AMERICAN JOURNAL OF HUMAN BIOLOGY 18:461–469 (2006)

Original Research Article

Asian and Non-Asian Origins of Mon-Khmer- and Mundari-Speaking Austro-Asiatic Populations of India

VIKRANT KUMAR,[1] BANRIDA T. LANGSITEH,[2] SILPAK BISWAS,[1] J. PRADEEP BABU,[1]
T. NAGESWARA RAO,[1] K. THANGARAJ,[3] A.G. REDDY,[3] L. SINGH,[3] AND B. MOHAN REDDY[1]*
[1]*Biological Anthropolgy Unit, Indian Statistical Institute, Habsiguda, Hyderabad 500 007, India*
[2]*Department of Anthropology, North Eastern Hill University, Shillong 793 014, India*
[3]*Center for Cellular and Molecular Biology, Hyderabad 500 007, India*

ABSTRACT In the present study, we analyzed 1,686 samples from 31 tribal populations of India for the mitochondrial DNA 9-base-pair deletion/insertion polymorphism, and characterized them based on the relevant mitochondrial DNA coding-region single nucleotide polymorphisms and hypervariable region I motifs, to test the genetic origins of the ethnically and linguistically heterogeneous Austro-Asiatic tribes of India. A comparative analysis of our results with the existing data suggests multiple origins of Austro-Asiatic tribes in India, and particularly the Asian and non-Asian origins of the Mon-Khmer and the Mundari populations. We also identified a novel subclade of haplogroup B in the Mon-Khmer Khasi tribes that distinguishes them from the Nicobarese, indicating two different waves of migration of the Mon-Khmer tribes in India. Am. J. Hum. Biol. 18:461–469, 2006.

The Indian subcontinent has been peopled by different linguistic groups (Austro-Asiatics, Dravidians, Tibeto-Chinese, and Indo-Europeans) at different time periods. About 30 tribal populations, representing three broad subfamilies, the Munda, the Khasi-Khmuic, and the Mon-Khmer (Diffloth, 2005), represent the Austro-Asiatic linguistic family in India. While the Mundari populations are ethnically Austroloids, the groups speaking Khasi-Khmuic and the Mon-Khmer languages have Mongoloid features and are probably of Asian origin, making this linguistic group ethnically more heterogeneous compared to other linguistic groups in India. Therefore, it may be logical to hypothesize that these populations are genetically heterogeneous and may have different genetic origins. Further, Austro-Asiatic populations were hypothesized to be the first to have arrived in India (for further details, see Kumar and Reddy, 2003), but the probable route(s) by which these people entered India is a contentious issue.

The length variation in mitochondrial DNA (mtDNA) due to the 9-base-pair (bp) deletion/insertion (del/ins) in the intergenic region between cytochrome oxidase subunit II and tRNAlysine (COII/$tRNA^{Lys}$) and its characterization, based on mtDNA hypervariable region I (HVR1), have been widely used for examining genetic relationships of human populations carrying this variation, and for tracing migrations (Hertzberg et al., 1989; Redd et al., 1995; Soodyall et al., 1995, 1996; Clark et al., 2000; Prasad et al., 2001). It was demonstrated to be a useful marker for elucidating the evolutionary history of many populations such as Pacific groups (Hertzberg et al., 1989; Stoneking and Wilson, 1989; Horai and Hayasaka, 1987, Ballinger et al., 1992; Harihara et al., 1992; Lum et al., 1994; Melton et al., 1995; Redd et al., 1995), Nicobarese (Prasad et al., 2001), Southeast Asians including Pacific (Yao et al., 2000), and Northeast Indian populations (Clark et al., 2000).

Many recent studies showed the presence of the 9-bp deletion, particularly in many tribal populations of India (Watkins et al., 1999; Clark et al., 2000; Reddy et al., unpublished findings), including the Nicobarese,

who are Austro-Asiatic (Watkins et al., 1999; Prasad et al., 2001; Thangaraj et al., 2005a), and certain tribes such as the Yanadi show a consistently high frequency (>40%) of the 9-bp deletion, even when sampled from different locations (Reddy et al., unpublished findings; Watkins et al., 1999). However, Majumder (2001) studied some Indian populations (which included only three Mundari groups with very small sample sizes) and, based on the absence of the 9-bp deletion among them, speculated that Austro-Asiatics (Mundari) might have arrived in India through its western borders, from Africa. The basis to suggest this is no longer valid, because the 9-bp deletion is not confined to Asian populations, and its presence is universally reported (Vigilant, 1990; Chen et al., 1995; Soodyall et al., 1996). Another study (Prasad et al., 2001), confined to only the Nicobarese belonging to the Mon-Khmer branch of the Austro-Asiatic linguistic group, suggested that the Nicobarese might have come from Southeast Asia during the process of westward colonization. However, most of the above studies (except that of Thangaraj et al., 2005a) did not characterize the 9-bp del/ins samples based on the mtDNA coding region single-nucleotide polymorphisms (SNPs).

