AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 109:147–158 (1999)

Multiple Origins of the mtDNA 9-bp Deletion in Populations of South India

W.S. WATKINS,[1]* M. BAMSHAD,[2] M.E. DIXON,[1] B. BHASKARA RAO,[3] J.M. NAIDU,[3] P.G. REDDY,[4] B.V.R. PRASAD,[3] P.K. DAS,[5] P.C. REDDY,[6] P.B. GAI,[7] A. BHANU,[8] Y.S. KUSUMA,[3] J.K. LUM,[1] P. FISCHER,[2] AND L.B. JORDE[1]
[1]*Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112*
[2]*Department of Pediatrics, University of Utah, Salt Lake City, Utah 84112*
[3]*Department of Anthropology, Andhra University, Visakhapatnam, India*
[4]*Department of Anthropology, University of Madras, Madras, India*
[5]*Department of Anthropology, Utkal University, Bhubaneswar, India*
[6]*Department of Anthropology, Sri Venkateswara University, Tirupati, India*
[7]*Department of Anthropology, Karnatak University, Dharwad, India*
[8]*Department of Anthropology, University of Calicut, Kerala, India*

KEY WORDS mitochondria; evolution; 9-bp deletion; control region

ABSTRACT The origins and genetic affinities of the more than 500 tribal populations living in South Asia are widely disputed. This may reflect differential contributions that continental populations have made to tribal groups in South Asia. We assayed for the presence of the intergenic COII/tRNALys 9-bp deletion in human mtDNA in 646 individuals from 12 caste and 14 tribal populations of South India and compared them to individuals from Africa, Europe, and Asia. The 9-bp deletion is observed in four South Indian tribal populations, the Irula, Yanadi, Siddi, and Maria Gond, and in the Nicobarese. Length polymorphisms of the 9-bp motif are present in the Santal, Khonda Dora, and Jalari, all of whom live in a circumscribed region on the eastern Indian coast. Phylogenetic analyses of mtDNA control region sequence from individuals with the 9-bp deletion indicate that it has arisen independently in some Indian tribal populations. Other 9-bp deletion haplotypes are likely to be of Asian and African origin, implying multiple origins of the 9-bp deletion in South India. These results demonstrate varying genetic affinities of different South Indian tribes to continental populations and underscore the complex histories of the tribal populations living in South Asia. Am J Phys Anthropol 109:147–158, 1999.

The genetic diversity of caste and tribal populations on the Indian subcontinent is well-established (Mujumder, 1998). This results, in part, from three major migration events. Descendants of Paleolithic expansions into India are believed to have contributed substantially to tribal populations, which currently account for approximately 7.5% of the population (Majumder and Mukherjee, 1993). A second immigration event, 10,000 years ago, spread proto-Dravidian-speaking Neolithic farmers throughout the Indian subcontinent and was followed by a third wave of Indo-European-speaking "Caucasoids" entering from West-Central Asia 3,500 years ago. Throughout India,

Grant sponsor: NSF; Grant numbers: SBR-9514733, SBR-9512178; Grant sponsor: NIH; Grant number: PHS MOI-0064; Grant sponsor: Primary Children's Center Foundation.

The data used in these analyses are available from the authors by request. The use of the term "caste" for population subdivision in this paper reflects a historical social hierarchy that is used here for population analyses.

*Correspondence to: W.S. Watkins, Department of Human Genetics, 15 E 2030 N, Room 2100, University of Utah, Salt Lake City, Utah 84112-0533. E-mail: swatkins@genetics.utah.edu

Received 10 September 1998; accepted 26 February 1999.

Indo-European speakers introduced the Hindu caste system, a social hierarchy that strictly governs marriage and mating practices. However, Indo-European-speaking migrants had less influence on South India, and thus South Indian populations remain linguistically and genetically distinct from North Indian populations (Majumder, 1998).

Numerous studies have examined the genetic relationships between Indian populations. Analyses using classical protein polymorphisms suggest that geographic proximity may be an essential determinant of genetic affinities between the tribal populations of Andhra Pradesh (Papiha et al., 1997; Sirajuddin et al., 1994). Indian populations have also been examined at the molecular level using selectively neutral mitochondrial DNA (mtDNA) and Y-chromosome markers (Bamshad et al., 1996, 1998; Cavalli-Sforza et al., 1994; Mountain et al., 1995; Passarino et al., 1996b). These sensitive molecular methods have detected sex-specific gene flow leading to genetic stratification of Hindu castes from the same geographical area (Bamshad et al., 1998).

A 9-bp deletion in the maternally inherited mitochondrial genome has been useful for examining genetic relationships between human populations (Redd et al., 1995; Soodyall et al., 1995, 1996). The 9-bp deletion is a length variant in the intergenic region between the cytochrome c oxidase subunit II and the mitochondrial tRNA for lysine (COII/$tRNA^{Lys}$). The deletion motif results from the loss of one of two CCCCCTCTA tandem repeats (Cann and Wilson, 1983; Wrischnik et al., 1987).

Originally identified as an "Asian-specific" polymorphism (Wrischnik et al., 1987), this 9-bp deletion motif has been found in several populations. There is a clinal increase in the frequency of the 9-bp deletion, with a corresponding decrease in mtDNA diversity among Pacific Islanders (Redd et al., 1995). The 9-bp deletion is nearly fixed in Samoans, Maoris, Niueans (Hertzberg et al., 1989), Native Hawaiians, and other Polynesian groups (Lum and Cann, 1998). The 9-bp deletion is found at substantially lower frequencies in a geographic cline along the eastern Asia coast (Ballinger et al., 1992; Harihara et al., 1992). Phylogenetic analyses of African mtDNA sequences suggest multiple independent origins of the 9-bp deletion in Central Africa (Soodyall et al., 1996). Aboriginal Australians with the 9-bp deletion have control region sequences distinct from those of Asians with the deletion (Betty et al., 1996). A triplication of the 9-bp repeat and a T to C transition with subsequent expansion of the homopolymeric C tract have been reported (Lum and Cann, 1998; Wrischnik et al., 1987). However, few data regarding the 9-bp deletion motif exist for populations living on the Indian subcontinent.

