives from the E group through two intermediate haplotypes. Although we cannot directly compare the Tharu triplication with that of Shields et al. (1992) (the latter mtDNA haplotype is not known), we believe that they have a distinct origin. The Tharu triplication is on an mtDNA molecule that carries a mutation (an A-to-G transition at nt 15487) which is very frequent (20%) in eastern Tharus (Passarino et al. 1992) and which has never been reported for any other East Asian group. The triplication, therefore, must have occurred in a molecule with the mutation that is peculiar to the Tharus.

To summarize, Ballinger et al. (1992) postulated at least three deletion events and two cases of partial triplication, one of which had been described elsewhere (Cann and Wilson 1983). This survey has found what is probably a new deletion event (haplotype 2^{Tharu}), together with a complete triplication. The origin of the latter is most likely independent of that found in Siberia by Shields et al. (1992). Taken together, these findings suggest that this region of two 9-bp repeats is a hot spot of rearrangements.

Apart from sporadic cases, all the identified deletion mtDNA haplotypes cluster in the C and in the D* groups (fig. 4). The C group is considered to be the oldest one because it includes the most divergent haplotypes, which are spread throughout Southeast Asia (Ballinger et al. 1992). The D* group is also frequent in Asia and, in addition, encompasses all the deletion haplotypes found in Native Americans (Torroni et al. 1992) and in Papua New Guinea (Stoneking et al. 1990; Ballinger et al. 1992.) On the other hand, no deletion haplotype of the Tharus falls within the D* group (fig. 5). This suggests that the Tharu mtDNA originated from an Oriental mtDNA that differed from that of the populations who migrated toward the Pacific Islands, and northward, through Siberia, to the Americas.

With regard to the question of a possible genetic heterogeneity within the Tharus, it should be pointed out that the mtDNA deletion haplotypes of the central

Tharus differ from those found in the eastern Tharus (fig. 5). While it is clear that a larger sample of deletion cases is required to answer the question, the observation is in line with the heterogeneity revealed by other markers. Two examples of this heterogeneity are provided both by differences in the frequency of the *Hae*II morph 5, which accounts for 25% of the central Tharus and for 9% of the eastern Tharus (Brega et al. 1986; Passarino et al. 1992), and by the extreme heterogeneity of the α-thalassemia gene (G. Modiano, unpublished data).

Irrespective of the possible occurrence of an intra-Tharu heterogeneity, these two Tharu subgroups, although in close vicinity to an overwhelming Caucasoid population (i.e., Hindus), display both morphological (Bista 1980) and genetic (Chopra 1970; Brega et al. 1986; Passarino et al. 1992, and present paper) Oriental features. This strongly implies that they have a common origin, and the implication is further supported by the identical and particularly high frequency of the $DdeI_{10394}$, $AluI_{10397}$ (+ +) combination in each of the Tharu groups (table 6).

In summary, (1) 74% of the Tharus (including all those individuals with the classical *Hae*II morph 5 Oriental mtDNA marker and three individuals with the 9-bp deletion) have the $DdeI_{10394}$, $AluI_{10397}$ (+ +) phenotype; (2) all of the few encountered $DdeI_{10394}$, $AluI_{10397}$ (+ −) mtDNAs carry the East Asian deletion; and (3) one-third of the 22% $DdeI_{10394}$, $AluI_{10397}$ (− −) mtDNAs (7% of the total sample) are *Hpa*I/*Hinc*II morph 1, one of the first recognized Oriental mitochondrial markers (Denaro et al. 1981; Blanc et al. 1983).

Although more extensive studies are required to precisely quantify the value of the $DdeI_{10394}$, $AluI_{10397}$ (+ +) phenotype as an indicator of Oriental affinities, we conclude that a substantial part of the Tharu mtDNAs are East Asian varieties. Thus, popular stories claiming a predominant Hindu maternal ancestry for the Tharus seem to be disproved by genetic data.

Acknowledgments

We are most grateful to Dr. S. Shresta for his valuable help in the blood collection. Thanks are also due to William Cooke, English lecturer at the University of Pavia, for revision of the English of the manuscript. This work received funds from MURST 40%, CNR Target Project on Biotechnology and Bioinstrumentation, CNR PF Ingegneria Genetica (to A.S.S.-B.); and by funds of "Grandi progetti d'Ateneo" of the University of Rome "La Sapienza" (to G.M.). G.P. was supported by a CNR fellowship.

References

Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

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

Bista DB (ed) (1980) People of Nepal, 4th ed. Ratna Pustak Bhandar, Kathmandu, Nepal

Blanc H, Chen K-H, D'Amore MA, Wallace DC (1983) Amino acid change associated with the major polymorphic Hinc II site of Oriental and Caucasian mitochondrial DNAs. Am J Hum Genet 35:167–176

Brega A, Gardella R, Semino O, Morpurgo G, Astaldi Ricotti GB, Wallace DC, Santachiara Benerecetti AS (1986) Genetic studies on the Tharu population of Nepal: restriction endonuclease polymorphisms of mitochondrial DNA. Am J Hum Genet 39:502–512

Brown MD, Voljavec AS, Lott MT, Torroni A, Yang C-C, Wallace DC (1992) Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. Genetics 130:163–173

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cann RL, Wilson AC (1983) Length mutations in human mitochondrial DNA. Genetics 104:699–711

Chopra VP (1970) Studies on serum groups in the Kuman region, India. Humangenetik 10:35–43

Denaro M, Blanc H, Johnson MJ, Chen K-H, Wilmsen E, Cavalli Sforza LL, Wallace DC (1981) Ethnic variation in HpaI endonuclease cleavage patterns of human mitochondrial DNA. Proc Natl Acad Sci USA 78:5768–5772

Harihara S, Momoki H, Suutou Y, Shimizu K, Omoto K (1992) Frequency of a 9bp deletion in the mitochondrial DNA among Asian populations. Hum Biol 64:161–166

Hertzberg M, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Horai ST, Matsunaga E (1986) Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum Genet 72:105–177

Modiano G, Morpurgo G, Terrenato L, Novelletto A, Di Rienzo A, Colombo B, Purpura M, et al (1991) Protection against malaria morbidity: near-fixation of the α-thalasse-

mia gene in a Nepalese population. Am J Hum Genet 48: 390–397

Passarino G, Semino O, Pepe G, Shrestha SL, Modiano G, Santachiara Benerecetti AS (1992) MtDNA polymorphisms among Tharus of eastern Terai. Gene Geogr 6:139–147

Shields GF, Hecker K, Voevoda MI, Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in native Beringians. Am J Hum Genet 50:758–765

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Terrenato L, Shrestha SL, Dixit KA, Luzzatto L, Modiano G, Morporgo G, Arese P (1988) Decreased malaria morbidity in the Tharu people compared to sympatric populations in Nepal. Ann Tropical Med Parasitol 82:1–11

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Wallace DC, Torroni A (1992) American Indian prehistory as written in the mitochondrial DNA: a review. Hum Biol 64:403–416

Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Res 15:529–542