The above studies, dealing with Austro-Asiatic populations in India, represent only a very small proportion of these populations, and hence are not adequately representative. Moreover, the Khasi subgroups, which are the only populations representing the Khasi-Khmuic branch of the Austro-Asiatic family from mainland India, have not yet been studied. The importance of these populations is further highlighted by the fact that these are the only Austro-Asiatic populations that presently inhabit the northeastern region of India, which probably served as a corridor for migrations into and out of India. To resolve the status of Austro-Asiatic tribes in the peopling of India, it is imperative to include as many Austro-Asiatic groups as possible, and to represent the maximum heterogeneity inherent among them. For this purpose, we sampled individuals from 15 Mundari groups and seven transitional groups (probably former Austro-Asiatic groups, now speaking different languages) from Chota-Nagpur and surrounding regions of India. Further, to capture the genetic heterogeneity within a population, most groups were sampled from more than one area. We also sampled 433 individuals from all eight subgroups of the Khasi, along with the Garo (a Tibeto-Burman group) from northeast India. With this exhaustive coverage of Austro-Asiatic groups and the comparison of these data with published material on other pertinent populations, we test if the ethnically distinct Austro-Asiatic tribes of India have genetically distinct origins and migration histories.

MATERIALS AND METHODS

Sampling

In the present study, most of the tribes of the Austro-Asiatic linguistic family, along with some neighboring populations which are considered to be transitional groups, were included. The names of populations, their linguistic affiliations, the areas of sampling, and sample sizes are given in Table 1. We analyzed a total of 1,686 samples from 31 tribal groups, of which 23 are Austro-Asiatic. Fifteen of 23 Austro-Asiatic populations speak Mundari languages, and all 15 groups inhabit the Chota-Nagpur area (the adjoining districts of the states of West Bengal, Orissa, Jharkhand, and Chattisgarh). The remaining eight Austro-Asiatic groups are subgroups of the Khasi from northeast India (Meghalaya), which are affiliated with the Khasi-Khmuic branch of the Austro-Asiatic linguistic family. Of the remaining eight groups, seven are from the Chota-Nagpur area. These populations, although now speaking non-Austro-Asiatic languages and at present identified with neighboring non-Austro-Asiatic groups, are popularly believed to be former Austro-Asiatic groups whose language, over a period of time, might have assimilated into the dominant languages spoken by neighboring populations. We therefore categorized them as transitional groups. The Garo is the other group from northeast India, which speaks a Tibeto-Burman language but inhabits contiguous areas of the Khasi. We also used comparative data from the available sources for the Nicobarese (Prasad et al., 2001; Thangaraj et al., 2005a; Watkins et al., 1999) and populations from other parts of India (Thangaraj et al., 2005a; Clark et al., 2000; Watkins et al., 1999; Kivisild et al., 2003; Metspalu et al., 2004), China (Melton et al., 1998; Yao et al., 2002; Kivisild et al., 2002), and Thailand (Fucharoen et al., 2001).

TABLE 1. Names, linguistic affiliations, areas of sampling, sample sizes and frequency of 9-bp del/ins motifs in Austro-Asiatic and transitional populations of India[1]

Name of poplations	Linguistic affiliations	Area of sampling	Sample size	9-bp del/ins (proportion)
Birjia	AA (Mundari)	Jharkhand	24	0
Asur	AA (Mundari)	Jharkhand	60	0
Korwa	AA (Mundari)	Jharkhand and Chattisgarh	52	0
Munda	AA (Mundari)	Jharkhand and Orissa	54	0
Bhumij	AA (Mundari)	Jharkhand and West Bengal	91	1 (0.01)
Mahali	AA (Mundari)	Jharkhand and West Bengal	32	0
Mudi	AA (Mundari)	Jharkhand and West Bengal	46	0
Savar	AA (Mundari)	Jharkhand and West Bengal	66	0
Santhal	AA (Mundari)	Jharkhand, Orissa, and West Bengal	111	1 (0.01)[2]
Kharia	AA (Mundari)	Jharkhand, Orissa, and West Bengal	52	1 (0.02)[2]
Birhor	AA (Mundari)	Jharkhand, Orissa, and West Bengal	58	0
Juang	AA (Mundari)	Orissa	52	0
Kolho	AA (Mundari)	Orissa	84	1 (0.01)[2]
Lodha	AA (Mundari)	West Bengal	70	0
Lyngngam	AA (Khasi-Khmuic)	Meghalaya	82	0
Nongtrai	AA (Khasi-Khmuic)	Meghalaya	29	0
Maram	AA (Khasi-Khmuic)	Meghalaya	72	6 (0.08)
Khynriam	AA (Khasi-Khmuic)	Meghalaya	94	0
Pnar	AA (Khasi-Khmuic)	Meghalaya	69	0
Bhoi	AA (Khasi-Khmuic)	Meghalaya	34	0
War-Khasi	AA (Khasi-Khmuic)	Meghalaya	31	0
War-Jaintia	AA (Khasi-Khmuic)	Meghalaya	22	0
Nagesia	Dravidian	Chattisgarh	15	1 (0.067)
Oraon	Dravidian	Jharkhand and Chattisgarh	97	0
Paharia	Dravidian	Jharkhand and West Bengal	11	1 (0.09)
Kanwar	IE	Chattisgarh	45	1 (0.02)
Pando	IE	Chattisgarh	30	0
Bathudi	IE	Orissa	38	2 (0.05)[2]
Bhuyan	IE	Orissa	78	2 (0.03)[3]
Garo	Tibeto-Burman	Meghalaya	87	1 (0.01)
Total			1,686	18 (0.01)

[1]AA, Austro-Asiatic; IE, Indo-European.
[2]Individuals with inserted motifs.
[3]Only one individual with inserted motifs.