The presence of this marker within India could help to identify evolutionary relationships that predate caste or tribal divisions and may be useful for tracking migration events into the subcontinent. To better understand the genetic relationships and the distribution of the mtDNA 9-bp deletion in populations from South India, we assayed 317 caste members and 329 tribal Indians for the 9-bp deletion motif. Here we report the presence of the mtDNA 9-bp deletion and polymorphic variants in populations of South India and the Nicobar Islands, including a de novo 9-bp deletion in tribal South Indians.

MATERIALS AND METHODS

Study subjects

All studies were approved by the Institutional Review Board at the University of Utah and the administration of Andhra University. Caste participants were selected from populations spanning the Hindu caste hierarchy. All caste members live in Andhra Pradesh and speak Telegu, a Dravidian language. Caste populations include: Brahmin (n = 59), Jalari (n = 28) Kapu (n = 58), Kshatriyas (n = 12), Madiga (n = 29), Mala (n = 26), Relli (n = 19), Vysya (n = 10), Wadabahija (n = 23), and Yadava (n = 53).

The tribal populations used in this study are geographically dispersed throughout South India (Fig. 1). They include: Chenchu (n = 11), a proto-Australoid tribe of hunters and gatherers from Andhra Pradesh; Gadaba (n = 30), a proto-Australoid, Mundari-speaking tribe of horticulturists from Andhra Pradesh and Orissa; Irula (n = 34), a proto-

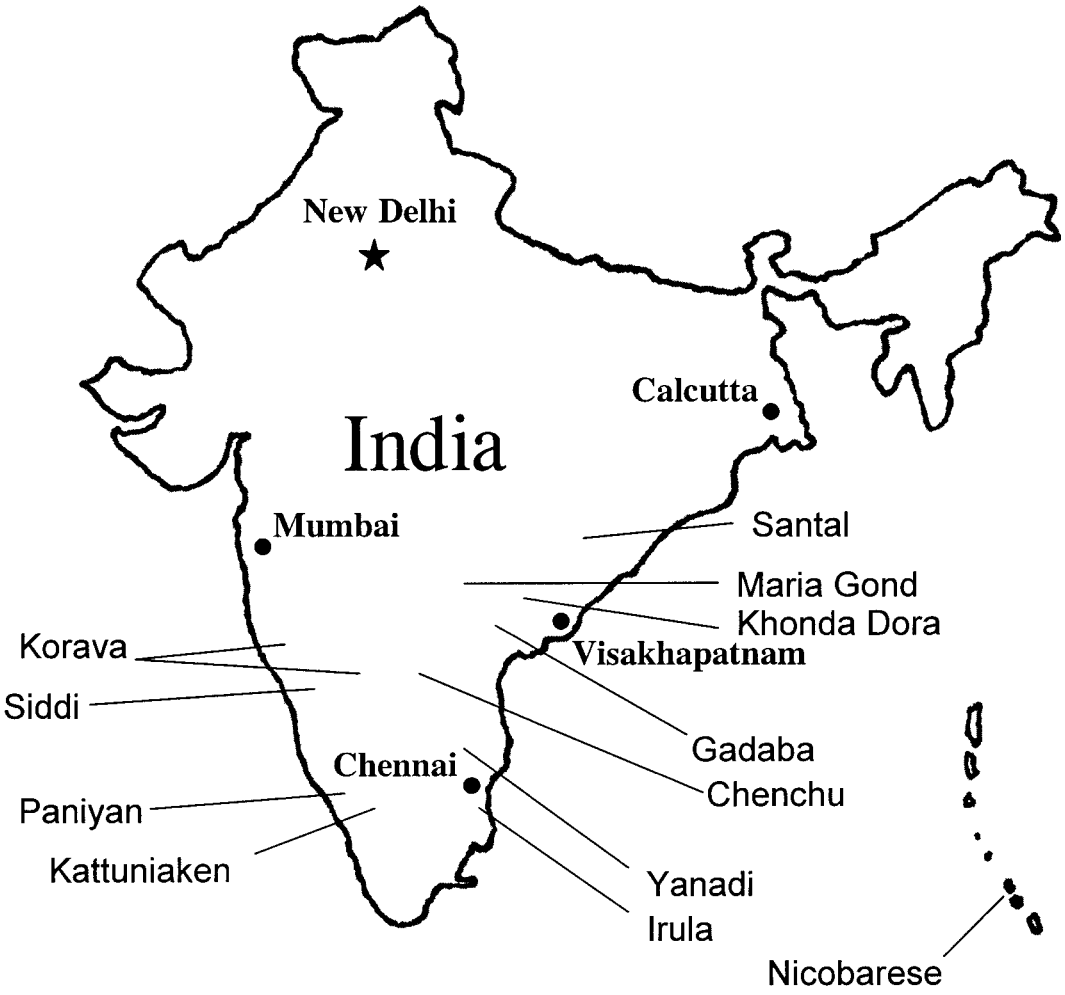

Fig. 1. A map of India indicates the approximate locations of tribal populations participating in the study. Caste samples were collected in Visakhapatnam, within the state of Andhra Pradesh on the eastern coast of South India.

Australoid, Telegu-speaking tribe of hunters and gatherers from the coastal region of Southern Andhra Pradesh; Kattunaicken (n = 22), a proto-Asian, Kannada-speaking tribe of hunters and gatherers from Kerala; Khonda Dora (n = 27), a proto-Asian, Telegu-speaking tribe of hunters and gatherers from Northern Andhra Pradesh; Korava (n = 24), a proto-Australoid, Telegu- and Tamil-speaking tribe of musicians and agriculturists from Karnataka; Maria Gond 1, 2, and 3 (n = 20, 22, and 22, respectively), three distinct geographical isolates of a proto-Asian tribe of agriculturists and food gatherers from Madya Pradesh; Paniyan (n = 24), a Negrito tribe of agriculturists from Kerala; Santal (n = 16), a proto-Australoid, Mundari-speaking tribe of food-gatherers from Orissa; Siddi (n = 24), a Negrito tribe from Karnataka brought to India in the 15th century as slaves; Yanadi (n = 24), a proto-Australoid, Telegu-speaking tribe of hunters and gatherers from Andhra Pradesh; and Nicobarese (n = 29), an Asian tribe of hunters and gatherers from the Nicobar Islands who speak dialects of an Austronesian language. African and Asian samples have been previously described (Jorde et al., 1995). An addi-

tional 24 Kenyans, 33 Mbuti, 12 Alur, 17 Hema, and 18 Nande individuals are included in the African data set (unpublished data).

DNA analysis

Whole blood was collected in either 8-ml or 0.5-ml EDTA collection tubes by venous blood draw or finger prick, respectively. In most cases, a hair sample was also collected. Samples were transported to the Laboratory of Biological Anthropology at Andhra University where DNA was extracted using a Puregene (Minneapolis, MN) kit according to the manufacturer's specifications. To test individuals for the 9-bp deletion, 25 ng of genomic DNA were amplified using primers 9DELA (ATGCTAAGTTAGCTTTACAG) and 9DELB (ACAGTTTCATGCCCATCGTC). The 9DELA primer was end-labeled using γ^{32}-PATP and polynucleotide kinase. The mtDNA sequences were amplified in 1 × PCR buffer (10 mM Tris, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$) using 200 µM dNTPs, 20 pmol of primers 9DELA and 9DELB, 1 pmol of radiolabeled 9DELA primer, and 1 unit of *Taq* DNA polymerase in a total reaction volume of 25 µl. Samples were cycled using a standard three-step PCR profile with the annealing temperature for the first five cycles set to 58°C and then lowered to 54°C for an additional 25 cycles. PCR products were resolved on a 5% urea/formamide denaturing polyacrylamide gel and visualized by autoradiography.

For sequencing of the 9-bp deletion motif, amplification products were ligated into the pCR2.1 plasmid (Invitrogen, Carlsbad, CA), using conditions specified by the manufacturer. Ligation mixes were used to transform XL1-Blue cells (Stratagene, La Jolla, CA), and transformants were selected on ampicillin/IPTG/X-gal plates. DNA was prepared from clones using a standard miniprep protocol. Clones were then sequenced using the universal and reverse primer sites in the vector.

To generate the HVS-1 sequence, a 1.1-kb amplicon from the mtDNA control region was amplified using conditions described above with primers UPL15996, RPH408, and H16401 (Bamshad et al., 1996). The sequence for HVS-1 was generated from the UPL15996 and H16401 primer sites using ABI (Foster City, CA) Dye-primer or dRhodamine sequencing reagents and an ABI 377 automated DNA sequencer. Sequence data were compared and edited using the Sequencher software package (Genecodes, Ann Arbor, MI).

Statistical analysis

A data set of 411 bp of the HVS-1 sequence for 673 individuals scored for the 9-bp deletion was assembled. This data set included the HVS-1 sequence for 122 tribal (Irula, Yanadi, Maria Gond, Khonda Dora, and Nicobarese) and 250 caste individuals. An additional 143 Africans, 61 Asians, 89 Caucasians (Northern Europeans) (Jorde et al., 1995), and 8 Australian Aborigines were included. After removing 194 redundant haplotypes to reduce the data set to 479 unique sequences, genetic distances were calculated by the Kimura two-parameter model with a transition/transversion ratio of 10:1 using the DNADIST program in the PHYLIP program package, and a neighbor-joining (N-J) tree of individuals was created by using NEIGHBOR (Felsenstein, 1993).

Support for monophyly of clusters with the 9-bp deletion was assessed using a modification of Templeton's test as implemented in PHYLIP. The test provides a statistical measure of differences between alternate arrangements of a phylogenetic tree. An N-J tree of 479 unique haplotypes with 906 steps was created and used as a minimum-step "best" tree. Branches containing Indian deletion haplotypes were moved to other continental 9-bp deletion clusters, and new (longer) trees were compared to the "best" N-J tree, using the DNAPARS program in PHYLIP. Rearranged trees in which the number of steps exceeded that of the best tree by more than 1.96 standard deviations were rejected.

An N-J network of populations was constructed from 664 individuals subdivided into 10 populations. Genetic distance matrices were calculated as described above, and the genetic distance, d_a, between populations X and Y was estimated as:

$$d_a = d_{xy} - (d_x + d_y)/2$$

(equation 10.21 in Nei, 1987). An N-J tree was constructed using the NEIGHBOR pro-

gram, as described above. To assess branch support of the original tree, 100 bootstrap replicates were generated with SEQBOOT, and population trees were compared using the CONSENSE program (Felsenstein, 1993).

A maximum-likelihood network of the 27 unique haplotypes with the 9-bp deletion was constructed using the PUZZLE program (Strimmer and von Haeseler, 1997). Nucleotide frequencies and the transition/transversion ratio were estimated from the data. A gamma correction ($\alpha = 0.08$) was estimated from the data and used to compensate for the nonuniform mutation rate in the mtDNA control region (Strimmer and von Haeseler, 1997; Tamura and Nei, 1993; Yang and Kumar, 1996)).