DNA analysis

About 5 ml of intravenous blood were collected from each individual in a tube containing EDTA as anticoagulant. All blood samples were collected with the informed written consent of every donor. DNA was isolated following standard protocol (Maniatis et al., 1989). Primers (8296-5′-ATG CTA AGT TAG CTT TAC AG-3′ and 8297-5′-ACA GTT TCA TGC CCA TCG TC-3′), flanking the 9-bp del/ins motif, present between COII and $tRNA^{Lys}$, were used for amplification. The amplicons were size-fractionated using 6% polyacrylamide gel electrophoresis. Samples showing the 9-bp del/ins were further characterized by sequencing HVRI and typing SNPs in the coding regions of mtDNA, using the protocol described elsewhere (Macaulay et al., 1999; Thangaraj et al., 1999; Quintana-Murci et al., 2004). Phylogenetic relationships between the observed haplotypes were reconstructed by using the NETWORK program (Bandelt et al., 1999; http://www.fluxus-engineering.com).

RESULTS

Nine-base-pair del/ins frequency

In total, 1,686 individuals from 31 tribal groups were screened for the 9-bp del/ins polymorphism, and only 18 (0.01) individuals showed the presence of the 9-bp del/ins (Table 1), of whom seven were from northeast India (six from subtribes of the Khasi, and one from the Garo), with the rest from the Chotanagpur region (four from Austro-Asiatic groups, i.e., one each from the Bhumij, Kolho, Santhal, and Kharia, and the remaining seven from the transitional groups). The analysis of HVR1 of 17 9-bp del/ins samples (one sample of Kolho from Orissa could not be sequenced) showed 11 different haplotypes (Table 1). Out of the 10 Mundari and transitional-group 9-bp del/ins samples, we found nine of them to have different HVRI motifs (except for one Kharia and one Bathudi individual sharing the same motif), suggesting multiple origins for this polymorphism in the Mundari groups, and also implying a probable greater antiquity of these populations.

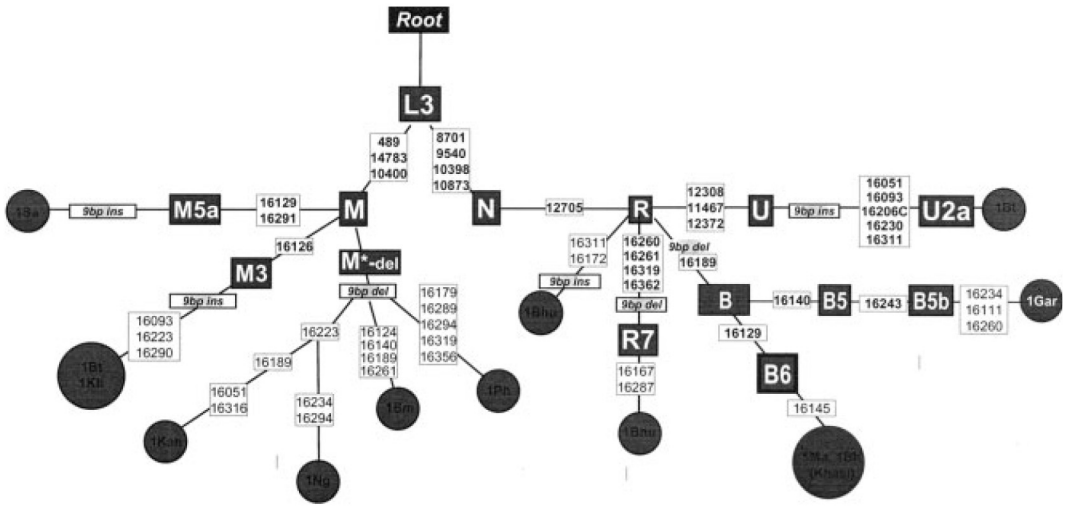

Fig. 1. Median-joining network of 9-bp del/ins individuals. Sa, Santhal; Bt, Bathudi; Kh, Kharia; Ng, Nagesia; Kan, Kanwar; Bm, Bhumij; Ph, Paharai; Bhu, Bhuiyan; Ma, Maram; Bh, Bhoi; Gar, Garo.