To estimate coalescent dates for mtDNA haplotypes, the average nucleotide diversity for a 9-bp deletion cluster was calculated using the Kimura two-parameter model with a gamma correction of $\alpha = 0.08$ that was estimated directly from the data. Calculations were performed using the Arlequin software package (Schneider et al., 1997). A time to coalescence was then estimated by:

$$T = d/2\lambda,$$

where d is the genetic diversity, λ is the nucleotide substitution rate, and T is the coalescent time (Nei, 1987; Soodyall et al., 1996). Although this model makes several assumptions such as selective neutrality and mutation/drift equilibrium, it provides a useful comparison of the relative ages of the deletion clusters.

RESULTS

Mitochondrial 9-bp deletion polymorphisms are found at varying frequencies in South Indian populations (Table 1). A 9-bp deletion motif is found at a high frequency in the Irula and Yanadi tribes and at lower frequencies in the Maria Gond, Siddi, and Nicobarese. The overall frequency of the 9-bp deletion motif in the Indian tribal population is 0.12. Two individuals from caste populations (Brahmin and Yadava) had the 9-bp deletion, although the Brahmin appeared to be heteroplasmic for the deletion (data not shown). We also surveyed more than 350 additional individuals from worldwide populations and found 15 of 179 Africans, 7 of 78 Asians, 0 of 8 Australian Aborigines, and 0 of 90 Caucasians with the 9-bp deletion.

TABLE 1. Frequency of 9-bp deletion motifs in South Indian populations[1]

	n	1	2−	2	2+
Caste populations					
Brahmin	59	0.02		0.92	0.07
Jalari	28		0.11	0.75	0.14
Kapu	58			1.00	
Kshatriyas	12			1.00	
Madiga	29			1.00	
Mala	26			1.00	
Relli	19		0.05	0.95	
Vysya	10			1.00	
Wadabahija	23			1.00	
Yadava	53	0.02		0.98	
Caste totals	317	0.01	0.01	0.96	0.02
Tribal populations					
Chenchu	11			1.00	
Gadaba	30			1.00	
Irula	34	0.44		0.56	
Kattunaicken	22			1.00	
Khonda Dora	27			0.74	0.26
Korava	24			1.00	
Maria Gond 1	20			1.00	
Maria Gond 2	22			1.00	
Maria Gond 3	22	0.09		0.91	
Nicobarese	29	0.24		0.76	
Paniyan	24			1.00	
Santal	16			0.81	0.19
Siddi	24	0.04		0.96	
Yanadi	24	0.68		0.32	
Tribal totals	329	0.12		0.85	0.03

[1] 1, [CCCCCTCTA]; 2−, two repeats, contracted; 2, $[CCCCCTCTA]_2$; 2+, two repeats, expanded.

Expanded repeat motifs are found in the Brahmin and Jalari castes and in the Khonda Dora and Santal tribes. Cloning and sequencing of 2 individuals with expanded alleles from the Khonda Dora and Santal revealed population-specific differences. Figure 2 shows the Cambridge reference sequence for the 9-bp repeat motif and the sequences from the four expanded alleles. All expanded alleles could be generated by a T to C transition, followed by expansion of the homopolymeric C tract. The expansion occurs in the 3′ repeat in the Khonda Dora samples, while the 5′ tract undergoes expansion in Santal individuals. Individuals from the Jalari and the Relli castes have a length variant intermediate between 1–2 repeats in length.

We identified 27 unique HVS-1 haplotypes with the 9-bp deletion, including 7 haplotypes from the Indian subcontinent and 5 Nicobarese haplotypes (Table 2). Four

Sample	5'–repeat	3'–repeat	Length
CRS	CCCCCTCTA	CCCCCTCTA	18bp
KD1	CCCCCTCTA	CCCCCCCCCCTA	21bp
KD7	CCCCTTCTA	CCCCCCCCCCCTA	22bp
ST8	CCCCCCCCCCTA	CCCCCTCTA	21bp
ST9	CCCCCCCCCCCTA	CCCCCTCTA	22bp

Fig. 2. Sequence of the expanded 9-bp repeat motif in 4 individuals from the Khonda Dora (KD) and Santal (ST) tribes. The 3′ repeat has expanded in the KD population, while the 5′ repeat has expanded in the ST population. The repeat motifs are compared to the Cambridge Reference Sequence (CRS), base pairs 8272-8289. Numbers after KD and ST are laboratory identifications.

deletion haplotypes that shared a unique transversion at base 16256 were found in the Irula (I) and Yanadi (YN). Several additional polymorphisms (16017, 16126, 16184, 16290, 16296, and 16325) found in Indians were not present on other 9-bp deletion haplotypes. All Yanadi with the 9-bp deletion had the I20 haplotype. Nicobarese (NIC) 9-bp deletion haplotypes were distinct from tribal and caste Indian haplotypes. Some 9-bp deletion haplotypes occurred in several individuals with the same HVS-1 sequence.

A comparison of the 194 shared HVS-1 sequences in individuals without the 9-bp deletion revealed little haplotype sharing between the major population groups. Indians did not share haplotypes with Africans or Asians, but shared four haplotypes with Caucasians. Within the Indian subcontinent, 6 of 50 unique tribal haplotypes were shared with caste populations. The Nicobarese shared a single haplotype with Asians and no haplotypes with caste or tribal Indians. Thus, haplotype sharing suggests only limited levels of very recent female gene flow between the major (sub) continental groups surveyed.

An N-J tree of 479 unique HVS-1 sequences (tree not shown) placed Asian, African, and Irula/Yanadi 9-bp deletion haplotypes in three separate clusters. Rearrangements of the tree were evaluated using a modification of Templeton's test. Trees in which the tribal Irula and Yanadi with the deletion were moved into Asian or African deletion clusters were significantly worse ($P < 0.05$), rejecting the hypothesis that tribal Irula and Yanadi are monophyletic with Asian or African deletion clusters. All Nicobarese deletion haplotypes, with one exception, fell within the Asian deletion cluster.

A maximum-likelihood tree was constructed by quartet puzzling for the 27 unique 9-bp deletion haplotypes (Fig. 3). The estimated transition/transversion ratio for the 27 haplotypes was 4.4; the mean pairwise sequence divergence was 0.029. All Irula and Yanadi 9-bp deletion haplotypes clustered together and were distinct from other deletion haplotypes with high statistical support (97%). Deletion haplotypes of the Maria Gond and Yadava clustered together. Most Nicobarese clustered together and were more closely related to Asians than other tribal Indians.

An N-J network constructed from genetic distances between populations indicated that tribal Indians with the 9-bp deletion were more closely related to tribal Indians without the deletion than to any other group (Fig. 4). Africans with the 9-bp deletion cluster with nondeletion Africans, and Asians and Nicobarese with the 9-bp deletion, clustered together. The genetic distances between caste, Asian, and Caucasian groups were relatively small, a result illustrated by short branch lengths with lower branch support on the N-J network. Altering the transition/transversion ratio or changing the mutation rate bias did not change the topology of the tree. Table 3 shows the genetic distances between the three major deletion clusters and (sub) continental populations. The three deletion groups were well-sepa-

TABLE 2. HVS-1 haplotypes for African, Asian, Indian, and Nicobarese with the 9-bp deletion[1]

	16017[2]	16093	16126[2]	16129	16140	16147	16148	16172	16183	16184[2]	16187	16188	16189	16195	16217	16223	16230	16235	16241	16242	16254	16256[2]	16260	16261	16266	16267	16274	16275	16287	16290[2]	16293	16294	16295	16296[2]	16311	16316	16320	16321	16325[2]	16360	16362	16388	16390	f
CRS	T	T	T	G	T	C	C	T	A	C	C	C	T	T	T	C	A	A	A	C	A	C	C	C	C	C	G	A	C	C	A	C	C	C	T	A	C	C	T	C	T	G	G	
African																																												
AFP12ID+							T	C			T	A	C			T	G			T															C		T						A	0.035
PED17+							T	C			T	G	C			T	G																	C		T							0.028
P10469A							T	C			T	A	C			T	G			T												T		C		T						A	0.007
PED59							T	C			T	G	C																					C		T							0.007
PED60							T	C			T	G	C			T	G				G													C		T							0.007
XH21		C					T	C			T	G	C			T	G																	C		T							0.007
AFN15							T	C	G			A	C			T	G																	C		T							0.007
AFN19							T	C			T	G	C			T	G																	C		T						A	0.007
Asian																																											
CAM191									C	A			C		C			G																									0.016
CAM376					C								C												A	T					G											0.016
CAM377				A	C								C											T	A																	0.016
CAM177													C		C				G																							0.016
C2698													C											T							G											0.016
C2732						T			C	A			C			T																								C			0.016
TAW17				A									C	G	C									T			A															0.016
Indian																																											
I20+										T						T						G												C					C			0.152
I4+										T						T						G										A		C					C			0.020
I19										T												G												C					C			0.010
I21										T						T						G												C					C		A	0.010
MG3-11			C													T													T													0.010
MG3-21		C	C																								A		T													0.010
Y15	C		C																									G			T		T					C			0.010
Nicobarese																																											
NIC39+					C				C				C												A															T			0.087
NIC10					C				C				C												A																		0.043
NIC24					C				C				C							A					A			G	A							T		T		T			0.043
NIC31					C				C				C							A					A			G											T			0.043
NIC44					C				C				C			T							T								T											0.043

[1] P, AFP, Mbuti Pygmy; PED, Pedi; AFN, Nande; XH, Sotho-Tswana; CAM, Cambodian; C, Chinese; TAW, Taiwanese; I, Irula; MG, Maria Gond; Y, Yadava; NIC, Nicobarese. +, haplotype found in multiple individuals; f, population-specific frequency of the HVS-1 haplotype.
[2] Unique polymorphisms found on South Indian deletion haplotypes.