Haplogrouping of 9-bp del/ins individuals

Seventeen 9-bp del/ins individuals were further characterized into haplogroups on the basis of mtDNA coding region SNPs (Fig. 1). While a Garo (Tibeto-Burman) sample falls into haplogroup B5b, the six Khasi samples (5 Maram and 1 Bhoi) fall into haplogroup B6, a new subclade of haplogroup B, identified in this study (Fig. 1). It is interesting that the Nicobarese samples (Prasad et al., 2001; Thangaraj et al., 2005a), which belong to the Mon-Khmer branch, unlike Khasis, who belong to the Khasi-Khmuic branch of the Austro-Asiatic linguistic family, fall into either the B5 or B5a or B5a1 haplogroups (Fig. 2), except for one sample, which falls into haplogroup M. Although all the 9-bp-deleted Mundari samples fall broadly into haplogroup M, each one of them falls into a different subhaplogroup (Fig. 1). The Bhumij sample falls into an unclassified haplogroup M*, whereas the Santhal and Kharia samples fall into the M5a and M3 haplogroups, respectively. The remaining seven samples of transitional groups fall into both the M and N haplogroups (Fig. 1). Within haplogroup N, two Bhuiyan samples fall into haplogroups R and R7, and one Bathudi sample falls into haplogroup U2a, whereas in haplogroup M, except for one Bathudi sample which falls into haplogroup M3, all samples fall into haplogroup M*, but each with a different HVR1 motif.

The presence of haplogroup M3 with the same HVR1 motif in the Bathudi and Kharia tribal groups, and the sharing of haplogroup B6 with the same HVRI motif between the Maram and Bhoi, strongly suggest gene flow between the adjacent populations. Further, the 9-bp del/ins in one Santhal sample of our study shares the same haplogroup, as well as HVRI motif, with one sample of the Oraon and Yadava, while one Bhuyian sample shares its haplogroup and the HVRI motif with one Kanwar sample of the previous study (Thangaraj et al., 2005a). Similarly, a Kanwar sample in our study shares the same HVRI motif and haplogroup with that of a Munda, while the Paharia sample shares the same HVRI motif, except at position 16356, with another Munda sample (Roychoudhury et al., 2001).

When we compare our results with other groups of India and Southeast Asia, we find that all the Khasi-Khmuic 9-bp del/ins individuals of this study share the same haplogroup (newly identified haplogroup B6) with the Han Chinese of Guangdong in South China (Fig. 2), suggesting close genetic affinities between these two groups. Although the 9-bp del/ins samples of the Khasi and Nicobarese, which are both linguistically Austro-Asiatic and ethnically Mongoloid groups, fall into the broad haplogroup B, further classification shows that only the Nicobarese belong to haplogroup B5a, while all the 9-bp del/ins samples of the Khasi fall into haplogroup B6. The Nicobarese, which have 16140-16189-16266A HVRI motifs, share them with the Han Chinese