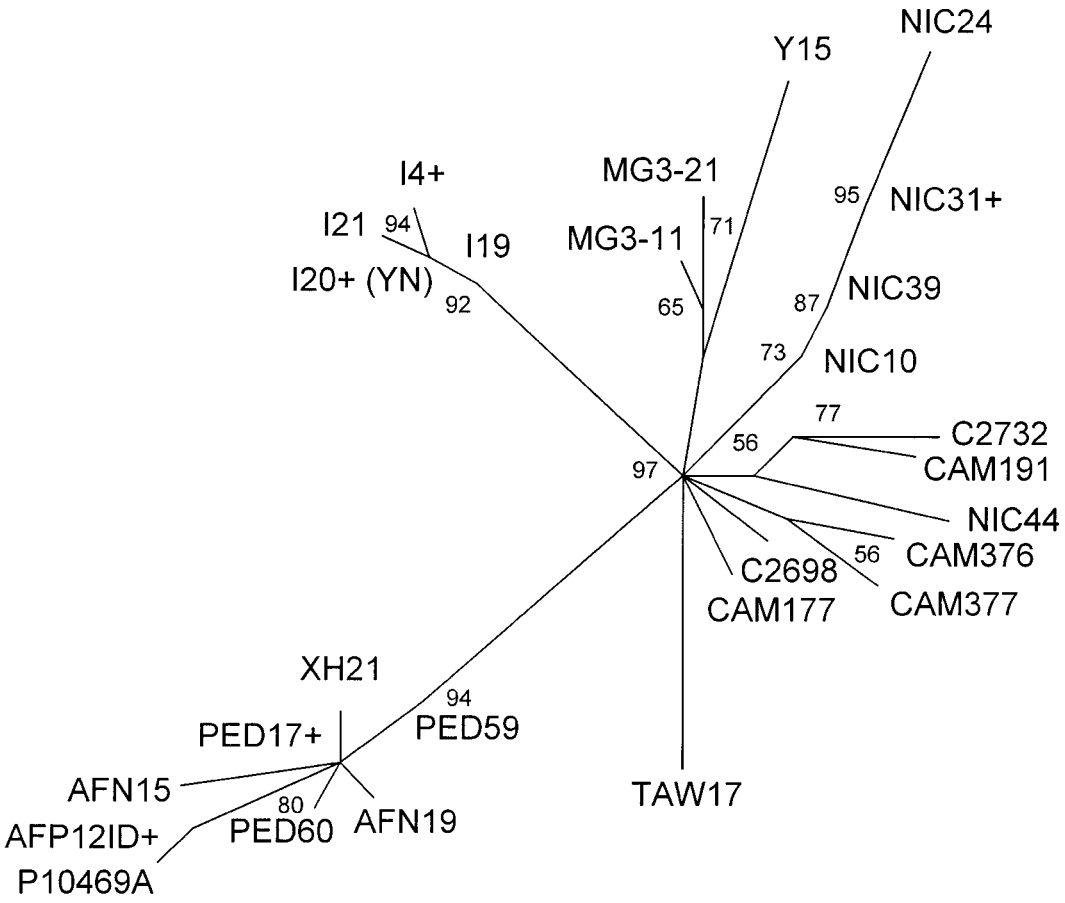

Fig. 3. A maximum-likelihood network comparing 27 unique 9-bp deletion haplotypes. Branch support values are shown as a percentage, and branch lengths reflect genetic distances. Indian Irula (I) and Yanadi (YN) haplotypes cluster with high branch support. Maria Gond (MG) and Yadava (Y) haplotypes are grouped together, distinct from the Irula and Yanadi. African haplotypes include Mbuti (P, AFP), Nande (AFN), and Sotho-Tswana (PED, XH). Asian haplotypes include Cambodian (CAM), Chinese (C), and Taiwanese (TAW).

rated from each other by large genetic distances. The distances between the Indian deletion group and the caste, tribal, and Asian continental populations were similar. The Asian deletion group was closer to Asians, castes, and tribal Indians. A comparison of the Asian deletion group and the Indian deletion group reveals that this result may reflect the small sample size, drift, and bottleneck effects, and a unique transversion on the Indian deletion haplotypes. The Asian deletion cluster was equidistant to castes and Asians but farther from tribal Indians. The genetic distance between caste and Asian continental populations was small.

To further characterize the evolutionary history of the 9-bp deletion in Indian tribal populations, coalescent dates for the 9-bp deletion clusters were estimated from the HVS-1 sequence. Individuals not included within the main Asian or Indian deletion clusters by the phylogenetic analysis previously described were not included in the estimates. The average nucleotide diversity between haplotypes was calculated for each deletion cluster. Diversity estimates were then used to calculate a coalescent time for each cluster, using an mtDNA control region substitution rate of 1.142×10^{-7} per site per year (Stoneking et al., 1992). A nucleotide

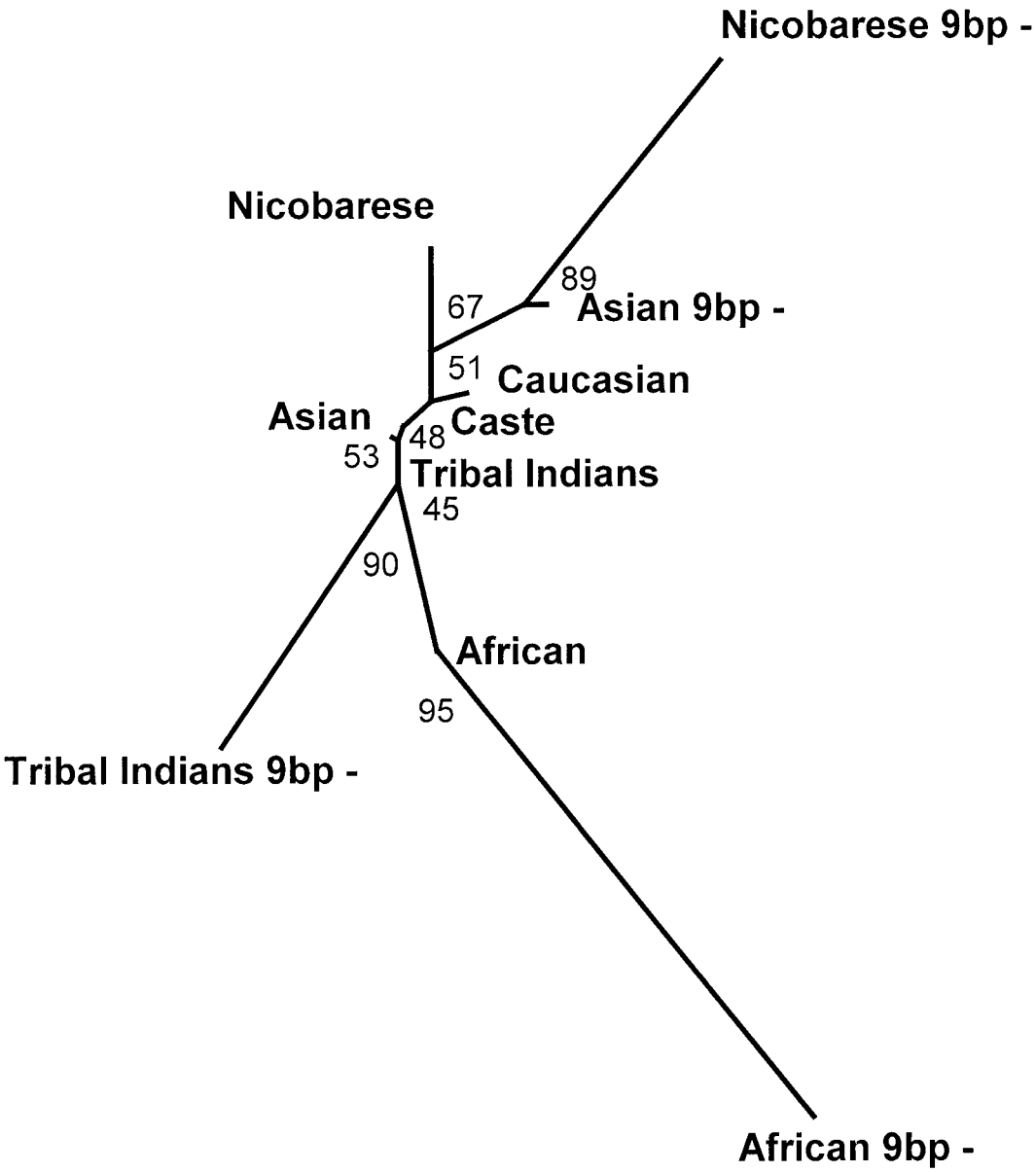

Fig. 4. An N-J tree of worldwide populations with African, Asian, South Indian, and Nicobarese 9-bp deletion groups (9 bp -). Bootstrap support indicated as percentages at the branch points.

diversity estimate of 0.000757 ± 0.000895 suggested a coalescent time of approximately 3,300 years for the Irula and Yanadi deletion cluster. This contrasts with more ancient coalescent times of approximately 71,000 years for the Asian and 36,000 years for the African deletion haplotypes, based on diversity estimates of 0.01629 ± 0.01034 and 0.00834 ± 0.00506, respectively. The Asian coalescent date is in general agreement with estimates of 58,000 years made by Redd et al. (1995). We estimate a more recent coalescent for the African cluster than previously reported (72,000 years) (Soodyall et al., 1996), but this result may be due to our more limited sampling of Africans.

TABLE 3. Genetic distances between deletion clusters and continental populations

	African 9-bp−	Asian 9-bp−	Indian 9-bp−	African	Asian	Caste
Asian 9-bp−	0.0223					
Indian 9-bp−	0.0247	0.0173				
African	0.0123	0.0084	0.0109			
Asian	0.0187	0.0059	0.0085	0.0040		
Caste	0.0182	0.0058	0.0092	0.0039	0.0005	
Tribal	0.0170	0.0072	0.0087	0.0036	0.0011	0.0013

DISCUSSION

The mitochondrial intergenic COII/tRNA^{Lys} 9-bp deletion marker is present at a high frequency and is found on at least four different HVS-1 haplotype backgrounds in the coastal Irula and Yanadi tribal populations of South India. Phylogenetic analyses indicate that Irula and Yanadi with the deletion are more similar to each other and to tribal populations without the deletion than they are to Asians or Africans with the 9-bp deletion. Tribal Indians with the deletion share haplotypes with other Indians without the deletion, but do not share haplotypes with Asians or Africans. These data support an independent origin of the 9-bp deletion on the Indian subcontinent.

A frequency cline of mtDNA restriction site polymorphisms, *Dde*I$_{10394}$/*Alu*I$_{10397}$ (+ +), increases from north to south in Indian populations and may be an indicator of mitochondrial haplotypes that predate Indo-European invasions into the subcontinent (Passarino et al., 1996a,b). Preliminary data (not shown) indicate that most tribal Indians with or without the 9-bp deletion have the *Dde*I$_{10394}$/*Alu*I$_{10397}$ (+ +) haplotype. In contrast, the Yadava individual and 6 of 7 Asians with the 9-bp deletion did not have the (+ +) haplotype. These observations provide further support for a de novo 9-bp deletion in South Indian tribal populations.

Coalescence estimates differ substantially for African, Asian, and Indian deletion clusters. The Indian 9-bp deletion has a very recent coalescent date, whereas the African and Asian 9-bp deletion coalescence dates are substantially older. A young coalescent for the tribal Indian 9-bp deletion reflects the limited diversity of these haplotypes and is consistent with a recent de novo deletion event. Limited diversity could also result from a severe population bottleneck. It is unlikely that the 9-bp deletion found in the Irula and Yanadi results from recent Asian admixture, based on the large genetic distance between Asian and Indian 9-bp deletion groups.

The other Indian 9-bp deletion haplotypes suggest multiple origins of this motif in South India and underscore the genetic diversity present on the Indian subcontinent. Two individuals from the Maria Gond and a Yadava caste member had the deletion, and a Brahmin was heteroplasmic for this marker. The caste and Maria Gond 9-bp deletion haplotypes are distinct from those of the Irula and Yanadi. Based on phylogenetic analysis, it is likely that the haplotypes were introduced by a founding Asian deletion haplotype. The 9-bp deletion is also found in the Siddi population. An African-specific *Hpa*I$_{3592}$ mtDNA restriction enzyme site was detected in the Siddi population. This finding is consistent with historical reports of an African origin for the Siddi (Russell and Lal, 1916).

Nearly 25% of Nicobar Islanders have the deletion sequence. Since phylogenetic analyses place Nicobarese and Asian 9-bp deletion haplotypes together, it is likely that the Nicobarese deletion haplotypes are Asian-derived. This finding is consistent with linguistic and morphometric data suggesting an Asian origin of the Nicobarese (Das and Rath, 1991).