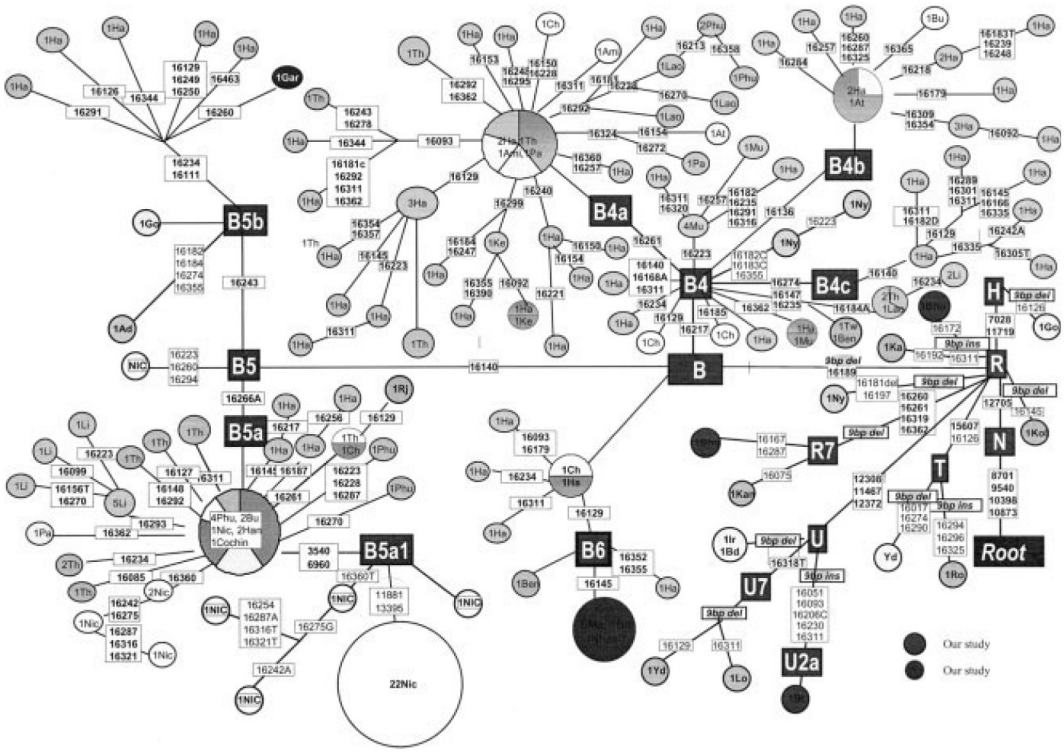

Fig. 2. Median-joining network of 9-bp del/ins of South, Southeast, and East Asia individuals of N haplogroup. Sources for comparative data: Ha, Han Chinese; Pa, Paiwan; Ke, Kanet; Li, Lisu; Phu, Phuthai; Th, Thai; Nic, Nicobarese; Ch, Chong; At, Atayal; Mu, Mussur; Bu, Bunun; Lao, Lao Song; Yd, Yadava; Go, Gond; Ka, Kafthodi; Kol, Koli; Ro, Roidas; Ny, Nishi; Kan, Kanwar; I, Irula; Bd, Baduga; Lo, Lodhe; Ben, Bengali; Rj, Rajput; Ad, Adi; Bhu, Bhuiyan; Bt, Bathudi; Ma, Maram; Bh, Bhoi; Gar, Garo; Tw, Tiwari. Ami, Atayal, Bunun, and Paiwan from Melton et al. (1998). Han Chinese from Yao et al. (2002) and Kivisild et al. (2002). Cochin and Kanet from Metspalu et al. (2004). Nicobarese from Prasad et al. (2001), Thangaraj et al. (2005a), and Watkins et al. (1999). Chong, Lao Song, Lisu, Mussur, Phuthai, and Thai from Fucharoen et al. (2001). Nishi from Clark et al. (2000) and Thangaraj et al. (2005a). Adi from Clark et al. (2000). Yadava, Gond, Kafthodi, Koli, Roidas, Nishi, Kanwar, Irula, Baduga, Lodhe, Bengali, Rajput, and Tiwari from Thangaraj et al. (2005a).

(from the provinces of Yunnan, Wuhan, Xinjiang, and Liaoning) and the Paiwan of China, with the Phuthai, Lisu, and Thai groups of Thailand, and surprisingly, with one sample from Kerala in India, but not with the Khasi-Khmuic populations of our study. Thus the two geographically and linguistically divergent groups of India show not only widely different frequencies of broad haplogroup B, but also the presence of different subhaplogroups, which argues strongly for different maternal origins and probably two different waves of migration into India: one by the Khasi, and the other by the Nicobarese. Further, the haplogroups and HVR1 motifs of 9bp-del/ins samples of the Mundari and neighboring transitional groups have not been found either in the other Austro-Asiatic linguistic subgroups or in the East Asian or Southeast Asian groups (Figs. 2, 3).

DISCUSSION

The comprehensive coverage of Austro-Asiatic groups in our study (23 in all: 15 Mundari, and 8 Khasi-Khmuic), besides the seven transitional groups from in and around Chota-Nagpur plateau, and the comparative analyses of our results with the existing data from different parts of India and Central, East, and Southeast Asia, including the Pacific and Africa, provide an adequate basis and insights to derive formidable conclusions. However, out of a total of 1,686 individuals from 31 tribal groups screened for the 9-bp del/ins, only 18 (0.01) showed the 9-bp del/ins (Table 1), of whom only 4 and 6 individuals are from the Mundari and Khasi-Khmuic tribes, respectively, suggesting a negligible presence of this polymorphism in the Austro-Asiatic popula-

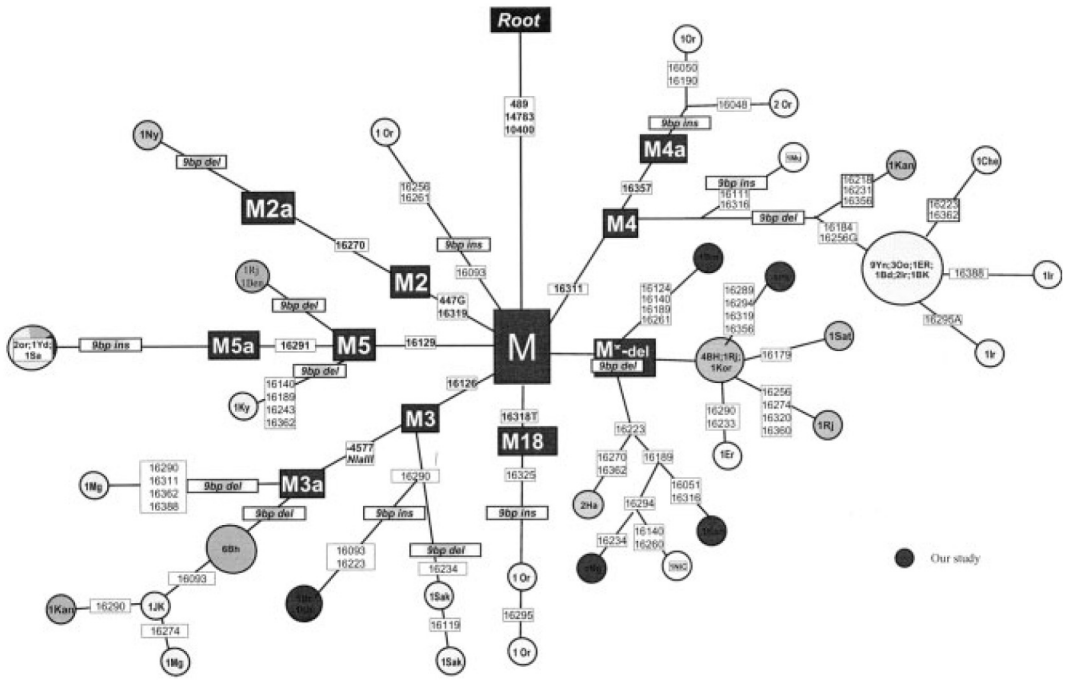

Fig. 3. Median-joining network of 9-bp del/ins of South, Southeast, and East Asia individuals of M haplogroup. Sources for comparative data: Sat, Satnami; Rj, Rajput; Bh, Bharia; Kor, Kori; Or, Oraon; Oo, Oorali; Yn, Yanadi; Er, Erukula; Bd, Baduga; Kan, Kanwar; Sak, Sakkili; Ny, Nishi; Mu, Munda; Yd, Yadava; Ben, Bengali; Che, Chenchu; Ng, Nagesia; Bm, Bhumij; Ph, Paharia; Bt, Bathudi; Kh, Kharia; Mg, Maria Gond; I, Irula; JK, Jenu Kurumba; Bk, Betta Kurumba; Nic, Nicobarese; Ha, Han Chinese; Ky, Koya. Han Chinese from Yao et al. (2002). Satnami, Rajput, Bharia, Kori, Oraon, Oorali, Yanadi, Erukula, Baduga, Kanwar, Sakkili, Nishi, Munda, Yadava, and Bengali from Thangaraj et al. (2005a). Irula and Maria Gond from Watkins et al. (1999). Jenu Kurumba and Betta Kurumba from Clark et al. (2000). Nicobarese from Prasad et al. (2001). Chenchu and Koya from Kivisild et al. (2003).

tions of this study, although the Mon-Khmer Nicobarese (studied earlier) displayed a relatively much higher frequency of the 9-bp del/ins. Nevertheless, most of the 9-bp del/ins are on different HVRI backgrounds in the Indian samples. For example, all the Mundari and non-Mundari transitional samples showing the 9-bp deletion have different HVRI motifs (except for one Kharia and one Bathudi individual sharing the same motif), suggesting multiple origins for this polymorphism. Further, for this polymorphism to have recurred at least nine times in these groups, the populations need to be quite old, and therefore the frequency of the 9-bp del/ins polymorphism suggests a greater antiquity of the Mundari population. Most of the haplogroups found in the Mundari and the transitional groups fall either into the M* (unclassified) or M5a or M3 subclades of superhaplogroup M, while the rest fall into either haplogroup R or R7 or U2a of superhaplogroup N. Haplogroups M3 and U2a were also reported from the populations of Sindh and North-West Frontier Province from Pakistan (Kivisild et al., 1999; Cordaux et al., 2003; Quintana-Murci et al., 2004), which served as a major western corridor for populations to enter India. Further, all haplogroups found in the Mundari and transitional groups of our study were suggested to have comigrated from Africa through the southern coastal route around 50,000–70,000 years BP (Kivisild et al., 2003; Thangaraj et al., 2003, 2005b; Metspalu et al., 2004). Most significantly, none of the samples of the Mundari and neighboring transitional groups fall into haplogroup B or its subclades, or into any other haplogroup with a putative origin and common occurrence in East and Southeast Asia. All these strongly suggest non-Asian/African origin of the Mundari groups. Given the linguistic similarity of the Mundari populations with some of those in Southeast Asia, and assuming that these are among the first to arrive in the region from