In N-J networks of global populations, tribal Indians with the 9-bp deletion show more similarity to other tribal Indians without the deletion than to other groups. Additionally, the relatively small differences in genetic distances between Asians, castes, and Caucasians suggest strong Asian and Caucasian components in modern populations on the Indian subcontinent. This finding is consistent with other reports indicat-

ing admixture in caste populations (Bamshead et al., 1998; Cavalli-Sforza et al., 1994). The small genetic distance between Asians, Indians, and Caucasians, and historical accounts of Asian/Caucasian immigration, suggest that additional population-specific markers will be highly useful for identifying genetic affinities between populations within India.

In addition to the 9-bp deletion, expansions and contractions of this sequence motif produce other polymorphic variants in South Indian populations. Expanded 9-bp repeats in the Khonda Dora and Santal tribes are population-specific. Mapping of the Khonda Dora expansion onto the individual HVS-1 N-J tree reveals a clustering of individuals with the expansion motif, indicating that this polymorphism identifies related mtDNA haplotypes within the Khonda Dora tribe. Among Indian castes, the Brahmin and Jalari have expanded repeats. HVS-1 sequence analysis indicates that Brahmin expanded alleles are not related to the Khonda Dora (HVS-1 data for the Jalari were not obtained). Expansion events also exist in appreciable frequencies in individuals from Borneo, the Philippines, and Samoa, who share closely related control region sequences (Lum and Cann, 1998).

The social and genetic features established and maintained in the highly endogamous Indian caste and tribal populations provide a unique opportunity for examining human evolution and population history. The development of panels of easily typed population-specific markers will improve our ability to understand the genetic relationships between the endogamous groups that comprise the populations of South India. Further analysis of the intergenic COII/tRNA^{Lys} variants in additional Indian populations, along with other mitochondrial, nuclear, and Y-chromosome markers, will provide information for resolving the complex genetic relationships between South Indian tribal populations and help to clarify questions about early inhabitants of the Indian subcontinent.

ACKNOWLEDGMENTS

The authors thank the individuals who participated in the study and acknowledge D. Rao, R. Pramuk, and members of the faculty and staff at Andhra University for their assistance and helpful discussions.

LITERATURE CITED

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139–152.

Bamshad M, Fraley AE, Crawford MH, Cann RL, Busi BR, Naidu JM, Jorde LB. 1996. mtDNA variation in caste populations of Andhra Pradesh, India. Hum Biol 68:1–28.

Bamshad M, Watkins WS, Dixon ME, Rao BB, Naidu JM, Prasad BVR, Reddy PG, Sung S, Rananayagam A, Hammer MF, Jorde LB. 1998. Female gene flow causes genetic stratification of the Hindu caste system. Nature 395:651–652.

Betty DJ, Chin-Atkins AN, Croft L, Sraml M, Easteal S. 1996. Multiple independent origins of the COII/tRNA(Lys) intergenic 9-bp mtDNA deletion in Aboriginal Australians. Am J Hum Genet 58:428–433.

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

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton: Princeton University Press. p 523.

Das HH, Rath R. 1991. The tribals of Andaman and Nicobar Islands. New Delhi: Ashish Publishing House. p 172.

Felsenstein J. 1993. PHYLIP (phylogeny inference package), version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

HariHara S, Hirai M, Suutou Y, Shimizu K, Omoto K. 1992. Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. Hum Biol 64:161–166.

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

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, Soodyall H, Jenkins T, Rogers AR. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear gentic data. Am J Hum Genet 57:523–538.

Lum JK, Cann RL. 1998. mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. Am J Phys Anthropol 105:109–119.

Majumder PP, Mukherjee BN. 1993. Genetic diversity and affinities among Indian populations: an overview. New York: Plenum Press. p 255–275.

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL. 1995. Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992.

Mujumder PP. 1998. People of India: biological diversity and affinities. Evol Anthropol 6:100–110.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press. p 512.

Papiha SS, Singh BN, Lanchbury JS, Mastana SS, Rao YS. 1997. Genetic study of the tribal populations of Andhra Pradesh, South India. Hum Biol 69:171–199.

Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS. 1996a. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59:927–934.

Passarino G, Semino O, Modiano G, Bernini LF, Santachiara Benerecetti AS. 1996b. mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals. Ann Hum Biol 23:121–126.

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M. 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615.

Russell RV, Lal RBH. 1916. Tribes and castes of the central provinces of India. Delhi: Cosmo Publications. p 419.

Schneider S, J-M K, Roessli D, Excoffier L. 1997. Alrequin version 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sirajuddin SM, Duggirala R, Crawford MH. 1994. Population structure of the Chenchu and other South Indian tribal groups: relationships between genetic, anthropometric, dermatoglyphic, geographic, and linguistic distances. Hum Biol 66:865–884.

Soodyall H, Jenkins T, Stoneking M. 1995. "Polynesian" mtDNA in the Malagasy. Nat Genet 10:377–378.

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T. 1996. mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608.

Stoneking M, Sherry S, Redd A, Vigilant L. 1992. New approaches to dating suggest a recent age for the human mtDNA ancestor. Philos Trans R Soc Lond [Biol] 337:167–175.

Strimmer KS, von Haeseler A. 1997. PUZZLE 4.0, maximum likelihood analysis for nucleotide, amino acid, and two-state data. Max-Planck-Institute for Biochemistry: Martinsried, Germany.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526.

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.

Yang Z, Kumar S. 1996. Approximate methods for estimating the pattern of nucleotide substitution and the variation of subsitution rates among sites. Mol Biol Evol 13:650–659.