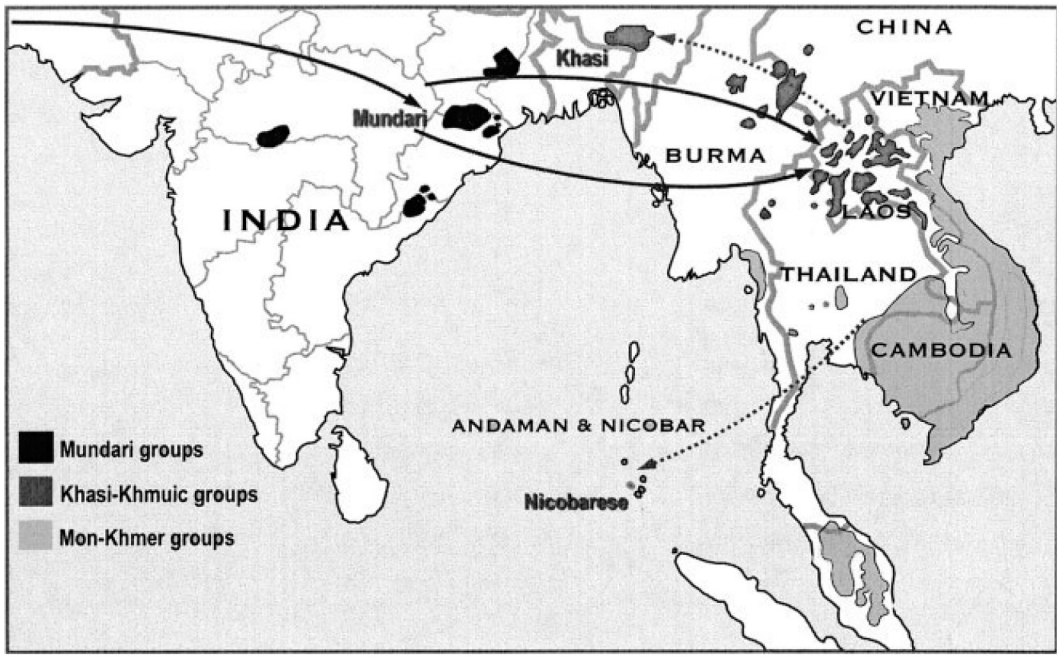

Fig. 4. Map showing present-day distribution of Austro-Asiatic groups (as depicted in van Driem, 2001) and schematic representation of routes of migration of different Austro-Asiatic linguistic subgroups of India. Dotted arrows indicate back-migration.

Africa, there are grounds to surmise the early migration of these populations into Southeast Asia, either through the northeastern corridor or via the sea route, as depicted in Figure 4. This hypothesis is substantiated by the composition of Y-haplogroups in the Mundari and Southeast Asian populations (Kumar et al., unpublished findings).

The Asian subclades of haplogroup B were named B4, characterized by the 16217 mutation, and B5, characterized by the 16140 mutation (Yao et al., 2002). Individuals who lack both these mutations are identified as haplogroup B. However, our analysis suggests a novel subclade (haplogroup B6), defined by the presence of a transition at 16129 (Figs. 1, 2) and the absence of 16217 and 16140 mutations, which characterize all six Khasi samples of our study, besides one individual each of the Han Chinese (Yao et al., 2002) and Bengalis of India (Thangaraj et al., 2005a).

The most interesting inference emerging from this study is the probable indication of multiple origins/migrations of Austro-Asiatic groups into the Indian subcontinent. What is surprising is that the 9-bp del/ins samples of the Khasi from northeastern India and the Nicobarese from the Nicobar Islands, who are ethnically similar and were earlier perceived to belong to the same Mon-Khmer subfamily, show vastly different frequencies of broad haplogroup B, and do not fall into the same subhaplogroup (Fig. 2). However, these findings are consistent with the current understanding of the linguistic affiliations of these groups. Diffloth (2005) classified the Khasi as belonging to the Khmuic and the Nicobarese to the Mon-Khmer branches of the Austro-Asiatic linguistic family. The Khasi samples share haplogroup B6 with the Han Chinese of Guangdong (a province in South China). The Khasi-Khmuic group is presently spread through Laos, Burma, and the regions of South China and Thailand bordering Burma. This suggests that haplogroup B6 originated in and around Laos/South China, and was brought to northeastern India by the initial Khasi-Khmuic migrants through Burma (Fig. 4). The absence of haplogroups B4 and B5 and its subclades in the Khasi, which are very common in East and Southeast Asia, suggests that the migration of these people might have taken place prior to the occurrence of these mutations in mtDNA that would have led to

the formation of these haplogroups. Kong et al. (2003) reported the age estimate of haplogroup B (to which haplogroup B6 belongs) as 50,800 ± 6,600 (SD) years. This indicates that the Khasi population might have entered India not earlier than ~50,000 BP through the northeast corridor of India.

The presence of haplogroup B6 in the Khasi due to gene flow from neighboring Tibeto-Burman groups is unlikely, because none of the 9-bp del/ins samples of the Tibeto-Burman Garo population of our study, which are culturally akin and geographically close to the Khasi, or the samples of other Tibeto-Burman populations from this region, fall into either haplogroup B6 or B. Since the Khasi are a matrilocal group, in which male moves to his spouse's house after marriage, the maternal gene flow into this population is restricted if not totally discounted.

Another wave of probable migration by the Mon-Khmer from Southeast Asia can probably be inferred from the fact that the HVR1 motif 16140-16189-16266A is shared by the Nicobarese, the Han Chinese from the provinces of Yunnan, Wuhan, Xinjiang, and Liaoning of China, and the Phuthai, Lisu, and Thai groups of Thailand, but not by the Khasi-Khmuic group of our study. The distribution of Mon-Khmer groups is currently restricted to Vietnam, South Laos, Cambodia, Thailand, and coastal southern Burma, which suggests that the Mon-Khmer Nicobarese might have migrated from Southeast Asia through Thailand and coastal southern Burma to Nicobar Island (Fig. 4). All these groups have either haplogroup B5 or B5a or B5a1. This migration is suggested to have occurred in the Neolithic era, during the demic expansion of agriculturalists (Thangaraj et al., 2003, 2005a; Underhill et al., 2000; Prasad et al., 2001). Thangaraj et al. (2005b) also found haplogroup F1a1a1, which was also observed in China, Malaysia, and Thailand (Ingman et al., 2000; Yao et al., 2002), and suggested their close relationship with Southeast Asians and their recent arrival from the Southeast during the past 18,000 years.

In a nutshell, one may surmise that the Austro-Asiatic populations of India have come in multiple waves of migration, and the ancestors of present-day Mundari groups might have been the first to arrive in India through the western Indian corridor, subsequently migrating to Southeast Asia. This was probably followed by the migration of the Khasi and later by the Nicobarese from Southeast Asia (Fig. 4). Overall, the results not only suggest distinct genetic origins of the Austro-Asiatic linguistic subgroups, but also indicate a non-Asian source of migration of the Mundari populations of India, which is further substantiated by analyses of Y-SNPs among Austro-Asiatic groups (Kumar et al., 2005).

ACKNOWLEDGMENTS

This work is essentially part of two planned projects of the Indian Statistical Institute (ISI), being carried out in collaboration with the Centre for Cellular and Molecular Biology (CCMB). We thank the directors of the ISI and CCMB for their support. Our thanks go to Mr. G. Chaubey for his suggestions and help in constructing the Median-Joining network.

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.

Bandelt H, Forster P, Rohl A. 1999. Median joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Chen YS, 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.

Clark VJ, Sivendren S, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M. 2000. The 9-bp deletion between the mitochondrial lysine tRNA and COII genes in tribal populations of India. Hum Biol 72:273–285.

Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. Eur J Hum Genet 11:253–264.

Diffloth G. 2005. The contribution of linguistic palaeontology to the homeland of Austro-Asiatic. In: Sagart L, Blench R, Sanchez-Mazas A, editors. The peopling of East Asia: putting together archaeology, linguistics and genetics. London: Routledge Curzon. p 77–81.

Fucharoen G, Fucharoen S, Horai S. 2001. Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46:115–125.

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-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510.

Horai S, Hayasaka K. 1987. Evolutionary implications of mitochondrial DNA polymorphisms in human populations. In: Vogel F, Sperling K, editors. Human genetics: proceedings of the 7th International Congress. Berlin: Springer. p 177–181.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713.

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R. 1999. Deep common ancestry of Indian and Western-Eur-

asian mitochondrial DNA lineages. Curr Biol 9:1331–1334.
Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751.
Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72:313–332.
Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73:671–676.
Kumar V, Reddy BM. 2003. Status of Austro-Asiatic groups in the peopling of India: an exploratory study based on the available prehistoric, linguistic and biological evidences. J Biosci 28:507–522.
Lum JK, Rickards O, Ching C, Cann RL. 1994. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590.
Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249.
Majumder PP. 2001. Ethnic populations of India as seen from an evolutionary perspective. J Biosci 26:533–545.
Maniatis T, Sambrook J, Fritsch EF. 1989. Molecular cloning: a laboratory manual,2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, Martinson J, Stoneking M. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 57:403–414.
Melton T, Clifford S, Martinson J, Batzer M, Stoneking M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63:1807–1823.
Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5:26.
Prasad BV, Ricker CE, Watkins WS, Dixon ME, Rao BB, Naidu JM, Jorde LB, Bamshad M. 2001. Mitochondrial DNA variation in Nicobarese Islanders. Hum Biol 73: 715–725.
Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K. 2004. Where West meets East: the complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet 74:827–845.
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.
Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan H, Usha Rani MV, Sil SK, Mitra M, Majumder PP. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. Hum Genet 109:339–350.
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, Wilson A. 1989. Mitochondrial DNA In: Hill AVS, Serjeantson SW, editors. The colonization of the Pacific: a genetic trail. Oxford: Clarendon Press. p 215–245.
Thangaraj K, Ramana GV, Singh L. 1999. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. Electrophoresis 20:1743–1747.
Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG, Hagelberg E. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. Curr Biol 2003:86–93.
Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, Mallick CB, Kumar N, Velavan TP, Suganthan R, Udaykumar D, Kumar R, Mishra R, Khan A, Annapurna C, Singh L. 2005a. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. Hum Genet 116:507–517.
Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L. 2005b. Reconstructing the origin of Andaman Islanders. Science 13:996.
Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361.
van Driem G. 2001. Languages of the Himalayas: an ethnolinguistic handbook of the Greater Himalayan region with a brief introduction to the symbiotic theory of language (2 volumes). Leiden: Brill.
Vigilant L. 1990. Control region sequence from African populations and the evolution of human mitochondrial DNA. Ph.D. thesis, University of California at Berkeley.
Watkins WS, Bamshad M, Dixon ME, Bhaskara Rao B, Naidu JM, Reddy PG, Prasad BV, Das PK, Reddy PC, Gai PB, Bhanu A, Kusuma YS, Lum JK, Fischer P, Jorde LB. 1999. Multiple origins of the mtDNA 9-bp deletion in populations of South India. Am J Phys Anthropol 109:147–158.
Yao YG, Watkins WS, Zhang ,YP. 2000. Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of East and Southeast Asia. Hum Genet 107:504–512.
Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651.